Environ Ecol Stat (2012) 19:309–326
DOI 10.1007/s10651-012-0188-1

Nonparametric mean estimation using partially ordered sets

Jesse Frey

Received: 18 March 2011 / Revised: 6 February 2012 / Published online: 15 March 2012

Abstract In ranked-set sampling (RSS), the ranker must give a complete ranking of the units in each set. In this paper, we consider a modification of RSS that allows the ranker to declare ties. Our sampling method is simply to break the ties at random so that we obtain a standard ranked-set sample, but also to record the tie structure for use in estimation. We propose several different nonparametric mean estimators that incorporate the tie information, and we show that the best of these estimators is substantially more efficient than estimators that ignore the ties. As part of our comparison of estimators, we develop new results about models for ties in rankings. We also show that there are settings where, to achieve more efficient estimation, ties should be declared not just when the ranker is actually unsure about how units rank, but also when the ranker is sure about the ranking, but believes that the units are close.

Keywords Imperfect rankings · Isotonic estimation · Judgment post-stratification · Ranked-set sampling · Ties in rankings

1 Introduction

Ranked-set sampling (RSS), proposed by McIntyre (1952, 2005), is a sampling scheme that typically outperforms simple random sampling in settings where precise measurements are costly, but small sets of units can easily be ranked or approximately ranked at negligible cost. McIntyre's work was motivated by applications in agriculture, but RSS has since been applied in many other areas, including environmental monitoring (Kvam 2003), forestry (Halls and Dell 1966), and medicine (Chen et al. 2005). Patil

J. Frey (✉)
Department of Mathematics and Statistics, Villanova University, Villanova, PA 19085, USA
e-mail: jesse.frey@villanova.edu

(1995) described a variety of applications of RSS in environmental monitoring and assessment.

RSS may be either balanced or unbalanced. To implement balanced RSS, one first decides on a set size m and a number of cycles n. One then draws nm independent simple random samples (sets) of size m and ranks the units in each set from smallest to largest. This ranking is done without making any precise measurements, and it need not be perfectly accurate. After putting the nm ranked sets in random order, one selects one unit from each set for precise measurement. From each of the first n sets, the unit ranked smallest in the set is selected for measurement. From each of the next n sets, the unit ranked second-smallest is selected for measurement, and so on. This yields a sample of nm independent values, with n values coming from each of the m possible in-set ranks. To implement unbalanced RSS, one drops the requirement that the sample include the same number of units from each in-set rank. Instead, one simply decides on a set size m and a vector $(n_1, \ldots, n_m)$ such that n_i gives the number of units with in-set rank i that are to be selected for precise measurement. For unbalanced RSS, the number of sets required is $\sum_{i=1}^{m} n_i$.

When ranking the units in a set, there will often be cases where the ranker is unsure about how two or more units should rank. However, implementing RSS requires a full ranking in each set. As a result, ties are not allowed, and if they occur, they must be broken at random. Ozturk (2011) described this difficulty, and he also proposed some variations on RSS in which rankers are permitted to declare ties. Under his first proposal, which is the most general, ties of any kind may be declared. Thus, the ranking process consists of dividing the units in each set into subsets $S_1, \ldots, S_k$ (with $k \leq m$) such that units in the same subset are tied, but each unit in set S_i is considered smaller than each unit in S_j if $i < j$. After this judgment subsetting process, a grouping of sets is carried out, and units are selected for precise measurement in such a way that exactly one unit is selected from each set. However, Ozturk (2011) provides few details about the grouping step, and since the grouping is done by the researcher, it appears that there is potential for bias to be introduced. An additional problem is that the theoretical results of Ozturk (2011) rely on the claim that having a particular subset of units be tied for ranks i to j tells us nothing about the average value for a unit drawn from that subset. We show in this paper that this claim does not always hold. In the remaining two proposals made by Ozturk (2011), there are restrictions on what ties may be declared, and in the method that Ozturk (2011) recommends for use, ties are actually forced. For example, one recommended scheme for set size $m = 6$ requires that the ranker split each set into three subsets: the smallest two, the middle two, and the largest two.

In this paper, we propose a sampling scheme and associated estimators that may be used no matter what ties are declared. The sampling scheme consists of breaking ties at random when deciding which units to select for precise measurement, but then also recording the tie structure. Thus, sampling with set size m and sample size vector $(n_1, \ldots, n_m)$, the full data would include not just the values $Y_{[i]j}$, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$, where $Y_{[i]j}$ is the precise measured value for the jth unit with rank i, but also indicator variables $I_{[i]jk}$ ($i = 1, \ldots, m$, $j = 1, \ldots, n_i$, $k = 1, \ldots, m$), where $I_{[i]jk}$ is 1 if the jth unit with rank i was tied for rank k and zero otherwise. Note that $\sum_{k=1}^{m} I_{[i]jk} \geq 1$ is the number of ranks for which the jth unit with rank i was tied.

In Sect. 2, we describe mean estimators that use data of this form. In Sect. 3, we discuss models that allow both for ties and for imperfect rankings, and we develop some new results about properties of these models. In Sect. 4, we use the models to compare the performance of the mean estimators discussed in Sect. 2, and in Sect. 5, we show that it is sometimes a good idea to declare a tie even when the ranker is sure about the correct ordering of two units. We show in Sect. 6 that the claim needed for implementing the most general method of Ozturk (2011) does not hold in general, and we conclude in Sect. 7 with a discussion.

2 Nonparametric mean estimators

Since data collected in the manner described in Sect. 1 constitutes a ranked-set sample, the standard RSS unbiased nonparametric mean estimator may still be used. This estimator ignores the tie information provided by the indicator variables $I_{[i]jk}$, and it is given by

$$\hat{\mu}_1 = \frac{1}{m}\sum_{i=1}^{m}\bar{Y}_{[i]},$$

where $\bar{Y}_{[i]} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{[i]j}$ is the sample mean for the ith (rank i) stratum.

One way to incorporate the tie information is to split each tied unit among the strata corresponding to the ranks for which the unit was tied. Specifically, if a measured unit is tied for r different ranks, then we assign it to each of those r strata, but with weight $1/r$ instead of 1. This strategy, which is a special case of a more general strategy proposed by MacEachern et al. (2004) in the context of judgment post-stratification (JPS), leads to the estimator

$$\hat{\mu}_2 = \frac{1}{m}\sum_{i=1}^{m}\bar{Y}'_{[i]},$$

where

$$\bar{Y}'_{[i]} = \frac{\sum_{l=1}^{m}\sum_{j=1}^{n_l} Y_{[l]j}\cdot\frac{I_{[l]ji}}{\sum_{k=1}^{m} I_{[l]jk}}}{\sum_{l=1}^{m}\sum_{j=1}^{n_l}\frac{I_{[l]ji}}{\sum_{k=1}^{m} I_{[l]jk}}}$$

is the weighted sample mean for the ith stratum.

It may happen that the sample means $\bar{Y}_{[1]},\ldots,\bar{Y}_{[m]}$ or $\bar{Y}'_{[1]},\ldots,\bar{Y}'_{[m]}$ don't come in the increasing order that the true stratum means must follow if the judgment strata are stochastically ordered. In this case, one might isotonize the estimates $\bar{Y}_{[1]},\ldots,\bar{Y}_{[m]}$ or $\bar{Y}'_{[1]},\ldots,\bar{Y}'_{[m]}$, using the sample sizes $n_1,\ldots,n_m$ or the weighted sample sizes $n'_i \equiv \sum_{l=1}^{m}\sum_{j=1}^{n_l}\frac{I_{[l]ji}}{\sum_{k=1}^{m} I_{[l]jk}}$, $i = 1,\ldots,m$, as the weights. By equation (4) in Wang et al. (2008), the isotonized versions of $\bar{Y}_{[1]},\ldots,\bar{Y}_{[m]}$ are given by

$$\bar{Y}_{[i],iso} = \max_{1 \le r \le i} \min_{i \le s \le m} \frac{\sum_{l=r}^{s} n_l \bar{Y}_{[l]}}{\sum_{l=r}^{s} n_l}$$

for $i = 1, \ldots, m$, and the isotonized versions of $\bar{Y}'_{[1]}, \ldots, \bar{Y}'_{[m]}$ are given by

$$\bar{Y}'_{[i],iso} = \max_{1 \le r \le i} \min_{i \le s \le m} \frac{\sum_{l=r}^{s} n'_l \bar{Y}'_{[l]}}{\sum_{l=r}^{s} n'_l}$$

for $i = 1, \ldots, m$. These isotonized in-stratum means may also be obtained by using the pool adjacent violators algorithm (PAVA), which is described in Chapter 1 of Robertson et al. (1988). The overall mean estimate is then the average of the isotonized in-stratum means. This isotonization strategy, proposed by Ozturk (2007) for the case of unbalanced RSS and by Wang et al. (2008) for JPS, leads to nonparametric mean estimators

$$\hat{\mu}_3 = \frac{1}{m} \sum_{i=1}^{m} \bar{Y}_{[i],iso}$$

and

$$\hat{\mu}_4 = \frac{1}{m} \sum_{i=1}^{m} \bar{Y}'_{[i],iso}.$$

Here $\hat{\mu}_3$ is the isotonized version of $\hat{\mu}_1$, and $\hat{\mu}_4$ is the isotonized version of $\hat{\mu}_2$.

Another way to incorporate the tie information is to consider the different ways that the ties might have been broken at random. This leads to Rao-Blackwellized versions of $\hat{\mu}_1$ and $\hat{\mu}_3$ that are thus guaranteed to be at least as good as $\hat{\mu}_1$ and $\hat{\mu}_3$, respectively, when assessed by mean squared error (MSE). These estimators are obtained by (i) considering all possible ways that the ties might have been broken at random and all possible ways that the choice of which rank to sample from each set might have been made, (ii) restricting attention to the cases where the units chosen for precise measurement would have been exactly those that were actually chosen for precise measurement, and then (iii) averaging together the resulting mean estimates. Since the units actually chosen for measurement could not be the units chosen for precise measurement unless their ranks matched the desired sample size vector $(n_1, \ldots, n_m)$, it is sufficient to average over all ways in which the ranks assigned to the units that were actually measured yield the desired sample size vector $(n_1, \ldots, n_m)$. We write $\hat{\mu}_5$ for the Rao-Blackwellized version of $\hat{\mu}_1$, and we write $\hat{\mu}_6$ for the Rao-Blackwellized version of $\hat{\mu}_3$. Rao-Blackwellizing $\hat{\mu}_2$ and $\hat{\mu}_4$ in this way does not lead to new estimators since $\hat{\mu}_2$ and $\hat{\mu}_4$ do not depend on how the ties were broken at random.

To illustrate how the estimators are computed, we consider an example. Suppose that we wish to estimate the average population for the 108 largest US cities. Using set size $m = 4$ and in-stratum sample sizes $(n_1, n_2, n_3, n_4) = (3, 3, 2, 2)$, we carried out the judgment subsetting process. The necessary 10 sets of size 4 were drawn by the author, and the judgment ranking with ties was done by a colleague. The sets were then put in random order, and the ties in judgment rankings were broken at random. From each of the first three sets, the unit with rank 1 was selected for precise measurement. From the each of the next three sets, the unit with rank 2 was selected, and so on. This process yielded the data shown in Table 1.

Each row of Table 1 corresponds to one set, and the cities in each set are listed according to their ranks after ties were broken at random. Cities that were declared tied are connected with ampersands, and the city that was selected for precise measurement is given in bold, together with its rank (after ties were broken at random) and, if applicable, the set of ranks that the city was tied for. We see from the table that just two of the cities selected for precise measurement were involved in ties. In set 6, Tulsa was tied for ranks 2 to 4, and in set 7, Minneapolis was tied for ranks 2 and 3.

Using the rank 2 for Tulsa and the rank 3 for Minneapolis, we find that $\bar{Y}_{[1]} = 253,838$, $\bar{Y}_{[2]} = 359,147$, $\bar{Y}_{[3]} = 531,009$, and $\bar{Y}_{[4]} = 825,911$. Thus, $\hat{\mu}_1 = 492,476$. If we split the tied units among the strata for the ranks for which they were tied, we obtain weighted sample sizes $(n'_1, \ldots, n'_4) = (3, 2.8\bar{3}, 1.8\bar{3}, 2.\bar{3})$. Computing the weighted sample means then gives $\bar{Y}'_{[1]} = 253,838$, $\bar{Y}'_{[2]} = 356,605$, $\bar{Y}'_{[3]} = 545,020$, and $\bar{Y}'_{[4]} = 763,584$. Thus, $\hat{\mu}_2 = 479,762$. Since $\bar{Y}_{[1]} < \bar{Y}_{[2]} < \bar{Y}_{[3]} < \bar{Y}_{[4]}$ and $\bar{Y}'_{[1]} < \bar{Y}'_{[2]} < \bar{Y}'_{[3]} < \bar{Y}'_{[4]}$, the isotonized estimators $\hat{\mu}_3$ and $\hat{\mu}_4$ satisfy $\hat{\mu}_3 = \hat{\mu}_1$ and $\hat{\mu}_4 = \hat{\mu}_2$ in this case. However, the Rao-Blackwellized estimators $\hat{\mu}_5$ and $\hat{\mu}_6$ differ from $\hat{\mu}_1$.

To compute the Rao-Blackwellized estimators, we must consider all the different ways that the ties might have been broken. Table 2 lists these possibilities. Among the six possibilities, there are only two for which the ranks assigned to the units chosen for precise measurement exactly match the desired sample size vector $(n_1, \ldots, n_4) = (3, 3, 2, 2)$. One of these ways gives the estimate 492, 476 (the value for $\hat{\mu}_1$ that we actually obtained), and the other gives the estimate 492, 653. Averaging these two estimates gives the Rao-Blackwellized estimate $\hat{\mu}_5 = 492,565$. Since neither of

Table 1 Judgment-ranked sets, tied values, and precise measured values for the Sect. 2 example

Set	Ranked units	Precise value
1	**Hialeah (1)**, Austin & Saint Paul, Chicago	218, 896
2	**Henderson (1)**, Tucson, Atlanta, Philadelphia	256, 445
3	**Anchorage (1)**, Long Beach, Portland, San Francisco	286, 174
4	Greensboro, **Fremont (2)**, Pittsburgh, Denver	205, 517
5	Irving, **Kansas City (2)**, Baltimore, San Jose	482, 299
6	Chula Vista, **Tulsa (2-2,3,4)** & Winston-Salem & Irvine	389, 625
7	Boise, Charlotte & **Minneapolis (3-2,3)**, Yonkers	385, 378
8	Milwaukee & Scottsdale, **Memphis (3)**, Riverside	676, 640
9	Colorado Springs & Modesto, Fort Wayne, **Newark (4)**	278, 154
10	Laredo, Albuquerque & Reno, **San Antonio(4)**	1, 373, 668

The cities are listed according to the post-tie-break full ranking, with cities that were declared tied connected by ampersands. Cities selected for measurement and the ranks for which they were tied are given in bold. True populations appear in the right-most column

Table 2 Different ways that the relevant ties might have been broken in the Sect. 2 example

Tulsa rank	Minneapolis rank	$(n_1, \ldots, n_4)$	Estimate
2	2	(3, 4, 1, 2)	N.A.
2	3	(3, 3, 2, 2)	492,476
3	2	(3, 3, 2, 2)	492,653
3	3	(3, 2, 3, 2)	N.A.
4	2	(3, 3, 1, 3)	N.A.
4	3	(3, 2, 2, 3)	N.A.

Only two of these possibilities could have lead to having the same set of units chosen for measurement that was actually chosen for measurement

the two ways leads to estimated in-stratum means that violate the increasing ordering that we would expect if the true stratum means are ordered, we also find that $\hat{\mu}_6 = \hat{\mu}_5 = 492,565$.

3 Models for rankings with ties

In this section, we describe two classes of models for ties in rankings. The first class was proposed by Fligner and MacEachern (2006), and the second class is similar, but different in some important ways. Both of these classes use the idea of perceived sizes that motivated the work of Dell and Clutter (1972) on imperfect rankings in RSS. Some alternate models for imperfect rankings in RSS were described by Bohn and Wolfe (1994) and by Frey (2007).

Consider a single set in which the ith unit has precise value Y_i. We assume that for each unit, there is also a perceived size (X_i for the ith unit), and that the pairs $(X_1, Y_1), \ldots, (X_m, Y_m)$ are independent draws from some bivariate distribution. In the Dell and Clutter (1972) model, the random variables $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_m$ satisfy $X_i = Y_i + \epsilon_i$, where $Y_1, \ldots, Y_m$ are independent draws from some distribution and $\epsilon_1, \ldots, \epsilon_m$ are independent normal errors with mean 0 and variance σ^2. The rankings are done according to the perceived sizes $X_1, \ldots, X_m$, and the precise measured values are the corresponding Y values. Thus, setting $\sigma^2 = 0$ gives perfect rankings, and letting σ^2 go to infinity gives rankings done at random. Fligner and MacEachern (2006) generalized the Dell and Clutter (1972) model by considering different bivariate distributions and nonadditive error structures. They also described a class of models for ties in rankings.

This class, which we call the *discrete perceived size* (DPS) class, involves discretizing the perceived size X. This can be done through an operation such as rounding X/c to the nearest integer or computing $\lfloor X/c \rfloor$, where $\lfloor \cdot \rfloor$ is the greatest integer function and $c > 0$ is a user-chosen model parameter. Such operations yield a discrete perceived size $X^\star$. Continuous perceived sizes are all distinct with probability 1, but discrete perceived sizes may be tied exactly. The DPS model for ties in rankings consists of declaring the ith and jth units tied whenever their discrete perceived sizes $X_i^\star$ and $X_j^\star$ satisfy $X_i^\star = X_j^\star$.

Another way to obtain ties is to declare the ith and jth units to be tied whenever $|X_i - X_j| < c$, where $c > 0$ is a user-chosen model parameter. Transitivity of ties is also required. Thus, the ith and jth units may be tied even if $|X_i - X_j| \geq c$, so long as there are other perceived sizes that bridge the gap. We refer to this class of models as *tied-if-close* (TIC) models. If c is small and the distribution of X is continuous, then very few ties will be declared, but as c goes to infinity, the chance of declaring an m-way tie involving all units in the set goes to 1. For a fixed c and a fixed distribution for X, there are more ties under a TIC model than under a DPS model.

Models in both the DPS and TIC classes have full support on the space of all possible tie structures in the sense that for any choice of a continuous distribution for X, any choice of a tie structure, and any sufficiently small value of c, that tie structure has a positive probability of occurring in a given set. However, models in each class can exhibit certain undesirable behavior when the parameters c and m are modified. For TIC models, adding an additional unit to the set can increase the number of ties among the units already in the set. This happens when the new unit fills a gap between two units already in the set. For example, with $m = 2$ and $c = 1$, units with $X_1 = 0.1$ and $X_2 = 1.6$ would not be declared tied. However, if we add a third unit with $X_3 = 0.9$, then a three-way tie will be declared since $|X_1 - X_3| < 1.0$ and $|X_2 - X_3| < 1.0$. For DPS models, the undesirable behavior comes when the value of c is increased. One would expect increasing c to lead to more ties, but this is not necessarily the case. To illustrate this behavior of DPS models, we first obtain a result about the probabilities of obtaining certain types of ties.

In this result, which is given below as Theorem 1, the values $\{p_l : l \in \mathbb{Z}\}$ are the break points that determine how the perceived size X is discretized. For example, if the discretization is done by setting $X^\star = \text{round}(X)$, then $\{p_l : l \in \mathbb{Z}\}$ is the set $\{\ldots, -1.5, -0.5, 0.5, 1.5, \ldots\}$.

Theorem 1 *Consider a DPS model in which X has continuous cumulative distribution function F. Suppose that two units are declared tied whenever both units fall into the same subinterval of the partition determined by the sequence of values $\{p_l : l \in \mathbb{Z}\}$. If $1 \leq i \leq j \leq m$, then the probability of a tie involving only ranks i to j is*

$$T_{ij} \equiv \frac{m!}{(i-1)!(j-i+1)!(m-j)!} \times \sum_{l=-\infty}^{\infty} F(p_l)^{i-1} \left(F(p_{l+1}) - F(p_l)\right)^{j-i+1} \left(1 - F(p_{l+1})\right)^{m-j}, \tag{1}$$

and the probability of a tie involving ranks i and j (possibly together with other ranks) is

$$T^\star_{ij} \equiv \sum_{l=1}^{i} \sum_{r=j}^{m} T_{lr}. \tag{2}$$

Proof In order to have a tie involving only ranks i to j, the set of m units must include $j - i + 1$ units with perceived sizes in some interval (p_l, p_{l+1}), $i - 1$ units with perceived sizes less than p_l, and $m - j$ units with perceived sizes greater than p_{l+1}. By a standard multinomial probability calculation, this probability is

$$\frac{m!}{(i-1)!(j-i+1)!(m-j)!} F(p_l)^{i-1} \left(F(p_{l+1}) - F(p_l)\right)^{j-i+1} \left(1 - F(p_{l+1})\right)^{m-j}$$

when l is given. Summing over all possible values for l leads to Eq. (1).

There is a tie involving ranks i and j (possibly together with other ranks) when there is a tie for a set of ranks that includes the ranks i to j. Such a tie occurs when there is a tie involving only ranks l to r for values l and r satisfying $l \leq i$ and $j \leq r$. Thus, Eq. (2) follows. □

To illustrate the behavior of DPS and TIC models when c changes, we took $m = 4$, and we supposed that the perceived sizes X follow gamma distributions with variance 1. We chose gamma distributions so that we could obtain distributions with fixed variance 1, but different shapes. Since the variance for the Gamma(α, β) distribution is $\alpha\beta^2$, we obtain variance 1 by setting $\beta = 1/\sqrt{\alpha}$. We considered $\alpha = 1/4, 1, 4$, and 16. Figure 1 shows the probability density functions for the four gamma distributions that we considered, and Figs. 2 and 3 show the probability of a tie for ranks 1 and 2 ($T^\star_{12}$), ranks 2 and 3 ($T^\star_{23}$), and ranks 3 and 4 ($T^\star_{34}$) for various choices of $c > 0$ under DPS and TIC models.

The values in Fig. 2 were obtained using Eq. (2), and the values in Fig. 3 were obtained through a simulation study in which 10,000 sets were generated for each choice of α and c. In the DPS model, the discrete perceived sizes were computed as $X^\star = \text{Round}(X/c + 1/2)$. In Fig. 2, which used the DPS model, we see that for the cases with $\alpha = 1/4$ and $\alpha = 1$, $T^\star_{12} > T^\star_{23} > T^\star_{34}$ for each choice of c. This ordering is expected since the perceived size X has a

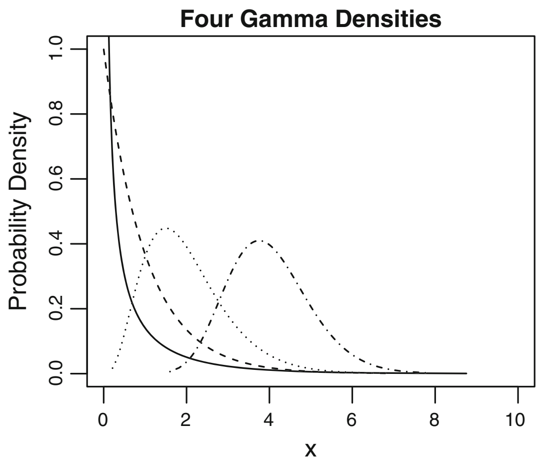

Fig. 1 The four gamma densities used as the perceived sizes in Figs. 2 and 3. The *curves* are for $\alpha = 0.25$ (*solid*), $\alpha = 1$ (*dashed*), $\alpha = 4$ (*dotted*), and $\alpha = 16$ (*dotdash*)

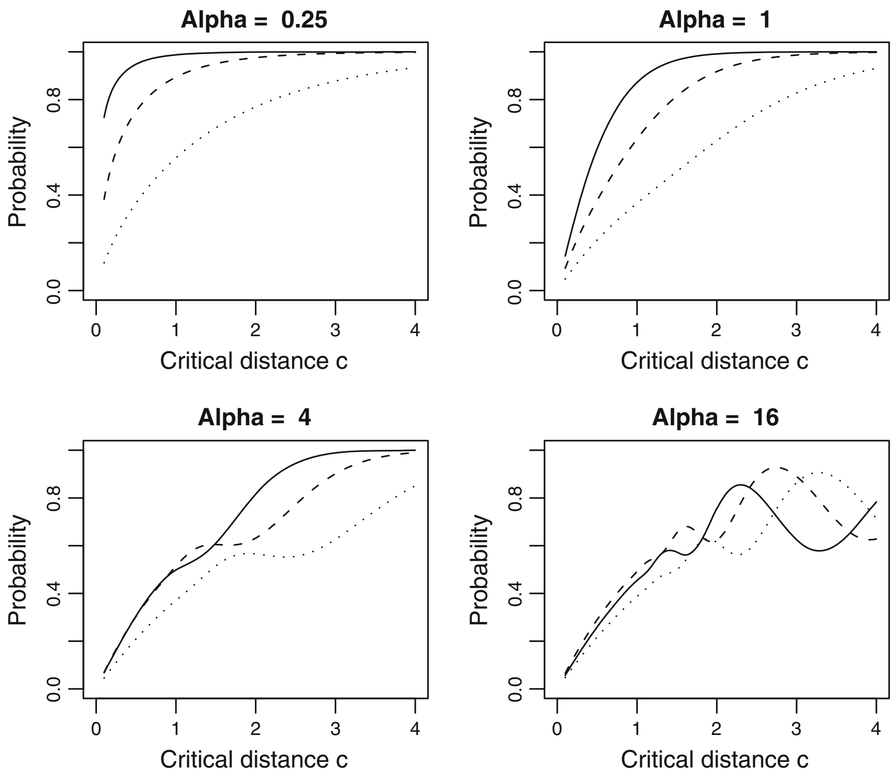

Fig. 2 Probabilities of different types of ties as a function of the constant c when ties are assigned according to a DPS model with $X \sim \text{Gamma}(\alpha, 1/\sqrt{\alpha})$. In each case, the discrete perceived sizes were obtained as $\text{Round}(X/c + 1/2)$. The *curves* are for ties involving ranks 1 and 2 (*solid*), ties involving ranks 2 and 3 (*dashed*), and ties involving ranks 3 and 4 (*dotted*) when $m = 4$

right-skewed distribution. However, for $\alpha = 4$ and $\alpha = 16$, this ordering does not always hold. Indeed, we see that for $\alpha = 4$ and $\alpha = 16$, increasing c can reduce the probability of a particular type of tie. This behavior occurs because values X_1 and X_2 that satisfy $\text{Round}(X_1/c + 1/2) = \text{Round}(X_2/c + 1/2)$ for one value of c may not satisfy the same equality if c is increased. We see from Fig. 3 that under a TIC model, increasing c always increases the chance of a tie involving any two ranks. We also see, by comparing Figs. 2 and 3, that for fixed c and α, each type of tie is more likely under the TIC model than under the DPS model.

As α increases, the $\text{Gamma}(\alpha, 1/\sqrt{\alpha})$ distribution becomes increasingly symmetric, and this increased symmetry is reflected in how the curves are ordered in the different plots of Fig. 3. When α is small so that the distribution is highly right-skewed, ties involving ranks 1 and 2 are most likely and ties involving ranks 3 and 4 are least likely. However, once α increases to 16, ties involving ranks 2 and 3 are more likely than ties involving ranks 1 and 2. In the limit as α goes to infinity and the distribution becomes symmetric, ties involving ranks 2 and 3 will be most likely, and ties involving ranks 1 and 2 will be exactly as likely as ties involving ranks 3 and 4.

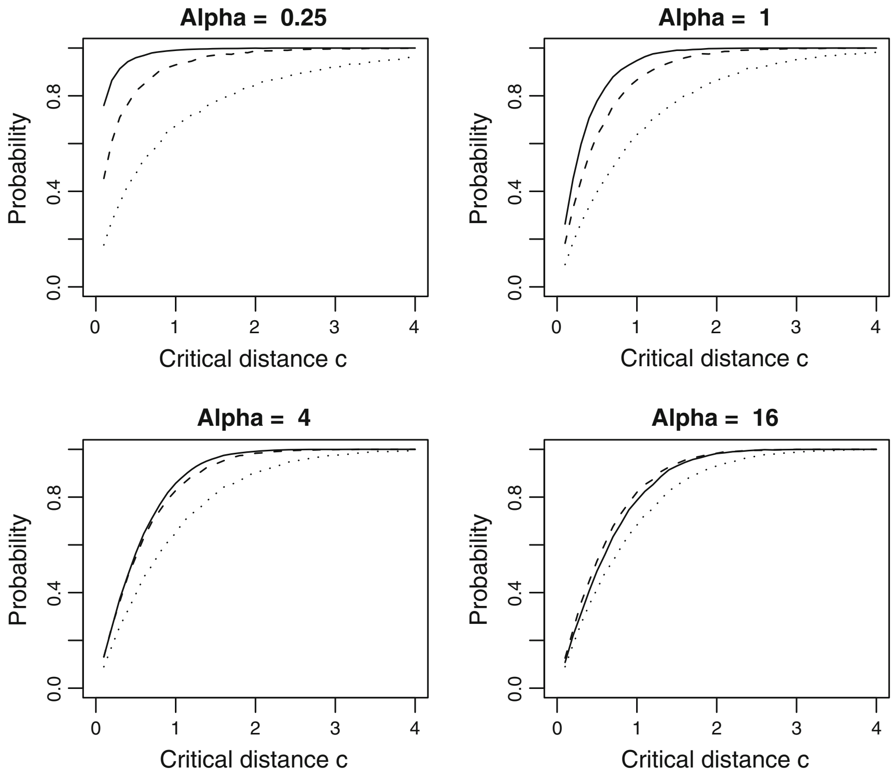

Fig. 3 Simulated probabilities of different types of ties as a function of the constant c when ties are assigned according to a TIC model with $X \sim \text{Gamma}(\alpha, 1/\sqrt{\alpha})$. The *curves* are for ties involving ranks 1 and 2 (*solid*), ties involving ranks 2 and 3 (*dashed*), and ties involving ranks 3 and 4 (*dotted*) when $m = 4$

4 Comparison of estimators

To compare the estimators from Sect. 2, we did a simulation study. We took the perceived size X and the precise measured value Y to have a joint bivariate normal distribution with correlation ρ, and we took $X \sim N(0, 1)$. Thus, we used a Dell and Clutter (1972) type model to relate the perceived size and the true precise value. We then generated the tie structure by using a DPS model in which the discrete perceived size was $X^\star = \text{Round}(X/c)$, where $c > 0$ is a model parameter. For a particular choice of ρ, c, the set size m, and the sample size vector $(n_1, \ldots, n_m)$, we compared the estimators in terms of mean squared error (MSE). Specifically, we compared the performance of each estimator to that of $\hat{\mu}_1$ by computing a simulated relative efficiency. For example, the efficiency of $\hat{\mu}_2$ relative to $\hat{\mu}_1$ was computed as

$$\text{Relative Efficiency} = \frac{\widehat{MSE}(\hat{\mu}_1)}{\widehat{MSE}(\hat{\mu}_2)},$$

where $\widehat{MSE}(\hat{\mu}_1)$ and $\widehat{MSE}(\hat{\mu}_2)$ are the simulated mean squared errors for the two estimators. With the relative efficiency defined in this way, values above 1 indicate an advantage for $\hat{\mu}_2$, and values below 1 indicate an advantage for $\hat{\mu}_1$.

We considered different choices for m, ρ, and c, and we also considered different choices of the sample size vector $(n_1, \ldots, n_m)$. In particular, we looked both at balanced cases where n_i is the same value n for all i and at unbalanced cases where the sample size varies from one rank to another. In balanced cases, only $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\mu}_4$ are distinct since the other three estimators are the same as $\hat{\mu}_1$. However, in the unbalanced case, all six estimators are different. Some representative relative efficiency results are given in Figs. 4 and 5 and Tables 3 and 4. Each value given in one of the figures or tables was based on 10,000 simulated samples.

Figures 4 and 5 show relative efficiencies for $\hat{\mu}_2$ and $\hat{\mu}_4$ in balanced cases. Figure 4 gives results for the case where $m = 4$, $n = 2$, $\rho \in [-1, 1]$, and $c \in \{1/2, 1, 2, 4\}$, and Fig. 5 gives results for the case where $m = 4$, $n = 4$, $\rho \in [-1, 1]$, and $c \in \{1/2, 1, 2, 4\}$. In the two figures, solid dots give simulated relative efficiencies for $\hat{\mu}_2$, and open dots give simulated relative efficiencies for $\hat{\mu}_4$. We see from the figures that $\hat{\mu}_2$ tends to be significantly more efficient than $\hat{\mu}_1$ when ρ is near 1 or -1, but slightly less efficient than $\hat{\mu}_1$ when ρ is near 0. The pattern is somewhat different for $\hat{\mu}_4$, which is significantly more efficient than $\hat{\mu}_1$ when ρ is close to 1, slightly less efficient than $\hat{\mu}_1$ when ρ is close to 0, and comparable in efficiency to $\hat{\mu}_1$ when ρ is close to -1. The performance of $\hat{\mu}_4$ is not symmetric in ρ since $\hat{\mu}_4$ involves isotonizing the weighted sample means $\bar{Y}'_{[1]}, \ldots, \bar{Y}'_{[m]}$ that were defined in Sect. 2.

The efficiency of $\hat{\mu}_2$ and $\hat{\mu}_4$ relative to $\hat{\mu}_1$ in the balanced case is affected not just by ρ, but also by c, which determines how likely it is that ties are declared. Both when

Table 3 Simulated efficiencies of $\hat{\mu}_4$ relative to $\hat{\mu}_1$ for different choices of m, n, ρ, and c

m	n	$\rho = 0.7$				$\rho = 1.0$			
		$c = 1/2$	$c = 1$	$c = 2$	$c = 4$	$c = 1/2$	$c = 1$	$c = 2$	$c = 4$
2	2	1.02	1.03	1.06	1.04	1.05	1.10	1.15	1.08
	6	1.02	1.04	1.07	1.02	1.06	1.13	1.16	1.04
	10	1.02	1.05	1.07	1.02	1.07	1.13	1.18	1.03
3	2	1.02	1.04	1.08	1.04	1.09	1.17	1.23	1.09
	6	1.03	1.06	1.09	1.03	1.09	1.20	1.27	1.05
	10	1.03	1.06	1.09	1.02	1.10	1.22	1.27	1.05
4	2	1.02	1.06	1.09	1.05	1.10	1.23	1.28	1.10
	6	1.03	1.06	1.10	1.03	1.12	1.27	1.33	1.06
	10	1.03	1.08	1.12	1.03	1.13	1.27	1.33	1.06
5	2	1.01	1.07	1.09	1.04	1.12	1.27	1.31	1.11
	6	1.02	1.07	1.11	1.03	1.13	1.30	1.36	1.06
	10	1.03	1.07	1.11	1.03	1.14	1.31	1.39	1.06

The rankings were done using a DPS model where (X, Y) is bivariate normal with correlation ρ, $X \sim N(0, 1)$, and the perceived size is $X^\star = \text{Round}(X/c)$. Each simulated efficiency is based on 10,000 samples

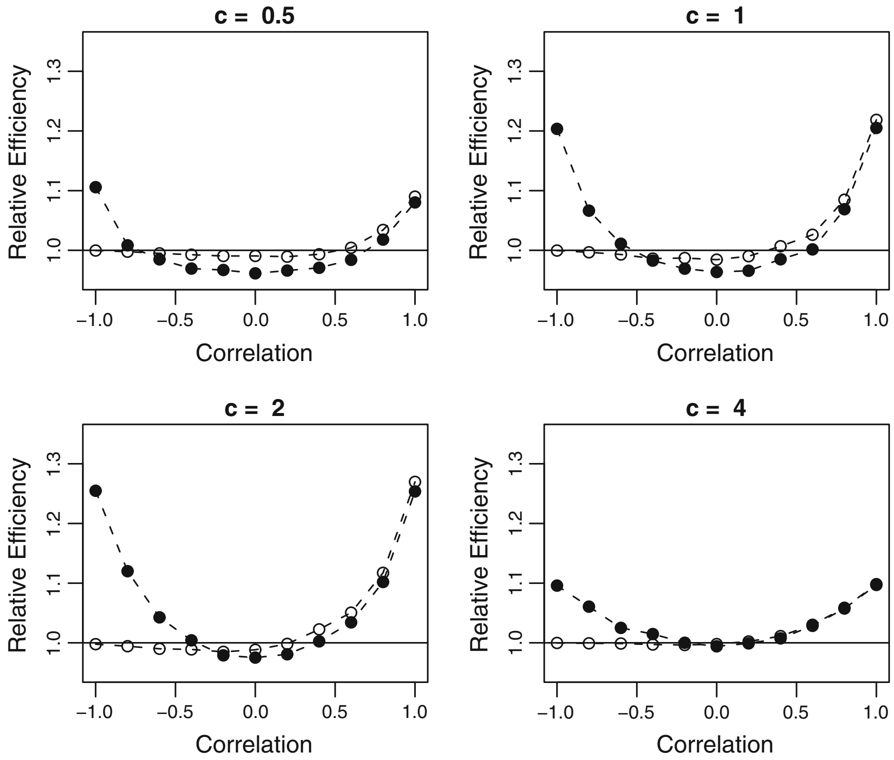

Fig. 4 Simulated efficiencies for $\hat{\mu}_2$ and $\hat{\mu}_4$ relative to $\hat{\mu}_1$ when $m = 4$ and $n = 2$. The rankings were done using a DPS model where (X, Y) is bivariate normal with correlation ρ, $X \sim N(0, 1)$, and the perceived size is $X^{\star} = \text{Round}(X/c)$. We drew 10,000 samples for each value of ρ and c. Here *solid dots* are for $\hat{\mu}_2$ and *open dots* for $\hat{\mu}_4$

c is small (1/2) and when c is large (4), the advantage of $\hat{\mu}_2$ or $\hat{\mu}_4$ over $\hat{\mu}_1$ is relatively small (no more than 10%), but when c is of moderate size (1 or 2), the advantage of $\hat{\mu}_2$ and $\hat{\mu}_4$ over $\hat{\mu}_1$ is large (as high as 20 or 30%). Comparing Figs. 4 and 5, which differ only in the total sample size used, we also see that there is a tendency (that holds more generally) for the best-case relative efficiencies of $\hat{\mu}_2$ and $\hat{\mu}_4$ to increase with increasing sample sizes.

In comparing the relative efficiencies for $\hat{\mu}_2$ and $\hat{\mu}_4$, we see that there is a slight advantage for $\hat{\mu}_4$ over $\hat{\mu}_2$ in those cases ($\rho > 0$) where the true in-stratum population means are ordered from smallest to largest. There is a substantial advantage for $\hat{\mu}_2$ over $\hat{\mu}_4$ when rankings are backwards or nearly backwards so that ρ is close to -1. However, since it seems unlikely that one would obtain highly accurate backwards rankings without becoming aware that the rankings were reversed, $\hat{\mu}_4$ is probably the better of the two choices.

Table 3 extends the results in Figs. 4 and 5 by expanding the list of choices for m and the common sample size n. We see from Table 3, which shows efficiencies only for $\hat{\mu}_4$, that the good performance of $\hat{\mu}_4$ holds not just for particular set sizes

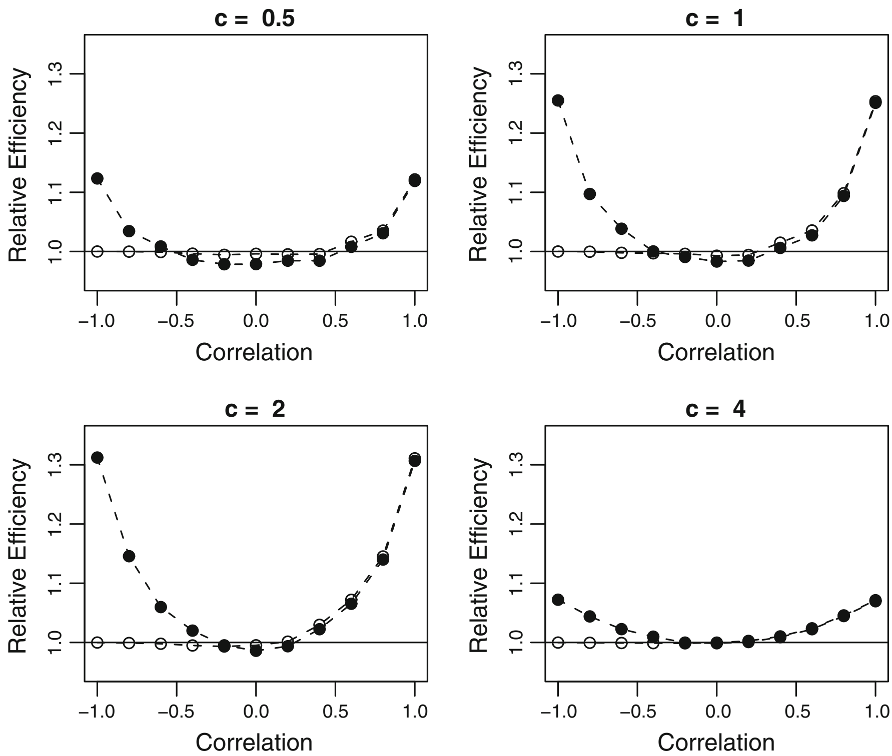

Fig. 5 Simulated efficiencies for $\hat{\mu}_2$ and $\hat{\mu}_4$ relative to $\hat{\mu}_1$ when $m = 4$ and $n = 4$. The rankings were done using a DPS model where (X, Y) is bivariate normal with correlation ρ, $X \sim N(0, 1)$, and the perceived size is $X^\star = \text{Round}(X/c)$. We drew 10,000 samples for each value of ρ and c. Here *solid dots* are for $\hat{\mu}_2$ and *open dots* for $\hat{\mu}_4$

and sample sizes, but across a range of set sizes and in-stratum sample sizes n. The relative efficiencies for $\hat{\mu}_4$ tend to increase as m, n, and ρ increase, and they tend to be higher for moderate values of c like 1 and 2 than for very low or very high values of c like $1/2$ and 4. Thus, the relative efficiency of $\hat{\mu}_4$ is highest when ties are common, but not extremely common. In all cases shown in Table 3, $\hat{\mu}_4$ outperformed $\hat{\mu}_1$.

Table 4 gives some results for a case with unbalanced sample sizes. Here $m = 4$ and $(n_1, \ldots, n_4) = (4, 2, 2, 4)$. In this case, all six estimators are different, and the table shows efficiencies for $\hat{\mu}_2$ through $\hat{\mu}_6$ relative to $\hat{\mu}_1$. In implementing the Rao-Blackwellized estimators $\hat{\mu}_5$ and $\hat{\mu}_6$, we did not systematically consider all ways in which the ties could have been broken. Instead, we repeatedly broke the ties at random until we obtained 10 different cases in which the ranks assigned to the precise measured values matched the desired sample sizes $(4, 2, 2, 4)$. We then averaged the estimates obtained in those 10 cases. This procedure yielded approximate versions of $\hat{\mu}_5$ and $\hat{\mu}_6$ that require less computation, but are only slightly less efficient than the estimators we would obtain by considering all possible ways that the ties might have been broken.

We see from Table 4 that, as is known theoretically, $\hat{\mu}_5$ is always at least as efficient as $\hat{\mu}_1$, and $\hat{\mu}_6$ is always at least as efficient as $\hat{\mu}_3$. However, the gains in efficiency

Table 4 Simulated efficiencies relative to $\hat{\mu}_1$ when $m = 4$ and $(n_1, \ldots, n_4) = (4, 2, 2, 4)$

ρ	c	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\mu}_5$	$\hat{\mu}_6$
0	1/2	1.02	1.11	1.11	1.01	1.11
	1	1.05	1.10	1.10	1.03	1.10
	2	1.11	1.12	1.13	1.09	1.14
	4	1.13	1.11	1.13	1.11	1.13
0.6	1/2	1.06	1.06	1.10	1.01	1.06
	1	1.11	1.06	1.14	1.03	1.08
	2	1.15	1.05	1.16	1.06	1.08
	4	1.11	1.07	1.11	1.09	1.08
0.8	1/2	1.07	1.03	1.09	1.01	1.04
	1	1.17	1.03	1.19	1.03	1.06
	2	1.23	1.04	1.23	1.07	1.07
	4	1.11	1.05	1.11	1.09	1.06
1	1/2	1.16	1.01	1.17	1.01	1.02
	1	1.34	1.01	1.35	1.03	1.03
	2	1.39	1.02	1.39	1.07	1.06
	4	1.10	1.02	1.10	1.08	1.04

The rankings were done using a DPS model where (X, Y) is bivariate normal with correlation ρ, $X \sim N(0, 1)$, and the perceived size is $X^\star = \text{Round}(X/c)$. We drew 10,000 samples for each value of ρ and c

from using these Rao-Blackwellized estimators are small in comparison to the gains in efficiency offered by $\hat{\mu}_2$ and $\hat{\mu}_4$. This, combined with the fact that the Rao-Blackwellized estimators require more computation than the other estimators, suggests that $\hat{\mu}_5$ and $\hat{\mu}_6$ are not good choices. When we compare $\hat{\mu}_2$, $\hat{\mu}_3$, and $\hat{\mu}_4$, we see that for all of the cases shown in Table 4, $\hat{\mu}_4$ was at least as efficient as $\hat{\mu}_2$ and $\hat{\mu}_3$, with the relative advantage for $\hat{\mu}_4$ over $\hat{\mu}_3$ going as high as 20% or 30%. Thus, it appears that $\hat{\mu}_4$ is the best choice for an estimator. Comparing the relative efficiencies in Table 4 to those shown in Table 3 and Figs. 4 and 5, it appears that the gains in efficiency from using $\hat{\mu}_4$ rather than $\hat{\mu}_1$ with unbalanced samples are comparable in magnitude to the gains in the case of balanced sampling.

5 Is it sometimes better to declare a tie?

In Sect. 4, we compared the estimators from Sect. 2 in the case where each estimator used the same ranking information. In this section, we consider the position of a ranker who is deciding whether to declare a tie or not. We present simulation results that suggest that when the rankings are very good, it can be advantageous to declare ties not just when the ranker is actually unsure about how the units rank, but also when the ranker is sure about the ranking, but believes that the units are close.

As in Sect. 4, we assumed that (X, Y) is bivariate normal with correlation ρ, and we obtained discrete perceived sizes using $X \sim N(0, 1)$ and $X^\star = \text{Round}(X/c)$. We compared the performance of $\hat{\mu}_4$ under two strategies: obtaining a full ranking

Table 5 Simulated efficiencies for $\hat{\mu}_4$ with ties relative to $\hat{\mu}_4$ without ties for different choice of ρ and c when $m = 4$ and $(n_1, \ldots, n_4) = (3, 3, 3, 3)$

ρ	$c = 0.2$	$c = 0.4$	$c = 0.6$	$c = 0.8$	$c = 1.0$	$c = 1.2$	$c = 1.4$
0.8	1.00	1.01	1.02	1.01	1.00	0.99	0.98
0.9	1.01	1.02	1.03	1.03	1.02	1.00	0.97
1.0	1.03	1.05	1.06	1.05	1.04	1.00	0.96

The without-ties method uses rankings based on X, while the with-ties method uses ties assigned according to a DPS model with (X, Y) bivariate normal, $X \sim N(0, 1)$, $\mathrm{Corr}(X, Y) = \rho$, and $X^\star = \mathrm{Round}(X/c)$

Table 6 Tie probabilities for the DPS model used in creating Table 5

Parameter	$c = 0.2$	$c = 0.4$	$c = 0.6$	$c = 0.8$	$c = 1.0$	$c = 1.2$	$c = 1.4$
$T_{12}^\star$	0.0991	0.1903	0.2732	0.3478	0.4137	0.4690	0.5126
$T_{23}^\star$	0.1253	0.2358	0.3322	0.4153	0.4871	0.5523	0.6154
$T_{13}^\star$	0.0071	0.0270	0.0577	0.0966	0.1413	0.1904	0.2440
$T_{14}^\star$	0.0003	0.0020	0.0066	0.0150	0.0284	0.0481	0.0760

$T_{ij}^\star$ is the probability of a tie involving ranks i and j, possibly together with other ranks. By symmetry, $T_{34}^\star = T_{12}^\star$ and $T_{24}^\star = T_{13}^\star$. These tie probabilities were computed using Theorem 1

using X and obtaining a ranking with ties using $X^\star$. As in Sect. 4, we compared the two strategies by computing simulated relative efficiencies. Here we computed the relative efficiencies in such a way that values above 1 indicate an advantage for the ranking-with-ties strategy. Table 5 shows simulated relative efficiency results for $\rho = 0.8, 0.9, 1.0$ and c between 0.2 and 1.4. Table 6 shows the probabilities of various types of ties for different values of c. We see from Table 6 that as c increases, ties become more common.

We see from Table 5 that when c is either close to 0 or larger than 1.0, there is little advantage in declaring ties when the full ranking according to X is available. However, for intermediate values of c, an efficiency advantage of as much as 6% can be obtained by declaring some ties even when the full ranking according to X is available. Thus, in cases where the rankings are believed to be very good, a ranker may gain an advantage by declaring ties between units that are believed to be close, though not actually tied. Specifically, Table 5 suggests that in scenarios similar to those considered in this section, declaring a tie when values are within about half a population standard deviation of each other will maximize efficiency. The source of the advantage seen in Table 5 is presumably that when two or more units are very close, the average of the ranks reflects each unit's position in the set better than does the actual rank.

6 Clarifying an earlier result

The theoretical results obtained by Ozturk (2011) rely on the claim, stated as his Lemma 1, that having a particular subset of units be tied tells us nothing about the average value of units in that subset. We show here that this claim does not hold in general. Thus, there is a need for alternate sampling methods like the one proposed in this paper.

Suppose that, for fixed set size m, $\mu_{[i]}$ is the expected value of the ith judgment order statistic when ties are broken at random. For integers i and j satisfying $1 \leq i \leq j \leq m$, define $\mu_{[i:j]}$ to be the expected value of a unit chosen at random from a subset of units that are tied for ranks i to j (and no others). That is, $\mu_{[i:j]}$ is conditional on there being a tie for ranks i to j and no others. Lemma 1 of Ozturk (2011) then states that $\mu_{[i:j]}$ is simply the average of $\mu_{[i]}, \ldots, \mu_{[j]}$.

To show that Lemma 1 of Ozturk (2011) does not hold in general, we did a simulation study in which we used $m = 5$ and a DPS model for ties. We took the precise measured values Y to be equal to the perceived sizes $X \sim N(0, 1)$, and we took the discrete perceived sizes to be given by Round(X). For each choice of i and j with $1 \leq i \leq j \leq 5$, we estimated two quantities: the average $\frac{1}{j-i+1} \sum_{k=i}^{j} \mu_{[k]}$ when ties are broken at random and the value $\mu_{[i:j]}$ defined earlier. In estimating the first quantity, every single simulated set is relevant, but in estimating $\mu_{[i:j]}$, only sets in which there is a tie for ranks i to j (and no others) are relevant. Thus, in what follows, we refer to one expectation as the unconditional expected value and the second ($\mu_{[i:j]}$) as the conditional expected value. Table 7 gives the results, which were obtained by simulating 1,000,000 sets of size 5. In the course of the 1,000,000 runs, each type of tie occurred at least 10,000 times.

If Lemma 1 of Ozturk (2011) were true in full generality, then the upper and lower halves of Table 7 would coincide up to simulation error. However, the two halves of the table are quite different. For example, the entry −0.77 in the unconditional part indicates that the average of $\mu_{[1]}$ and $\mu_{[2]}$ is −0.77, but the entry −0.88 in the conditional part indicates that $\mu_{[1:2]} = -0.88$. Thus, having a tie for ranks 1 and 2 tells us something about the corresponding precise measured values.

Table 7 Simulated unconditional and conditional expected values for a unit selected at random from those with ranks (after breaking ties at random) i to j when $m = 5$

Type	i	j				
		1	2	3	4	5
Unconditional	1	−1.07	−0.77	−0.51	−0.27	0.00
	2		−0.46	−0.23	0.00	0.27
	3			0.00	0.23	0.51
	4				0.45	0.76
	5					1.07
Conditional	1	−1.38	−0.88	−0.53	−0.22	0.00
	2		−0.59	−0.25	0.00	0.22
	3			0.00	0.25	0.53
	4				0.59	0.88
	5					1.38

The conditional expected values are conditional on there being a tie for ranks i to j (and no others). We used a DPS model with $X \sim N(0, 1)$ and $X^{\star} = \text{Round}(X)$. The simulation study used 1,000,000 simulated sets

The differences between the upper and lower halves of Table 7 are actually not surprising. For example, consider the case with $i = j = 1$. The unconditional expected value is $\mu_{[1]} = -1.07$, and the conditional expected value is $\mu_{[1:1]} = -1.38$. Thus, the average value for a first judgment order statistic is -1.07, but the average value for a first judgment order statistic that is not tied for any other ranks is -1.38. This makes perfect sense since a unit that is judged to be smallest in its set will surely tend to be smaller than a unit that is judged to be only tied for smallest in its set. Other differences between the top and bottom halves of Table 7 can be explained with similar logic.

7 Discussion

We have considered a modification of RSS in which the ranker is allowed to declare ties. Unlike the recommended scheme of Ozturk (2011), this modification does not involve forced ties. Instead, the ranker may declare as many or as few ties as desired. We have developed nonparametric mean estimators that are appropriate for use with data obtained from this sampling scheme, and we have compared the estimators using models that allow both for ties and for imperfect rankings. Our comparisons suggest that the isotonized estimator $\hat{\mu}_4$, defined in Sect. 2, is the best estimator both for balanced sampling and for unbalanced sampling.

Two of the estimators that we have compared are the standard unbiased RSS mean estimator $\hat{\mu}_1$, which ignores the tie structure, and $\hat{\mu}_5$, a Rao-Blackwellized version of $\hat{\mu}_1$ that coincides with $\hat{\mu}_1$ under balanced sampling, but outperforms $\hat{\mu}_1$ under unbalanced sampling. The existence of $\hat{\mu}_5$ shows that $\hat{\mu}_1$ is inadmissible under squared error loss when tie information is available. Thus, we have shown here that in unbalanced RSS, ignoring tie information leads to inadmissible mean estimators. Frey (2011) obtained a result with a similar flavor by showing that $\hat{\mu}_1$ is inadmissible in unbalanced RSS when the rankings are done using a covariate.

Acknowledgments The author thanks the Associate Editor and the reviewers for helpful comments that have improved the paper.

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 H, Stasny EA, Wolfe DA (2005) Ranked set sampling for efficient estimation of a population proportion. Stat Med 24:3319–3329

Dell TR, Clutter JL (1972) Ranked-set sampling theory with order statistics background. Biometrics 28:545–555

Fligner MA, MacEachern SN (2006) Nonparametric two-sample methods for ranked-set sample data. J Am Stat Assoc 101:1107–1118

Frey J (2007) New imperfect rankings models for ranked set sampling. J Stat Plan Inference 137:1433–1445

Frey J (2011) A note on ranked-set sampling using a covariate. J Stat Plan Inference 141:809–816

Halls LK, Dell TR (1966) Trial of ranked-set sampling for forage yields. For Sci 12:22–26

Kvam PH (2003) Ranked set sampling based on binary water quality data with covariates. J Agricult Biol Environ Stat 8:271–279

MacEachern SN, Stasny EA, Wolfe DA (2004) Judgement post-stratification with imprecise rankings. Biometrics 60:207–215

McIntyre GA (1952) A method for unbiased selective sampling, using ranked sets. Aust J Agricult Res 3:385–390
McIntyre GA (2005) A method for unbiased selective sampling, using ranked sets. Am Stat 59:230–232 (originally appeared in Aust J Agricult Res 3:385–390)
Ozturk O (2007) Statistical inference under a stochastic ordering constraint in ranked set sampling. J Nonparametr Stat 19:131–144
Ozturk O (2011) Sampling from partially rank-ordered sets. Environ Ecol Stat 18:757–779
Patil GP (1995) Editorial: ranked set sampling. Environ Ecol Stat 2:271–285
Robertson T, Wright FT, Dykstra RL (1988) Order restricted statistical inference. Wiley, New York
Wang X, Lim J, Stokes SL (2008) A nonparametric mean estimator for judgment post-stratified data. Biometrics 64:355–363

Author Biography

Jesse Frey received his PhD in statistics from The Ohio State University in 2005. Since then, he has taught statistics at Villanova University. His research interests include ranked-set sampling, nonparametric statistics, and statistical computing.