Ann Inst Stat Math (2012) 64:439–456
DOI 10.1007/s10463-011-0326-9

Constrained nonparametric estimation of the mean and the CDF using ranked-set sampling with a covariate

Jesse Frey

Received: 1 February 2010 / Revised: 18 June 2010 / Published online: 2 March 2011

Abstract Ranked-set sampling (RSS) and judgment post-stratification (JPS) are related schemes in which more efficient statistical inference is obtained by creating a stratification based on ranking information. The rankings may be completely subjective, or they may be based on values of a covariate. Recent work has shown that regardless of how the rankings are done, the in-stratum cumulative distribution functions (CDFs) must satisfy certain constraints, and we show here that if the rankings are done according to a covariate, then tighter constraints must hold. We also show that under a mild stochastic ordering assumption, still tighter constraints must hold. Taking advantage of these new constraints leads to improved small-sample estimates of the in-stratum CDFs in all RSS and JPS settings. For JPS, the new constraints also lead to improved estimates of the overall CDF and the population mean.

Keywords Concomitant · Convexity · Judgment post-stratification · Maximum likelihood estimation · Stratified sampling · Woodruff confidence intervals

1 Introduction

Ranked-set sampling (RSS), proposed by McIntyre (1952, 2005), is a sampling scheme appropriate for use when it is inexpensive to rank or approximately rank small sets of units without actually measuring them. The rankings may be completely subjective, or they may be based on values of an easily available covariate (Stokes 1977), and they need not be completely accurate. The ranking information is used to guide the selection

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

of the units to be measured, and the resulting sample tends to be more informative than a simple random sample of the same size.

To implement balanced RSS, one first specifies a set size m and a number of cycles n. One then selects $N \equiv nm$ independent simple random samples (sets) of size m. The units in each of these N sets are ranked from smallest to largest without making any actual measurements. From each of the first n sets, the unit with rank 1 is selected for measurement. From each of the next n sets, the unit with rank 2 is selected for measurement, and so on. The ranked-set sample then consists of N independent values, with n values from each of the m possible in-set ranks. If the rankings are perfect, then these N values are independent order statistics. More generally, they are independent *judgment* order statistics. For some statistical problems, it is helpful to allow the number of measured values to vary from one rank to another. In this case, one may use unbalanced RSS. One simply specifies a set size m and a vector $\mathbf{n} \equiv (n_1, \ldots, n_m)$ such that n_i gives the number of units with rank i to be selected for measurement. The ranked-set sample then consists of $N \equiv \sum_{i=1}^{m} n_i$ independent judgment order statistics.

Takahasi and Wakimoto (1968) showed that for a fixed number of measured values N, balanced RSS is at least as efficient as simple random sampling (SRS) for estimating the population mean. Similar gains in efficiency are available for other statistical problems, including parametric point estimation (Stokes 1995), nonparametric estimation of the cumulative distribution function (CDF) (Stokes and Sager 1988), testing for a difference in location between two distributions (Fligner and MacEachern 2006), and nonparametric estimation of the population variance (MacEachern et al. 2002). However, in order to realize these gains, one must use the RSS sampling approach. Judgment post-stratification (JPS), proposed by MacEachern et al. (2004), is an alternate method that exploits the same ranking information used in RSS, but that starts with a simple random sample. As a result, researchers who use JPS retain the option of using SRS-based methods if needed.

To implement JPS, one first specifies a total sample size N and a set size m. One then draws a simple random sample of size N. The N units are each measured, and some ranking information is also collected. For each of the N units in the simple random sample, one obtains an additional $m - 1$ independent units, yielding a set of m units. The units in this set are ranked from smallest to largest, and the rank of the unit that was measured is recorded. The full data set then consists of the N measured values and the rank associated with each measured value. As in unbalanced RSS, the number of measured units with each particular rank may vary from one rank to another, and some ranks may not appear at all. In what follows, we let $\mathbf{n} \equiv (n_1, \ldots, n_m)$ be a vector giving the number of measured units with each rank. In RSS, $\mathbf{n}$ is specified ahead of time, but in JPS it is random. In fact, $\mathbf{n}$ follows a multinomial distribution with mass parameter N and probability vector $(1/m, \ldots, 1/m)$. JPS tends to be somewhat less efficient than balanced RSS. However, when the ranking information is good or the sample is large, it outperforms SRS for estimating population means and CDFs. JPS also offers more flexibility than RSS in that rankers may be permitted to declare ties (MacEachern et al. 2004).

When implementing RSS and JPS, appropriate precautions must be taken to avoid bias. For example, when implementing RSS using subjective rankings, the ranker must not know which ranked unit is to be selected from a particular set. Similarly, when implementing JPS using subjective rankings, the ranker must not know which unit in a particular set is the one on which a measurement has been made.

The standard nonparametric RSS and JPS mean estimators are obtained using exactly the procedure used in stratified random sampling or in standard post-stratification (see Lohr 1999), with the rank (1 to m) associated with each measured value being used as the stratification variable. However, recent work has shown that better estimators can be obtained by taking into account additional structure that need not exist for ordinary stratified sampling and post-stratification. One way to obtain better estimators is to take advantage of the fact that, whether the rankings are perfect or not, it is reasonable to assume that the in-stratum distributions are stochastically ordered in some way. Ozturk (2007) assumed that the in-stratum CDFs are stochastically ordered, and Wang et al. (2008) assumed that the in-stratum means are a nondecreasing function of the rank (1 to m). There are obstacles to formally testing such assumptions in a nonparametric context (Davidson and Duclos 2006), but the assumptions may be informally assessed via graphical techniques. For example, one might make side-by-side boxplots of the stratum-by-stratum samples.

Another way to obtain improved estimators involves looking at constraints that must be satisfied by the in-stratum CDFs at each particular point. Frey and Ozturk (2010) showed that the in-stratum CDFs for strata that arise from ranking information can be no more extreme, in a certain sense, than the CDFs for true order statistics from the overall distribution. In this paper, we show that if the rankings are done according to a covariate, then constraints tighter than those obtained by Frey and Ozturk (2010) must hold. We also show that under a mild stochastic ordering assumption, still tighter constraints must hold. By taking advantage of these new constraints, we obtain improved small-sample estimates of the in-stratum CDFs in all RSS and JPS settings, and we also obtain improved estimates of the overall CDF and the population mean in the JPS setting.

The paper is structured as follows. We derive the constraints in Sect. 2, and we use them to estimate the CDF in Sect. 3. We use the constraints to estimate the population mean in Sect. 4, and we use them to create Woodruff-type confidence intervals for population quantiles in Sect. 5. We give our conclusions in Sect. 6.

2 The constraints

Suppose that Y is the variable of interest and that X is the covariate used for ranking purposes. In this section, we derive constraints that must hold for the in-stratum CDFs of Y when the strata arise from ranking units according to X. In deriving the first set of constraints, we assume only that the distribution of X is continuous. Later, we obtain stronger constraints by assuming that the distribution of Y given $X = x$ is stochastically increasing in x. We obtain certain convexity results that hold for any set size m, and we then use these results to obtain specific computational results that

apply in the $m = 3$ case. We focus on small set sizes since these tend to be the most important cases in practice (Takahasi and Wakimoto 1968).

We first note that as long as X is a continuous random variable, we may assume that X is standard uniform. To see this, note that if X has continuous CDF F_X, then $U \equiv F_X(X)$ is standard uniform. Moreover, since F_X is continuous and nondecreasing, ranking units according to U is equivalent to ranking units according to X. Thus, in what follows, we assume without loss of generality that the covariate has a standard uniform distribution, and we emphasize this point by writing U for the covariate.

Let $S_u(y)$ be the conditional CDF for Y given that $U = u$. The unconditional CDF for Y is then given by $F(y) = P(Y \leq y) = \int_{u=0}^{1} S_u(y)\, \mathrm{d}u$. Let $Y_{[1]}, \ldots, Y_{[m]}$ be the measured Y values associated with the order statistics $U_{(1)}, \ldots, U_{(m)}$ in a particular set. By well-known properties of uniform order statistics, $U_{(i)}$ is distributed $Beta(i, m + 1 - i)$. Thus, the CDF for $Y_{[i]}$ is given by

$$F_{[i]}(y) = P(Y_{[i]} \leq y) = \int_{u=0}^{1} S_u(y) \cdot \frac{m!}{(i-1)!(m-i)!} u^{i-1}(1-u)^{m-i}\, \mathrm{d}u.$$

Fix the value y, and define $p_i \equiv F_{[i]}(y)$ for $i = 1, \ldots, m$. Define $K^c \subset [0, 1]^m$ to be the space of all possible values for the vector $(p_1, \ldots, p_m)$. Our first result is that K^c is convex.

Theorem 1 *The space K^c of all possible values for $(p_1, \ldots, p_m)$ is convex.*

Proof Let $\mathbf{p_1} \equiv (p_{11}, \ldots, p_{m1})$ and $\mathbf{p_2} \equiv (p_{12}, \ldots, p_{m2})$ be arbitrary points from K^c, and let $\lambda \in [0, 1]$ be an arbitrary constant. We need to show that $\lambda\mathbf{p_1} + (1-\lambda)\mathbf{p_2} \in K^c$. Let $S_u^{(1)}(y)$ and $S_u^{(2)}(y)$ be conditional CDFs for Y given $U = u$ that lead to the vectors $\mathbf{p_1}$ and $\mathbf{p_2}$, respectively. The function $T_u(y) \equiv \lambda S_u^{(1)}(y) + (1-\lambda) S_u^{(2)}(y)$ is then also a possible conditional CDF for Y given $U = u$, and the corresponding vector of CDF values is precisely $\lambda\mathbf{p_1} + (1-\lambda)\mathbf{p_2}$. For example, if $T_u(y)$ is the conditional CDF for Y given $U = u$, then $P(Y_{[1]} \leq y)$ satisfies

$$\begin{aligned}
P\left(Y_{[1]} \leq y\right) &= \int_{u=0}^{1} T_u(y) \cdot m(1-u)^{m-1}\, \mathrm{d}u \\
&= \int_{u=0}^{1} \left(\lambda S_u^{(1)}(y) + (1-\lambda) S_u^{(2)}(y)\right) \cdot m(1-u)^{m-1}\, \mathrm{d}u \\
&= \lambda \int_{u=0}^{1} S_u^{(1)}(y) \cdot m(1-u)^{m-1}\, \mathrm{d}u \\
&\quad + (1-\lambda) \int_{u=0}^{1} S_u^{(2)}(y) \cdot m(1-u)^{m-1}\, \mathrm{d}u \\
&= \lambda p_{11} + (1-\lambda) p_{12}.
\end{aligned}$$

Thus, $\lambda\mathbf{p_1} + (1-\lambda)\mathbf{p_2} \in K^c$, and the theorem is proved. □

Now suppose that $m = 3$, and consider the convex set K_r^c of possible values for the vector (p_1, p_2) when the overall CDF value $\bar{p} \equiv \frac{1}{3}(p_1 + p_2 + p_3)$ is fixed at r.

The following result shows that the boundary points of K_r^c arise when the conditional CDF $S_u(y)$ has a very specific form.

Theorem 2 *The boundary points of the set K_r^c are the points (p_1, p_2) achieved when $S_u(y)$ is an indicator function $I(u \in J)$, where J is either an interval $[a, a+r]$ for $a \in [0, 1-r]$ or a union $[a, 1] \bigcup [0, a+r-1]$ for $a \in (1-r, 1)$.*

Proof The set K_r^c is a two-dimensional convex set. Thus, it is the intersection of all half-planes that contain it, and a particular point in K_r^c is a boundary point for K_r^c if and only if it maximizes a function $a_1 p_1 + a_2 p_2$ over the set K_r^c for some real constants a_1 and a_2 that are not both zero. Writing $a_1 p_1 + a_2 p_2$ in terms of $S_u(y)$, we have that

$$\begin{aligned} a_1 p_1 + a_2 p_2 &= a_1 \int_{u=0}^{1} S_u(y) \cdot 3(1-u)^2 \, \mathrm{d}u + a_2 \int_{u=0}^{1} S_u(y) \cdot 6u(1-u) \, \mathrm{d}u \\ &= \int_{u=0}^{1} S_u(y) \left(a_1 \cdot 3(1-u)^2 + a_2 \cdot 6u(1-u) \right) \mathrm{d}u. \end{aligned} \tag{1}$$

Expanding the expression in parentheses inside (1), we obtain the function $Q(u) \equiv u^2(3a_1 - 6a_2) + u(-6a_1 + 6a_2) + 3a_1$, which is either quadratic in u or, if $a_1 = 2a_2$, linear in u. By choosing a_1 and a_2 appropriately, we can force $Q(\cdot)$ to be quadratic, to open in the direction of our choice (up or down), and to be symmetric about any real number that we choose.

For fixed a_1 and a_2, consider choosing $S_u(y)$ so that $a_1 p_1 + a_2 p_2$ is maximized. In order for $S_u(y)$ to yield an overall CDF value of r, we must have $\int_{u=0}^{1} S_u(y) \, \mathrm{d}u = r$. By (1), $a_1 p_1 + a_2 p_2 = \int_{u=0}^{1} S_u(y) Q(u) \, \mathrm{d}u$. Thus, to maximize $a_1 p_1 + a_2 p_2$, we need $S_u(y)$ to be big when $Q(\cdot)$ is big and small when $Q(\cdot)$ is small. Since $S_u(y)$ is a probability, $S_u(y)$ can be no larger than 1 and no smaller than 0. Thus, we maximize $a_1 p_1 + a_2 p_2$ by having $S_u(y)$ be 1 on some subset of $[0, 1]$ and 0 otherwise. Since $Q(\cdot)$ is either quadratic or linear, there will exist a set J so that (i) J is either an interval $[a, a+r]$ for $a \in [0, 1-r]$ or a union $[a, 1] \bigcup [0, a+r-1]$ for $a \in (1-r, 1)$, (ii) the total length of J is r, and (iii) the infimum of $Q(\cdot)$ on J is equal to the supremum of $Q(\cdot)$ on $[0, 1] \bigcap J^c$. It then follows that if $S_u(y) = I(u \in J)$, then $a_1 p_1 + a_2 p_2$ is maximized. Since we can force $Q(\cdot)$ to be quadratic and to be centered at any real number, it is also clear that for any set J of the form described in the theorem, there exist a_1 and a_2 such that $a_1 p_1 + a_2 p_2$ is maximized when $S_u(y) = I(u \in J)$. This proves the theorem. □

To plot the boundary for the set K_r^c when r is fixed, one simply lets the value a from Theorem 2 take on values in the interval $[0, 1]$ and plots the resulting points (p_1, p_2). The value $a = 0$ corresponds to perfect rankings, the value $a = 1 - r$ corresponds to perfectly wrong (backwards) rankings, and the limit as a approaches 1 from below is perfect rankings again. Thus, as a goes from 0 to 1, the corresponding points (p_1, p_2) trace out a path from perfect rankings to perfectly wrong rankings and back again.

The constraints determined by Theorems 1 and 2 are tighter than other constraints in the literature. Consider the constraints on p_2 when the overall CDF value is

$r = 0.5$ and $m = 3$. Frey and Ozturk (2010) showed that for any fixed value of r, $B(r; 1, 3) \le p_2 \le B(r; 3, 1)$, where $B(\cdot; \alpha, \beta)$ is the CDF for a $Beta(\alpha, \beta)$ distribution. With $r = 0.5$, this leads to constraints $1/8 \le p_2 \le 7/8$. Suppose now that the rankings are done using a covariate. The choice for $S_u(y)$ that maximizes p_2 when $r = 0.5$ is $S_u(y) = I(u \in [0.25, 0.75])$, and the choice for $S_u(y)$ that minimizes p_2 is $S_u(y) = I(u \in [0, 0.25] \bigcup [0.75, 1])$. Computing the corresponding values for p_2, we obtain constraints $5/16 \le p_2 \le 11/16$, which are tighter than the earlier bounds.

Suppose that we now add a stochastic ordering assumption. Ozturk (2007) assumed that the in-stratum CDFs are stochastically ordered so that $F_{[1]}(y) \ge F_{[2]}(y) \ge \cdots \ge F_{[m]}(y)$ at every point y, and Wang et al. (2008) assumed that the in-stratum means are a nondecreasing function of the rank. Fligner and MacEachern (2006) suggested that having the conditional distribution of the covariate U given $Y = y$ be stochastically increasing in y is an essential property for any reasonable model for rankings. Here, we also use conditional distributions, but it is more convenient to work with the conditional distribution of Y given $U = u$. Thus, we make the following stochastic ordering assumption.

Assumption 1 The distributions $S_u(y)$ are stochastically nondecreasing in u. That is, if $u_1 < u_2$, then $S_{u_1}(y) \ge S_{u_2}(y)$ for all y.

Fix the value y, and let $p_1, \ldots, p_m$ be defined as before. Define $K^s \subset [0, 1]^m$ to be the space of all possible values for the vector $(p_1, \ldots, p_m)$ under Assumption 1 about $S_u(y)$. The following theorem shows that, like the larger set K^c, K^s is convex.

Theorem 3 *The space K^s is convex.*

Proof We rely on the fact that a convex combination of nonincreasing functions is also nonincreasing. Let $\mathbf{p_1} \equiv (p_{11}, \ldots, p_{m1})$ and $\mathbf{p_2} \equiv (p_{12}, \ldots, p_{m2})$ be arbitrary points from K^s, and let $\lambda \in [0, 1]$ be an arbitrary constant. We need to show that $\lambda \mathbf{p_1} + (1 - \lambda)\mathbf{p_2} \in K^s$. Let $S_u^{(1)}(y)$ and $S_u^{(2)}(y)$ be conditional CDFs for Y given $U = u$ that are nonincreasing functions of u and that lead to the vectors $\mathbf{p_1}$ and $\mathbf{p_2}$, respectively. It then follows that the function $T_u(y) \equiv \lambda S_u^{(1)}(y) + (1 - \lambda) S_u^{(2)}(y)$ is also nonincreasing and a possible conditional CDF for Y given $U = u$. Thus, by calculations similar to those used in proving Theorem 1, $\lambda \mathbf{p_1} + (1 - \lambda)\mathbf{p_2} \in K^s$, and the theorem is proved. □

Now suppose that $m = 3$, and consider the convex set K_r^s of possible values for (p_1, p_2) when Assumption 1 holds and the overall CDF value $\bar{p} \equiv \frac{1}{3}(p_1 + p_2 + p_3)$ is fixed at r. The following result shows that the boundary points for K_r^s can only come when $S_u(y)$ has a very specific form.

Theorem 4 *The boundary points of the set K_r^s are the points (p_1, p_2) achieved when $S_u(y)$ is either* (i) *constant on the interval* $[0, 1]$ *or* (ii) *a nonincreasing step function that takes on only two values, one of which is 0 or 1.*

Proof As in the proof of Theorem 2, a point can be a boundary point for K_r^s only if it maximizes a function $a_1 p_1 + a_2 p_2$ over the set K_r^s for some real constants a_1 and

a_2 that are not both zero. Thus, the problem of determining the possible boundary points reduces to that of finding all functions $S_u(y)$ that are nonincreasing, satisfy $0 \le S_u(y) \le 1$ and $\int_{u=0}^{1} S_u(y)\, du = r$, and maximize $\int_{u=0}^{1} S_u(y) Q(u)\, du$ for some linear or quadratic function $Q(\cdot)$.

We first make two observations about the shape of such functions $S_u(y)$ on intervals where $Q(\cdot)$ is either increasing or decreasing. Suppose that $Q(\cdot)$ is *increasing* over the interval (a, b). In this case, we make the function $a_1 p_1 + a_2 p_2$ large by choosing $S_u(y)$ on the interval (a, b) to be large for large u and small for small u. Since $S_u(y)$ must be nonincreasing, the best we can do is to have $S_u(y)$ be constant on (a, b). Suppose that $Q(\cdot)$ is *decreasing* over the interval (a, b). In this case, we make the function $a_1 p_1 + a_2 p_2$ large by choosing $S_u(y)$ on the interval (a, b) to be large for small u and small for large u. Thus, the best form for $S_u(y)$ on the interval (a, b) is a step function that takes on only two values. For all values of u up to a certain point $c \in (a, b)$, $S_u(y)$ should take on the maximum value $S_a(y)$. For $u > c$, $S_u(y)$ should take on the value $S_b(y)$.

Since $Q(\cdot)$ is linear or quadratic, it must be either (i) increasing over the entire interval [0, 1], (ii) decreasing over the entire interval [0, 1], (iii) increasing and then decreasing on the interval [0, 1], or (iv) decreasing and then increasing on the interval [0, 1]. If $Q(\cdot)$ is increasing over the entire interval [0, 1], then we maximize $a_1 p_1 + a_2 p_2$ by choosing $S_u(y)$ to be constant on [0, 1], and if $Q(\cdot)$ is decreasing over the entire interval [0, 1], then any maximizing choice of $S_u(y)$ must be a step function that is 1 up to a certain point, and then 0 beyond that point. These two results follow from our observations in the previous paragraph. Now consider the other two cases. Suppose that $Q(\cdot)$ is increasing and then decreasing on the interval [0, 1]. Any maximizing $S_u(y)$ must be constant over the interval where $Q(\cdot)$ is increasing. Suppose that $S_u(y) = v$ on this interval. It then follows that on the interval where $Q(\cdot)$ is decreasing, $S_u(y)$ must be equal to v up to a certain point and then 0 beyond that point. Thus, $S_u(y)$ must be a step function of the form described in the theorem. Suppose that $Q(\cdot)$ is decreasing and then increasing on [0, 1]. Any maximizing $S_u(y)$ must be constant over the interval where $Q(\cdot)$ is increasing. Suppose that $S_u(y) = v$ on this interval. It then follows that on the interval where $Q(\cdot)$ is decreasing, $S_u(y)$ must be equal to 1 up to a certain point and then v beyond that point. Thus, the theorem is proved.

□

The constraints determined by Theorems 3 and 4 are more stringent than those determined by Theorems 1 and 2. To illustrate this, we created Fig. 1. Figure 1 shows the spaces K_r^c and K_r^s for $r = 0.2, 0.4, 0.6$, and 0.8 and $m = 3$. It also, for comparison purposes, shows the space K_r of possible values for (p_1, p_2) under the $\bar{p} = r$ assumption and the constraints obtained by Frey and Ozturk (2010). Recall that $p_i \equiv F_{[i]}(y)$ so that $r = \bar{p} = F(y)$ is the overall CDF value. In each plot of Fig. 1, the outer-most set of bounds shows the Frey and Ozturk (2010) constraints, and the inner-most set of bounds is for rankings done according to a covariate that satisfies Assumption 1. We see from the figure that all of the sets are convex and that the new sets of bounds are substantially tighter than the bounds obtained by Frey and Ozturk (2010). The dashed reference line in each plot is the line $p_1 = p_2$.

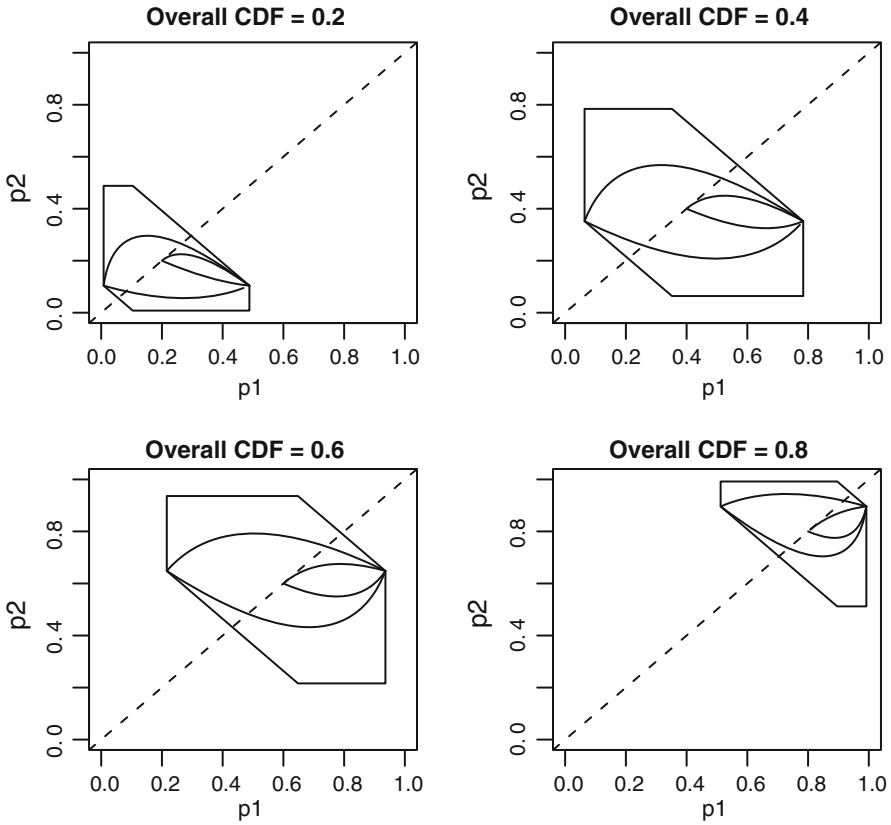

Fig. 1 Slices from the space of possible vectors (p_1, p_2, p_3) when the set size is $m = 3$. Each plot shows the potential values for $p_1 \equiv F_{[1]}(y)$ and $p_2 \equiv F_{[2]}(y)$ when the overall CDF value $\bar{p} = F(y)$ is fixed at some particular value r. The value $p_3 \equiv F_{[3]}(y)$ is determined by the constraint that $r = \frac{1}{3}(p_1 + p_2 + p_3)$. In each plot, the regions are, from largest to smallest, the Frey and Ozturk (2010) region K_r, the new region K_r^c, and the new region K_r^s. The *dashed reference line* is the line $p_1 = p_2$

3 Constrained estimation of the CDF

The standard nonparametric estimate of the population CDF F under either RSS or JPS is

$$\hat{F}(y) = \frac{1}{m} \sum_{i=1}^{m} \hat{F}_{[i]}(y),$$

where $\hat{F}_{[i]}(y)$ is the empirical distribution function (EDF) for the measured values that were given rank i. If any of the strata are not represented, as may occur with JPS, then $\hat{F}(y)$ is the average of $\hat{F}_{[i]}(y)$ over all the strata with nonzero sample sizes. The estimator $\hat{F}$ is unbiased for F, but the vector $(\hat{F}_{[1]}(y), \ldots, \hat{F}_{[m]}(y))$ need not be a possible value for $(F_{[1]}(y), \ldots, F_{[m]}(y))$, the true vector of in-stratum CDF values

at the point y. Thus, it makes sense to replace $(\hat{F}_{[1]}(y), \ldots, \hat{F}_{[m]}(y))$ with a vector of estimates that is a possible value for $(F_{[1]}(y), \ldots, F_{[m]}(y))$. In this section, we describe a general strategy for doing this in the case where $m = 3$. This strategy works with either of the two sets of constraints developed in Sect. 2.

To obtain estimates, we follow the approach used by Frey and Ozturk (2010). Suppose that we want to estimate the vector $(F_{[1]}(y), \ldots, F_{[m]}(y))$ for some fixed real number y. Let $\mathbf{n} \equiv (n_1, \ldots, n_m)$ be the vector of in-stratum sample sizes, and let $\mathbf{x} \equiv (x_1, \ldots, x_m)$ be the vector of counts of the number of values in each stratum that are less than or equal to y. If $\mathbf{p} \equiv (p_1, \ldots, p_m)$ is a candidate vector of in-stratum CDF values, then the likelihood function is given by

$$L(\mathbf{p}|\mathbf{n}, \mathbf{x}) = \prod_{i=1}^{m} \binom{n_i}{x_i} p_i^{x_i}(1 - p_i)^{n_i - x_i} \quad \text{for } 0 \le p_i \le 1, \quad i = 1, \ldots, m.$$

Thus, the log likelihood function can be written as

$$l(\mathbf{p}|\mathbf{n}, \mathbf{x}) = c + \sum_{i=1}^{m} \{x_i \log(p_i) + (n_i - x_i) \log(1 - p_i)\}$$
$$\text{for } 0 < p_i < 1, \quad i = 1, \ldots, m,$$

where c does not depend on $\mathbf{p}$. If we maximize L over the entire space $[0, 1]^m$, then we obtain the standard EDF-based vector of in-stratum estimates $(\hat{F}_{[1]}(y), \ldots, \hat{F}_{[m]}(y))$. However, if we maximize L only over the set K^c or the set K^s, then we obtain estimates that satisfy the corresponding constraints. Thus, the covariate-based estimate is defined by

$$\left(\hat{F}^c_{[1]}(y), \ldots, \hat{F}^c_{[m]}(y)\right) \equiv \arg\max_{\mathbf{p} \in K^c} l(\mathbf{p}|\mathbf{n}, \mathbf{x}),$$

and the corresponding estimate for the overall CDF is

$$\hat{F}^c(y) \equiv \frac{1}{m} \sum_{i=1}^{m} \hat{F}^c_{[i]}(y).$$

Similarly, the stochastic ordering covariate-based estimate is defined by

$$\left(\hat{F}^s_{[1]}(y), \ldots, \hat{F}^s_{[m]}(y)\right) \equiv \arg\max_{\mathbf{p} \in K^s} l(\mathbf{p}|\mathbf{n}, \mathbf{x}),$$

with the corresponding estimate for the overall CDF being

$$\hat{F}^s(y) \equiv \frac{1}{m} \sum_{i=1}^{m} \hat{F}^s_{[i]}(y).$$

Since the sets K^c and K^s are each convex and the log likelihood $l(\mathbf{p}|\mathbf{n}, \mathbf{x})$ is concave, finding either the estimates $(\hat{F}^c_{[1]}(y), \ldots, \hat{F}^c_{[m]}(y))$ or the estimates $(\hat{F}^s_{[1]}(y), \ldots, \hat{F}^s_{[m]}(y))$ requires maximizing a concave function over a convex set. This maximization is complicated, however, by the fact that we aren't able to write down nice general expressions for the boundaries of K^c and K^s. However, using Theorems 2 and 4, we are able to find these estimates when $m = 3$. The procedure we use involves two types of maximization, as described below.

First, given a particular value r for the overall CDF $\bar{p} \equiv \frac{1}{3}(p_1 + p_2 + p_3)$, we maximize $l(\mathbf{p}|\mathbf{n}, \mathbf{x})$ over either K^c_r or K^s_r by (i) finding the maximum of $l(\mathbf{p}|\mathbf{n}, \mathbf{x})$ under the constraint $\bar{p} = r$ and (ii) noting whether that maximum is in the set (K^c_r or K^s_r). If this maximum is in the set, then it is also the restricted maximum. Otherwise, the restricted maximum must occur on the boundary of the set. Using either Theorem 2 (for K^c_r) or Theorem 4 (for K^s_r), we obtain a 120-point list of boundary points for the set. We then evaluate $l(\mathbf{p}|\mathbf{n}, \mathbf{x})$ at each boundary point and approximate the restricted maximum using the point among the 120 that maximizes $l(\mathbf{p}|\mathbf{n}, \mathbf{x})$. Second, we search for the overall CDF value r for which the maximum of the likelihood over K^c_r or K^s_r is largest. Procedure 2 in the appendix of Frey and Ozturk (2010) is one method for doing this. The key to this procedure is that the functions $r \to \arg\max_{\mathbf{p} \in K^c_r} l(\mathbf{p}|\mathbf{n}, \mathbf{x})$ and $r \to \arg\max_{\mathbf{p} \in K^s_r} l(\mathbf{p}|\mathbf{n}, \mathbf{x})$ are concave functions defined on the interval [0, 1].

Example 1 To illustrate the estimation ideas just described, we consider the case where $\mathbf{n} = (10, 10, 10)$ and $\mathbf{x} = (8, 7, 1)$. Suppose that we wish to estimate (p_1, p_2, p_3) while assuming that $\bar{p} = 0.5$. Figure 2 shows the bounds on (p_1, p_2) in this setting. The three sets shown are the set K_r determined by the Frey and Ozturk (2010) bounds, the set K^c_r, and the set K^s_r. The solid dots in the plot are the unrestricted estimate, the estimate under the Frey and Ozturk (2010) constraints, the estimate under the ranking-by-covariate constraints, and the estimate under the stochastic ordering constraints. The unrestricted estimate lies outside of all the sets, and this forces the other three estimates to lie on the boundaries of the respective sets. The dashed curves in the figure show some contours of the likelihood function. Computing the estimates shown in Fig.2 is just one part of the overall estimation process. We would also need to determine the value of the overall CDF for which the likelihood is maximized. Completing this second level of maximization leads to the estimates $(0.8, 0.7, 0.1)$ (unrestricted EDF-based estimate), $(0.773, 0.668, 0.149)$ (Frey and Ozturk 2010), $(0.809, 0.602, 0.187)$ (covariate-based), and $(0.828, 0.575, 0.193)$ (stochastic ordering). The corresponding estimates of the overall CDF are 0.533, 0.530, 0.533, and 0.532, respectively.

To compare the performance of the new CDF estimators to that of other estimators in the literature, we computed pointwise mean squared errors (MSEs) for each estimator under different types of rankings, different average in-stratum sample sizes, and different values for the overall CDF. We compared the performance of the estimators both under JPS and under RSS. Several models for imperfect rankings are available in the literature. Dell and Clutter (1972) developed a model in which units in the same set are ranked according to a perceived size, where the perceived size is the sum of the true size and an independent error term. Bohn and Wolfe (1994) developed a model in

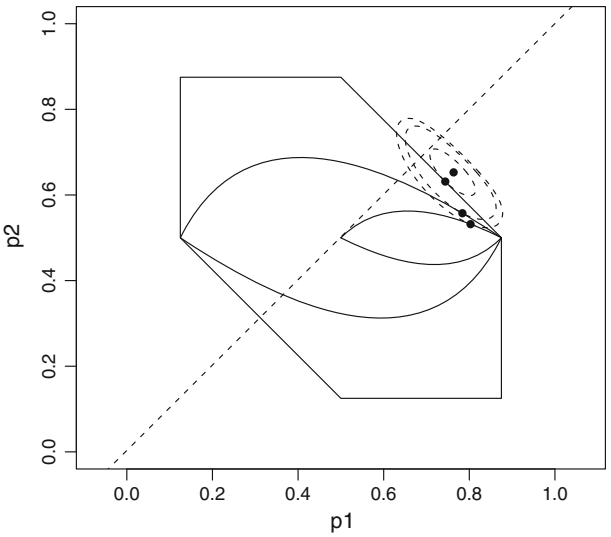

Fig. 2 Maximum likelihood estimates of the vector $(F_{[1]}(y), F_{[2]}(y))$ when the overall CDF is fixed at $r = 0.5$, the sample size vector is $\mathbf{n} = (10, 10, 10)$, and the vector of counts is $\mathbf{x} = (8, 7, 1)$. The *solid lines* show the boundaries of the various sets, and the *dashed curves* show contours of the likelihood function. The *solid dots* show the maximum likelihood estimates under the different sets of constraints. The *dashed reference line* is the line $p_1 = p_2$

which the in-stratum CDFs are written as convex combinations of the distributions for true order statistics from the overall distribution. In our comparisons, we used a related model in which the in-stratum CDFs are convex combinations of the distributions for true order statistics and the overall CDF. Specifically, we assumed that the CDF $F_{[i]}$ for the ith stratum satisfies

$$F_{[i]}(y) = \lambda F_{(i)}(y) + (1 - \lambda) F(y), \tag{2}$$

where $F_{(i)}(y)$ is the CDF for a true ith order statistic from the parent distribution $F(y)$ and $\lambda \in [0, 1]$ is a parameter. Here, $\lambda = 1$ gives perfect rankings, $\lambda = 0$ gives random rankings, and λ values between 0 and 1 give rankings that lie between perfect and random rankings, but satisfy Assumption 1 from Sect. 2.

Some detailed results for one special case are given in Fig. 3, which shows MSEs for JPS under perfect rankings with set size $m = 3$ and total sample size $N = 9$. The four plots show MSEs as a function of the overall CDF value when estimating the overall CDF and the three in-stratum CDFs. The curves in each plot correspond to the standard EDF-based estimator (solid), the Frey and Ozturk (2010) estimator (dashed), the covariate-based estimator (dot-dash), the stochastic ordering estimator (dotted), and the isotonic estimator developed by Ozturk (2007) (long dash). The Ozturk (2007) estimator is obtained by isotonizing the EDF-based estimates for the in-stratum CDFs. We see from the figure that the EDF-based estimator is typically the worst of the estimators, while the stochastic ordering estimator is usually the best. Neither the Frey and Ozturk (2010) estimator nor the covariate-based estimator makes a stochastic

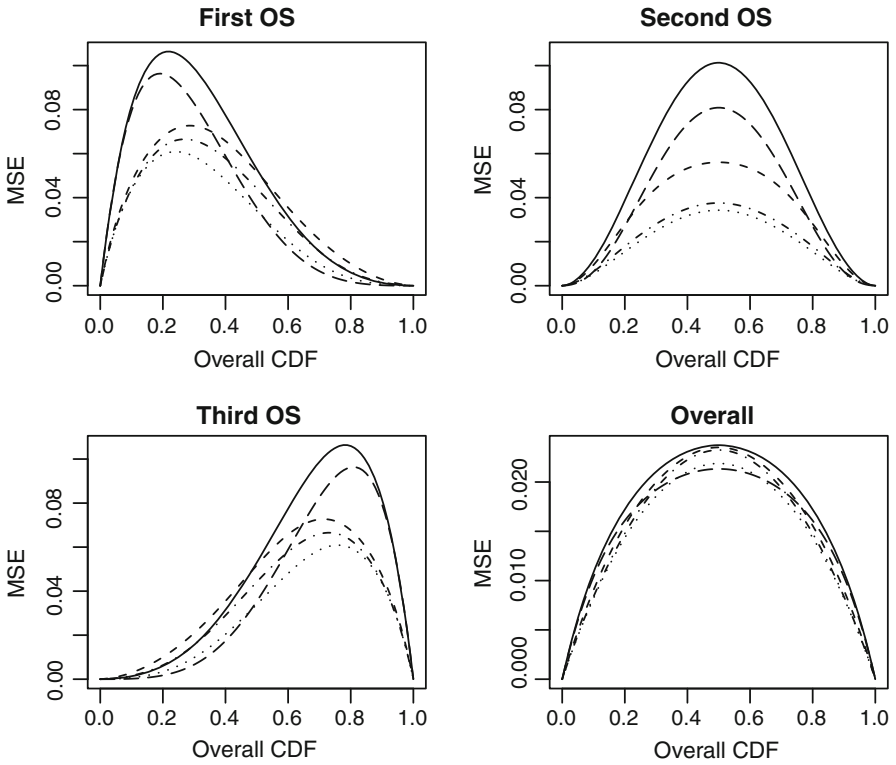

Fig. 3 Mean squared errors as a function of the overall CDF value for JPS under perfect rankings with $N = 9$. The methods are the EDF-based method (*solid*), the Frey and Ozturk (2010) method (*dashed*), the covariate-based method (*dot-dash*), the covariate-based method with stochastic ordering (*dotted*), and the Ozturk (2007) method (*long dash*)

ordering assumption, and the covariate-based estimator is consistently the better of the two.

If we compute the area under curves like those shown in Fig. 3, we obtain integrated MSEs (IMSEs). These IMSEs provide a measure of overall efficiency for an estimator, and we can compare estimators by computing ratios of IMSEs. For example, the efficiency of $\hat{F}^c$ relative to $\hat{F}$ can be computed as

$$\text{Relative efficiency} = \frac{\text{IMSE}(\hat{F})}{\text{IMSE}(\hat{F}^c)}.$$

Table 1 shows calculated efficiencies relative to the EDF-based estimator for several different estimators under JPS and balanced RSS. We considered different types of rankings ($\lambda = 0, 1/3, 2/3, 1$) and different total sample sizes ($N = 3, 9, 15$). The table shows relative efficiencies both for estimation of the overall CDF (F) and for

Table 1 Calculated efficiencies for the new CDF estimators and the Frey–Ozturk (2010) CDF estimator relative to the standard EDF-based estimator for $m = 3$, JPS and balanced RSS, and different types of rankings

Type	Parameter	λ	$N = 3$			$N = 9$			$N = 15$		
			F-O	Cov	Sto	F-O	Cov	Sto	F-O	Cov	Sto
JPS	F	0	1.03	1.03	1.05	1.06	1.12	1.19	1.03	1.08	1.15
		1/3	1.03	1.03	1.05	1.07	1.13	1.19	1.03	1.09	1.14
		2/3	1.02	1.02	1.06	1.07	1.12	1.18	1.04	1.09	1.13
		1	1.01	1.01	1.06	1.07	1.10	1.15	1.05	1.09	1.11
	$F_{[1]}$	0	1.49	1.49	1.64	1.34	1.61	2.14	1.16	1.40	1.94
		1/3	1.44	1.44	1.71	1.37	1.63	2.31	1.20	1.44	2.02
		2/3	1.31	1.32	1.57	1.38	1.61	2.14	1.28	1.50	1.90
		1	1.14	1.15	1.33	1.27	1.44	1.70	1.25	1.45	1.62
	$F_{[2]}$	0	1.49	1.60	1.80	1.34	2.10	3.05	1.16	1.78	3.01
		1/3	1.47	1.58	1.74	1.35	2.12	2.82	1.18	1.82	2.70
		2/3	1.45	1.55	1.68	1.43	2.25	2.70	1.27	2.00	2.54
		1	1.42	1.52	1.62	1.61	2.47	2.77	1.52	2.35	2.66
RSS	F	0	1.00	1.00	1.00	0.98	0.98	0.99	0.99	0.99	1.00
		1/3	1.00	1.00	0.99	0.98	0.99	0.99	0.99	0.99	0.99
		2/3	1.00	1.00	0.99	0.98	0.98	0.98	0.98	0.99	0.99
		1	1.00	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	$F_{[1]}$	0	1.85	1.83	2.22	1.26	1.43	1.84	1.12	1.27	1.72
		1/3	1.82	1.82	2.30	1.29	1.46	1.94	1.15	1.31	1.76
		2/3	1.71	1.73	2.10	1.32	1.48	1.83	1.24	1.38	1.67
		1	1.46	1.52	1.67	1.24	1.39	1.53	1.22	1.35	1.47
	$F_{[2]}$	0	1.85	2.57	2.85	1.26	1.85	2.64	1.11	1.58	2.61
		1/3	1.84	2.52	2.68	1.27	1.87	2.40	1.13	1.62	2.34
		2/3	1.84	2.52	2.62	1.32	1.97	2.30	1.20	1.76	2.20
		1	1.84	2.58	2.71	1.45	2.15	2.39	1.39	2.03	2.30

Here, N is the total sample size

estimation of the in-stratum CDFs. By symmetry, the relative efficiencies for estimating $F_{[3]}$ are the same as the relative efficiencies for estimating $F_{[1]}$.

We see from Table 1 that under JPS, there is a clear ordering of the three estimators. The stochastic ordering estimator $\hat{F}^s$ and its associated in-stratum CDF estimators outperform $\hat{F}^c$, which, in turn, outperforms the Frey and Ozturk (2010) estimator. The efficiency of $\hat{F}^s$ relative to $\hat{F}$ is as high as 1.19, and the corresponding relative efficiency when estimating in-stratum CDFs is as high as 3.05. Under balanced RSS, none of the new estimators outperform the standard estimator when estimating the overall CDF. However, when estimating the in-stratum CDFs, the new estimators offer substantial gains in efficiency. The relative efficiency of the stochastic ordering estimator is not less than 1.47 in any of the scenarios shown, and it goes as high as

2.85. Comparing the top half of the table to the bottom half shows that the advantage of using the new estimators tends to be higher with JPS than with balanced RSS.

4 Constrained estimation of the population mean

The usual unbiased nonparametric estimate of the population mean under RSS or JPS is given by

$$\hat{\mu} = \int_y y \, \mathrm{d}\hat{F}(y).$$

If we replace $\hat{F}$ with $\hat{F}^c$ or $\hat{F}^s$, then we obtain alternate estimators, which we denote $\hat{\mu}_c$ and $\hat{\mu}_s$. These alternate estimators are no longer unbiased, and they are not more efficient than the standard estimator under balanced RSS. However, under unbalanced RSS or JPS, the new estimators tend to be more efficient than the standard estimator. We demonstrate this increased efficiency by presenting results from a simulation study. In addition to the new estimators, we consider the mean estimator proposed by Frey and Ozturk (2010) and the isotonic JPS mean estimator developed by Wang et al. (2008).

In the simulation study, we considered different types of rankings, different parent distributions, and different total sample sizes. Using the same model (2) for imperfect rankings that we used in Sect. 3, we considered parameter values $\lambda = 0, 1/3, 2/3$, and 1. We considered normal, exponential, $Gamma(5, 1)$, uniform, and $Beta(1/2, 1/2)$ parent distributions, and we used total sample sizes 3, 6, 9, 12, and 15. Thus, the average in-stratum sample size ranged from 1 to 5. The set size was fixed at $m = 3$, and 20,000 samples were simulated for each combination of parent distribution, type of rankings, and average sample size. For each combination of factor levels, we estimated the efficiency of each estimator relative to the standard nonparametric mean estimator $\hat{\mu}$ by computing ratios of estimated MSEs. For example, the estimated efficiency of the stochastic ordering mean estimator relative to the standard JPS mean estimator was computed as

$$\text{Relative efficiency} = \frac{\widehat{\text{MSE}}(\hat{\mu})}{\widehat{\text{MSE}}(\hat{\mu}_s)}.$$

Some results of the simulation study for JPS are presented in Table 2. In implementing the Wang et al. (2008) estimator in cases where there were empty cells, we used the following procedure. We first isotonized the means for the nonempty cells. We then replaced the mean for any empty cell with either the isotonized cell mean for the nearest nonempty cell or, if both adjacent cells were nonempty, the average of the isotonized cell means for those two cells. We then averaged the three cell means.

Table 2 shows that $\hat{\mu}_s$ is consistently better than $\hat{\mu}_c$, which, in turn, is consistently better than the Frey and Ozturk (2010) estimator. The Wang et al. (2008) estimator is the best estimator under perfect rankings ($\lambda = 1$) when $N = 3$, but it is outperformed by the stochastic ordering estimator when the rankings are closer to random or when

Table 2 Simulated efficiencies for the Frey and Ozturk (2010) mean estimator, the new mean estimators, and the Wang et al. (2008) mean estimator relative to the standard EDF-based JPS mean estimator for $m = 3$ under different types of rankings and different total sample sizes

Est.	Dist.	λ for $N = 3$				λ for $N = 9$				λ for $N = 15$			
		0	1/3	2/3	1	0	1/3	2/3	1	0	1/3	2/3	1
F-O	Norm.	1.05	1.06	1.04	1.02	1.10	1.10	1.10	1.08	1.05	1.05	1.07	1.06
	Exp.	1.07	1.07	1.07	1.05	1.12	1.13	1.13	1.13	1.08	1.07	1.09	1.11
	Gamma	1.06	1.06	1.05	1.02	1.10	1.11	1.10	1.09	1.06	1.06	1.07	1.08
	Unif.	1.03	1.03	1.02	0.99	1.07	1.07	1.06	1.04	1.03	1.03	1.04	1.02
	Beta	1.02	1.02	1.01	0.98	1.05	1.05	1.05	1.02	1.02	1.02	1.02	1.01
Cov	Norm.	1.05	1.06	1.04	1.02	1.15	1.16	1.16	1.14	1.10	1.11	1.13	1.12
	Exp.	1.07	1.07	1.07	1.05	1.16	1.18	1.19	1.21	1.12	1.12	1.16	1.17
	Gamma	1.06	1.06	1.05	1.03	1.15	1.16	1.17	1.15	1.11	1.11	1.13	1.13
	Unif.	1.03	1.03	1.02	0.99	1.12	1.13	1.12	1.08	1.09	1.09	1.10	1.07
	Beta	1.02	1.02	1.01	0.98	1.11	1.11	1.10	1.05	1.08	1.08	1.08	1.05
Sto	Norm.	1.06	1.10	1.12	1.14	1.18	1.22	1.23	1.23	1.14	1.15	1.16	1.15
	Exp.	1.06	1.11	1.15	1.17	1.18	1.23	1.24	1.28	1.15	1.16	1.18	1.19
	Gamma	1.05	1.10	1.13	1.15	1.17	1.22	1.23	1.25	1.14	1.15	1.17	1.16
	Unif.	1.04	1.07	1.08	1.09	1.17	1.20	1.19	1.17	1.13	1.14	1.14	1.10
	Beta	1.04	1.06	1.06	1.06	1.17	1.19	1.18	1.14	1.13	1.14	1.13	1.08
WLS	Norm.	1.01	1.06	1.11	1.18	1.13	1.14	1.13	1.11	1.11	1.09	1.07	1.03
	Exp.	1.01	1.06	1.12	1.18	1.12	1.13	1.11	1.08	1.12	1.09	1.07	1.02
	Gamma	1.01	1.06	1.11	1.18	1.12	1.14	1.12	1.10	1.12	1.10	1.07	1.03
	Unif.	1.01	1.05	1.08	1.13	1.13	1.14	1.13	1.12	1.11	1.10	1.08	1.04
	Beta	1.01	1.04	1.07	1.10	1.13	1.14	1.13	1.12	1.11	1.10	1.08	1.04

For each combination of distribution, overall sample size, and type of rankings, 20,000 runs were done

the sample size is larger. Overall, there seems to be a clear advantage for the stochastic ordering estimator $\hat{\mu}_s$.

5 An application

To illustrate the value of the in-stratum CDF estimators that we developed in Sect. 3, we apply them to obtain confidence intervals for population quantiles. Woodruff (1952) developed a method for obtaining confidence intervals for population quantiles under stratified random sampling and more complex survey sampling schemes. His basic strategy, which is also described in Lohr (1999), begins with finding, for every possible y, an approximate confidence interval for $F(y)$. The confidence interval for the population quantile $q_p = \inf\{y : F(y) \geq p\}$ then consists of all values y such that the confidence interval for $F(y)$ contains p. If each of the pointwise confidence intervals for $F(y)$ is a nominal $100(1-\alpha)\%$ confidence interval, then the corresponding confidence interval for q_p is also a nominal $100(1-\alpha)\%$ confidence interval.

To apply this idea under balanced RSS, we use a normal approximation. If the vector of in-stratum CDF values at y is $(F_{[1]}(y), \ldots, F_{[m]}(y))$ and the overall CDF value is $F(y)$, then the variance of $\hat{F}(y)$ is $V(\hat{F}(y)) = \frac{1}{m^2} \sum_{i=1}^{m} \left(F_{[i]}(y)(1 - F_{[i]}(y))\right)/n$. To obtain an EDF-based estimate of $V(\hat{F}(y))$, we could replace each $F_{[i]}(y)$ with $\hat{F}_{[i]}(y)$. This, combined with the asymptotic normality of $\hat{F}(y)$ as n increases, leads to a nominal $100(1-\alpha)\%$ confidence interval for $F(y)$ of the form

$$\hat{F}(y) \pm \frac{z_{\alpha/2}}{m} \sqrt{\sum_{i=1}^{m} \left(\hat{F}_{[i]}(y)(1 - \hat{F}_{[i]}(y))\right)/n},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile for the standard normal distribution. However, since the constraints derived in Sect. 2 lead to improved estimates of the in-stratum CDFs, it makes sense to replace each $F_{[i]}(y)$ with $\hat{F}^s_{[i]}(y)$, provided that Assumption 1 is believed to hold. This leads to a nominal $100(1-\alpha)\%$ confidence interval for $F(y)$ of the form

$$\hat{F}(y) \pm \frac{z_{\alpha/2}}{m} \sqrt{\sum_{i=1}^{m} \left(\hat{F}^s_{[i]}(y)(1 - \hat{F}^s_{[i]}(y))\right)/n}. \tag{3}$$

We then proceed as Woodruff (1952) proceeded to obtain the confidence interval for q_p. Specifically, if $(L(y), U(y))$ is the confidence interval for $F(y)$ given in Eq. (3), then our confidence interval for q_p is the interval $(\inf\{y : L(y) < p < U(y)\}, \sup\{y : L(y) < p < U(y)\})$. Note that this procedure does not involve finding a point estimate for q_p. Note also that whichever variance estimate we use, we continue to use $\hat{F}(y)$ as the center of the confidence interval for $F(y)$.

To compare the small-sample performance of these two confidence intervals, we computed the true coverage probabilities of nominal 95% confidence intervals for various population quantiles. We used set size $m = 3$, and we considered numbers of cycles n ranging from 4 to 8. To examine the impact of imperfect rankings, we used the same model (2) that we used in Sects. 3 and 4. What we found is that the confidence intervals based on the stochastic ordering estimator perform better. Figure 4 shows some results for the $n = 4$ case. The figure shows true coverage probabilities for nominal 95% confidence intervals as a function of the parameter λ in the imperfect rankings model. The four plots in the figure correspond to target quantiles $q_{0.2}$, $q_{0.3}$, $q_{0.4}$, and $q_{0.5}$, and the two curves in each plot show the coverage probability for the EDF-based interval (solid) and the stochastic ordering interval (dashed). We see that while both intervals tend to have true coverage probabilities that fall short of the nominal level, the stochastic ordering interval comes closer to the nominal level. When the rankings are perfect ($\lambda = 1$) or nearly perfect, the true level is quite close to the nominal level 95%.

Both Chen (2001) and Zhu and Wang (2005) also used RSS to make inference on population quantiles. However, they considered point estimation rather than interval estimation. Our work here also differs from that of Zhu and Wang (2005) in that we do not assume perfect rankings.

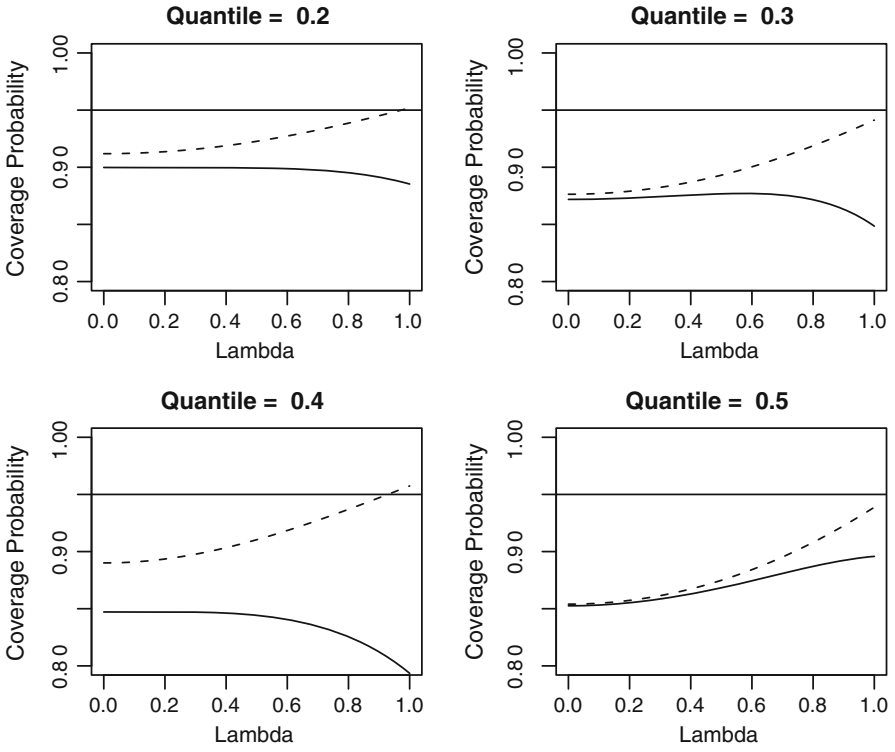

Fig. 4 Exact coverage probabilities as a function of λ for Woodruff-type 95% confidence intervals for different population quantiles. Here, $\mathbf{n} = (4, 4, 4)$. The *solid curves* give coverage probabilities for intervals based on the EDF-based estimator, and the *dashed curves* give coverage probabilities for intervals based on the stochastic ordering estimator

6 Conclusions

Frey and Ozturk (2010) showed that strata arising from ranking information must satisfy additional constraints that need not hold for strata that arise in other ways. We have shown here that when the rankings are done according to a covariate, tighter constraints must hold. We have also shown that if the relationship between the covariate and the variable of interest satisfies a mild stochastic ordering assumption, then even tighter constraints must hold. These constraints can be used to obtain better estimates of in-stratum CDFs under both RSS and JPS, and they lead to better estimates of the overall CDF and the population mean under JPS. Better estimates of the overall CDF and the overall mean are of obvious importance, and better estimates of the in-stratum CDFs are valuable in that they allow one to create statistical procedures that are calibrated to adjust for the effect of imperfect rankings. Ozturk (2007) used in-stratum CDF estimates to create calibrated RSS-based procedures for the two-sample location problem, and Ozturk (2008) used in-stratum CDF estimates to create calibrated RSS-based procedures for creating confidence intervals for population quantiles. In Sect. 5,

we used in-stratum CDF estimates to create Woodruff-type confidence intervals for population quantiles when the sample size is very small.

Since RSS and JPS are often implemented using rankings that are based on a covariate, the main results in this paper are widely applicable. In fact, they may be applied even in cases where no covariate is used explicitly. A sufficient condition for applying the constraints from Sect. 2 is that the rankings are based on a perceived size that behaves like an unrecorded covariate.

Acknowledgments The author thanks the referees for helpful suggestions that have improved the paper.

References

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*, 168–176.

Chen, Z. (2001). The optimal ranked-set sampling scheme for inference on population quantiles. *Statistica Sinica*, *11*, 23–37.

Davidson, R., Duclos, J.-Y. (2006). *Testing for restricted stochastic dominance*, Working paper. McGill University Department of Economics.

Dell, T. R., Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, *28*, 545–555.

Fligner, M. A., MacEachern, S. N. (2006). Nonparametric two-sample methods for ranked-set sample data. *Journal of the American Statistical Association*, *101*, 1107–1118.

Frey, J., Ozturk, O. (2010). Constrained estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics* (to appear).

Lohr, S. (1999). *Sampling: design and analysis*. Pacific Grove: Duxbury.

MacEachern, S. N., Ozturk, O., Wolfe, D. A., Stark, G. A. (2002). A new ranked set sample estimator of variance. *Journal of the Royal Statistical Society, Series B*, *64*, 177–188.

MacEachern, S. N., Stasny, E. A., Wolfe, D. A. (2004). Judgement post-stratification with imprecise rankings. *Biometrics*, *60*, 207–215.

McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*, *3*, 385–390.

McIntyre, G. A. (2005). A method for unbiased selective sampling, using ranked sets. *The American Statistician*, *59*, 230–232 (originally appeared in *Australian Journal of Agricultural Research*, *3*, 385–390).

Ozturk, O. (2007). Statistical inference under a stochastic ordering constraint in ranked set sampling. *Journal of Nonparametric Statistics*, *19*, 131–144.

Ozturk, O. (2008). Statistical inference in the presence of ranking error in ranked set sampling. *Canadian Journal of Statistics*, *36*, 577–594.

Stokes, S. L. (1977). Ranked set sampling with concomitant variables. *Communications in Statistics—Theory and Methods*, *6*(12), 1207–1211.

Stokes, S. L. (1995). Parametric ranked set sampling. *Annals of the Institute of Statistical Mathematics*, *47*, 465–482.

Stokes, S. L., Sager, T. W. (1988). Characterization of a ranked-set sample with application to estimating distribution functions. *Journal of the American Statistical Association*, *83*, 374–381.

Takahasi, K., Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, *20*, 1–31.

Wang, X., Lim, J., Stokes, L. (2008). A nonparametric mean estimator for judgment post-stratified data. *Biometrics*, *64*, 355–363.

Woodruff, R. S. (1952). Confidence intervals for medians and other position measures. *Journal of the American Statistical Association*, *47*, 635–646.

Zhu, M., Wang, Y.-G. (2005). Quantile estimation from ranked set sampling data. *Sankhyā*, *67*, 295–304.