Ann Inst Stat Math (2011) 63:769–789
DOI 10.1007/s10463-009-0255-z

Constrained estimation using judgment post-stratification

Jesse Frey · Omer Ozturk

Received: 11 August 2008 / Revised: 7 May 2009 / Published online: 24 July 2009
© The Institute of Statistical Mathematics, Tokyo 2009

Abstract In ranked-set sampling (RSS) and judgment post-stratification (JPS), more efficient inference is obtained by creating a stratification based on ranking information. Using this stratification exactly as is done in stratified sampling or standard post-stratification leads to the standard nonparametric estimators for RSS and JPS. However, we show that strata obtained from ranking information satisfy additional constraints that need not be met by ordinary strata. Specifically, the in-stratum cumulative distribution functions (CDFs) can be no more extreme, in a certain sense, than the CDFs for order statistics from the overall distribution. The additional constraints can be used to obtain better small-sample estimates of the in-stratum CDFs using either RSS or JPS. In the JPS case, the constraints also lead to better small-sample estimates of the overall CDF and the population mean.

Keywords Convexity · Maximum likelihood · Ranked-set sampling · Stratified sampling

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 units without actually measuring them. The ranking information is used to guide the selection of the

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

O. Ozturk (✉)
Department of Statistics, The Ohio State University, Columbus, OH 43210, USA
e-mail: omer@stat.osu.edu

units to be measured, and the result is a sample that tends to be more informative than a sample obtained via simple random sampling (SRS). The problem that motivated McIntyre (1952, 2005) was that of estimating average yields in agriculture. Measuring a yield is expensive because one must actually harvest the crops, but an expert may be able to produce an accurate ranking of the yields for a collection of plots on the basis of a visual inspection.

RSS may be either balanced or unbalanced. To implement balanced RSS, one first specifies both a number of cycles n and a set size m. The set size m is typically chosen to be between two and five so that sets of m units can easily be ranked accurately. One then draws $N = nm$ independent simple random samples (sets) of size m and ranks the units within each sample from smallest to largest using some method other than actual measurement. From each of the first n samples, the unit ranked smallest is selected for measurement. From each of the next n samples, the unit ranked second-smallest is selected for measurement, and so on. If the rankings are perfectly accurate, then the ranked-set sample consists of N independent measurements, with n values distributed like the smallest order statistic from a sample of size m, n values distributed like the second-smallest order statistic from a sample of size m, and so on. If the rankings are not perfectly accurate, then the sample consists of independent *judgment* order statistics.

To implement unbalanced RSS, one relaxes the requirement that the same number of each type of order statistic be chosen. One still specifies a set size m, but instead of specifying a number of cycles n, one specifies a vector $\mathbf{n} \equiv (n_1, \ldots, n_m)$, where n_i indicates the number of units with rank i to be selected for measurement. The total sample then consists of $N = \sum_{i=1}^{m} n_i$ independent order statistics or judgment order statistics.

RSS allows for more efficient statistical inference than does SRS in a wide variety of statistical problems. These problems include 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), nonparametric estimation of the population variance (MacEachern et al. 2002), and best linear unbiased estimation (Barnett and Moore 1997). However, one must use the RSS sampling approach described above in order to realize the benefits. Judgment post-stratification (JPS), proposed by MacEachern et al. (2004), is an alternate method that uses the same ranking information that is used in RSS, but that is based on a simple random sample. As a result, researchers who use JPS retain the option of using SRS-based methods if desired.

To implement JPS, one first specifies both a total sample size N and a set size m. As in RSS, the set size m is typically chosen to be between two and five so that the sets of m units can easily be ranked accurately. 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 the set are ranked from smallest to largest, and the rank of the unit from the simple random sample is recorded. The full data set then consists of both the N measured values and the rank associated with each measured value. Just as in unbalanced RSS, the number of measured units with each particular rank may differ from one rank to another. In fact, some ranks may not appear at all, particularly if N is small. In what follows, we let the vector $\mathbf{n} \equiv (n_1, \ldots, n_m)$, where $N = \sum_{i=1}^{m} n_i$, indicate the number of sampled values with each rank 1 to m. JPS may be thought of as a randomized version of balanced RSS; it is equivalent to drawing N independent simple random samples of size m, ranking the units in each sample from smallest to largest, and then deciding at random which one unit in each sample to measure.

Like RSS, JPS allows for more efficient statistical inference than does SRS. For example, it allows for more efficient estimation of population means (MacEachern et al. 2004) and more efficient estimation of contrasts in designed experiments (Du and MacEachern 2008). JPS also offers more flexibility than RSS in that rankers can be permitted to declare ties.

Neither RSS nor JPS requires perfectly accurate rankings in order to be effective, but more accurate rankings tend to lead to more efficient inference. In addition, recent work by Ozturk (2007) and Wang et al. (2008) has shown that if one is willing to assume that the in-stratum distributions are stochastically ordered, then even greater gains in efficiency can be obtained using in-stratum estimators that satisfy the same stochastic-ordering constraints. The stochastic-ordering assumption is certainly reasonable in many settings. However, in small-sample settings where only limited data are available and where the ranking information may be subjective information that does not translate easily into covariates, there is a danger that the rankings may be inaccurate in surprising ways. In this paper, we develop methods for obtaining better inference without making *any* additional assumptions at all.

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, we show in this paper that strata that arise from ranking information have a structure that is not present for ordinary strata. Specifically, the collection of in-stratum CDFs must satisfy, at each point on the real line, constraints that force them to be no more extreme, in a certain sense, than the CDFs for order statistics from the overall distribution. After deriving these constraints in Sect. 2, we show in Sect. 3 that when sample sizes are small, the constraints can be used to obtain better estimates of the in-stratum CDFs. In the JPS case, these better estimates of the in-stratum CDFs also lead to a better estimate of the overall CDF. In Sect. 4, we show that in the JPS case, we can also obtain a better estimate of the population mean. In Sect. 5, we discuss a small data example, and in Sect. 6, we give our conclusions and mention some possible extensions of the method.

2 The constraints

In this section, we derive the constraints that must hold for the in-stratum CDFs when the strata arise from ranking information. We then use these constraints to prove some useful results about the geometry of the space of possible judgment-order-statistic CDF values at a particular point. Our first theorem shows that, in a certain sense, the in-stratum CDF values can never be more extreme than the CDFs for order statistics from the overall distribution.

Theorem 1 *Assume that $m > 1$. Suppose that $\mathbf{p} \equiv (p_1, \dots, p_m)$ is the vector of in-stratum CDF values at the point t, and let $\bar{p} = \frac{1}{m}\sum_{i=1}^{m} p_i$ be the overall CDF value at t. Then, if I is any subset of the set $\{1, 2, \dots, m\}$, we have that*

$$\sum_{i \in I} p_i \leq \sum_{i=1}^{|I|} B(\bar{p}; i, m+1-i), \tag{1}$$

where $B(\cdot; \alpha, \beta)$ is the CDF for the beta distribution with parameters α and β.

Proof Let I be any subset of $\{1, 2, \dots, m\}$. Letting $X_{[i]}$ indicate a random ith judgment order statistic and $X_{(i)}$ indicate a random true ith order statistic, we have that

$$\begin{aligned}
\sum_{i \in I} p_i &= \sum_{i \in I} P(X_{[i]} \leq t) = E\left[\sum_{i \in I} I(X_{[i]} \leq t)\right] \\
&= E\left[E\left[\sum_{i \in I} I(X_{[i]} \leq t) | X_{(1)}, \dots, X_{(m)}\right]\right] \\
&= E\left[E\left[\# \text{ of values } X_{[i]},\ i \in I \text{ less than or equal to } t\ |X_{(1)}, \dots, X_{(m)}\right]\right].
\end{aligned}$$

The values $X_{[i]}, i \in I$ are a subset of $|I|$ values chosen from $X_{(1)}, \dots, X_{(m)}$. The smallest $|I|$ values in the list $X_{(1)}, \dots, X_{(m)}$ are $X_{(1)}, \dots, X_{(|I|)}$. Thus, the number of values $X_{[i]}, i \in I$ less than or equal to t can be no larger than the number of values $X_{(1)}, \dots, X_{(|I|)}$ less than or equal to t. This tells us that

$$\begin{aligned}
\sum_{i \in I} p_i &= E\left[E\left[\# \text{ of values } X_{[i]},\ i \in I \text{ less than or equal to } t\ |X_{(1)}, \dots, X_{(m)}\right]\right] \\
&\leq E\left[E\left[\# \text{ of values } X_{(i)},\ i = 1, \dots, |I| \text{ less than or equal to } t \right.\right. \\
&\qquad \left.\left.|X_{(1)}, \dots, X_{(m)}\right]\right] \\
&= \sum_{i=1}^{|I|} P(X_{(i)} \leq t) = \sum_{i=1}^{|I|} B(\bar{p}; i, m+1-i),
\end{aligned}$$

proving the theorem. □

Theorem 1 allows us to prove Theorem 2, which characterizes the space of all possible values for the in-stratum CDFs when the overall CDF has the value r. Theorem 2 shows that the constraints (1) given in Theorem 1 are sufficient as well as necessary. Moreover, if we define K_r to be the set of all vectors $\mathbf{p}$ such that the constraints (1) and the constraint $\bar{p} = r$ hold, then Theorem 2 identifies the vertices for K_r.

Theorem 2 *Assume that $m > 1$. Suppose that at the point t, the overall CDF has the value r. Then the possible values $\mathbf{p} \equiv (p_1, \dots, p_m)$ for the vector of in-stratum CDF values at the point t is precisely the set of all convex combination of the* *vectors* $(o_{\pi(1)}, \ldots, o_{\pi(m)})$, $\pi \in S_m$, *where* S_m *is the set of all permutations of the set* $\{1, 2, \ldots, m\}$ *and* o_i *is defined by* $o_i \equiv B(r; i, m+1-i)$, $i = 1, \ldots, m$.

Proof We first show that any convex combination of the vectors $(o_{\pi(1)}, \ldots, o_{\pi(m)})$, $\pi \in S_m$ can arise as a vector of in-stratum CDF values. We then show that no other vectors are possible by proving that the vectors $(o_{\pi(1)}, \ldots, o_{\pi(m)})$, $\pi \in S_m$, are the vertices for the space K_r of vectors $\mathbf{p}$ that satisfy the constraints (1) and yield an overall CDF value of r.

Let $(p_1, \ldots, p_m)$ be an arbitrary convex combination of the vectors $(o_{\pi(1)}, \ldots, o_{\pi(m)})$, $\pi \in S_m$. We can then write that

$$(p_1, \ldots, p_m) = \sum_{\pi \in S_m} c_\pi (o_{\pi(1)}, \ldots, o_{\pi(m)}),$$

where the c_π are nonnegative constants that sum to 1. To show that $(p_1, \ldots, p_m)$ can actually be realized as a vector of in-stratum CDFs, we suppose that with probability c_π, the true order statistics $X_{(1)}, \ldots, X_{(m)}$ in a particular set are judgment-ranked from smallest to largest as $X_{(\pi(1))}, \ldots, X_{(\pi(m))}$. We also assume that the judgment ranking of the true order statistics depends on the actual data only through the true ranking of the data values. The vector of in-stratum CDF values at the point t is then given by $\sum_{\pi \in S_m} c_\pi (o_{\pi(1)}, \ldots, o_{\pi(m)})$, which is precisely $(p_1, \ldots, p_m)$.

To show that no other vectors $\mathbf{p}$ can be realized as vectors of in-stratum CDF values, we show that the vectors $(o_{\pi(1)}, \ldots, o_{\pi(m)})$, $\pi \in S_m$ are the vertices of the convex set K_r consisting of all vectors $\mathbf{p}$ that satisfy both the constraints (1) specified in Theorem 1 and the constraint $\bar{p} = r$. We note first that the set of all vectors $\mathbf{p}$ that satisfy the condition $\bar{p} = r$ for fixed r is an $(m-1)$-dimensional convex polytope (see Ziegler 1995). Thus, any vertex of the set must achieve equality in at least $m-1$ of the inequalities given by (1), and the inequality determined by choosing I to be the full set $\{1, 2, \ldots, m\}$ may not count towards the $m-1$.

Suppose that $\mathbf{p}$ is a vertex of K_r, and suppose that two of the equalities satisfied by $\mathbf{p}$ involve the same number of indices. That is, suppose that there are nonidentical subsets I_1 and I_2 of $\{1, 2, \ldots, m\}$ such that $|I_1| = |I_2|$ and

$$\sum_{i \in I_1} p_i = \sum_{i=1}^{|I_1|} o_i = \sum_{i \in I_2} p_i. \tag{2}$$

Let j be a value in $I_1 \setminus I_2$. Then, by Theorem 1, we know that

$$\left(\sum_{i \in I_1} p_i \right) - p_j \leq \sum_{i=1}^{|I_1|-1} o_i.$$

Subtracting this new inequality from (2), we find that $p_j \geq o_{|I_1|}$. Consider the set I_3 that consists of I_2 together with j. The set I_3 includes $|I_2|+1$ indices, and we have that

$$\sum_{i \in I_3} p_i = \left(\sum_{i=1}^{|I_2|} o_i \right) + p_j \geq \left(\sum_{i=1}^{|I_2|} o_i \right) + o_{|I_2|}$$
$$> \left(\sum_{i=1}^{|I_2|} o_i \right) + o_{|I_2|+1} = \sum_{i=1}^{|I_3|} o_i,$$

where the strict inequality follows from the fact that $o_1 > \cdots > o_m$. The subset I_3 thus violates the constraints (1) from Theorem 1, providing a contradiction. This shows that the $m-1$ equalities must each involve a different number of indices.

Since there are only $m-1$ possible numbers of indices, and since $\mathbf{p}$ must satisfy $m-1$ different equalities, it must be true that for each $k = 1, \ldots, m-1$, there is a set of k in-stratum CDF values that sum to $\sum_{i=1}^{k} o_i$. The in-stratum CDF values that sum to $\sum_{i=1}^{k} o_i$ must be the k largest in-stratum CDF values, for otherwise the constraints (1) will be violated. Thus, the in-stratum CDF values, ordered from largest to smallest, must coincide exactly with the values $o_1, \ldots, o_m$. That is, the values $p_1, \ldots, p_m$ must be a permutation of the values $o_1, \ldots, o_m$. This shows that the vertices of the convex set K_r are precisely the vectors $(o_{\pi(1)}, \ldots, o_{\pi(m)})$, $\pi \in S_m$, proving the theorem. □

Theorem 2 shows that when the overall CDF value is fixed at $r \in [0, 1]$, the space K_r of possible values for $\mathbf{p}$ is a convex set. The next theorem shows that if we define $K \equiv \bigcup_{r \in [0,1]} K_r$ to be the set of all possible values for $\mathbf{p}$, then K is also convex. To prove the theorem, we need the following lemma. As before, $B(\cdot; \alpha, \beta)$ is the CDF for the beta distribution with parameters α and β.

Lemma 1 *For any $k = 1, \ldots, m-1$, the function $B(t; 1, m) + \cdots + B(t; k, m+1-k)$ is concave in t on the interval* $[0, 1]$.

Proof We proceed by showing that the second derivative of the function is nonpositive on the interval $[0, 1]$. Consider first the case $k = 1$. The first derivative of $B(t; 1, m)$ is the beta density

$$b(t; 1, m) = m(1-t)^{m-1}.$$

Differentiating again with respect to t, we find that the second derivative is $-m(m-1)(1-t)^{m-2}$, which is nonpositive for $t \in [0, 1]$. Thus, the lemma holds for $k = 1$.

Suppose that $k > 1$. The first derivative of $B(t; k, m+1-k)$ is the beta density

$$b(t; k, m+1-k) = \frac{m!}{(k-1)!(m-k)!} t^{k-1} (1-t)^{m-k},$$

which means that the second derivative of $B(t; k, m+1-k)$ with respect to t is

$$\frac{m!}{(k-1)!(m-k)!} \left((k-1) t^{k-2} (1-t)^{m-k} - (m-k) t^{k-1} (1-t)^{m-k-1} \right)$$
$$= \frac{m!}{(k-2)!(m-k)!} t^{k-2} (1-t)^{m-k} - \frac{m!}{(k-1)!(m-k-1)!} t^{k-1} (1-t)^{m-k-1},$$

a sum of two terms. However, the positive term in the second derivative of $B(t; k, m+1-k)$ is exactly the same as the negative term in the second derivative of $B(t; k-1, m+2-k)$. Thus, when we compute the second derivative of the sum $B(t; 1, m) + \cdots + B(t; k, m+1-k)$ with respect to t, all but one of the terms cancel out, and the final result is

$$-\frac{m!}{(k-1)!(m-k-1)!} t^{k-1} (1-t)^{m-k-1},$$

which is nonpositive for $t \in [0, 1]$. This proves the lemma. □

Theorem 3 *The space K of all possible vectors of in-stratum CDF values is a convex set.*

Proof We proceed by showing that any convex combination of two vectors of in-stratum CDF values must also be a vector of in-stratum CDF values. To do this, we show that the bounds specified in (1) behave in an appropriate way.

Let $\mathbf{p_1} \equiv (p_{11}, \ldots, p_{1m})$ and $\mathbf{p_2} \equiv (p_{21}, \ldots, p_{2m})$ be arbitrary vectors of in-stratum CDF values. Let $\bar{p}_1 = \frac{1}{m}\sum_{i=1}^{m} p_{1i}$ and $\bar{p}_2 = \frac{1}{m}\sum_{i=1}^{m} p_{2i}$ be the corresponding values for the overall CDF, and let I be an arbitrary subset of $\{1, 2, \ldots, m\}$. When the overall CDF is $\bar{p}_j$, the upper bound on $\sum_{i \in I} p_{ji}$ specified in (1) is $\sum_{i=1}^{|I|} B(\bar{p}_j; i, m+1-i)$. Thus, $\mathbf{p_1}$ and $\mathbf{p_2}$ satisfy

$$\sum_{i \in I} p_{1i} \le \sum_{i=1}^{|I|} B(\bar{p}_1; i, m+1-i) \quad \text{and} \quad \sum_{i \in I} p_{2i} \le \sum_{i=1}^{|I|} B(\bar{p}_2; i, m+1-i).$$

Let $\lambda \in (0, 1)$ be a constant, and let $\mathbf{p_3} \equiv \lambda \mathbf{p_1} + (1-\lambda)\mathbf{p_2}$ be a convex combination of $\mathbf{p_1}$ and $\mathbf{p_2}$. The corresponding value for the overall CDF is $\bar{p}_3 = \lambda \bar{p}_1 + (1-\lambda)\bar{p}_2$, and we have that

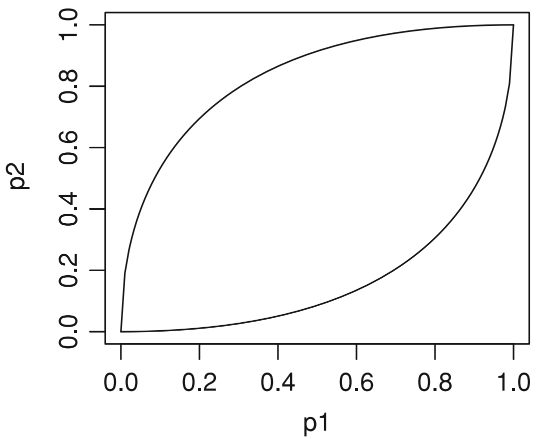
$$\begin{aligned}\sum_{i \in I} p_{3i} &= \lambda \sum_{i \in I} p_{1i} + (1-\lambda) \sum_{i \in I} p_{2i} \\ &\le \lambda \sum_{i=1}^{|I|} B(\bar{p}_1; i, m+1-i) + (1-\lambda) \sum_{i=1}^{|I|} B(\bar{p}_2; i, m+1-i).\end{aligned}$$

By Lemma 1, the function $\sum_{i=1}^{|I|} B(t; i, m+1-i)$ is a concave function of t. Thus,

$$\begin{aligned}\sum_{i \in I} p_{3i} &\le \lambda \sum_{i=1}^{|I|} B(\bar{p}_1; i, m+1-i) + (1-\lambda) \sum_{i=1}^{|I|} B(\bar{p}_2; i, m+1-i) \\ &\le \sum_{i=1}^{|I|} B(\lambda \bar{p}_1 + (1-\lambda)\bar{p}_2; i, m+1-i) = \sum_{i=1}^{|I|} B(\bar{p}_3; i, m+1-i),\end{aligned}$$

which proves the theorem. □

Fig. 1 The space of all possible vectors (p_1, p_2) when the set size is $m = 2$. That is, the enclosed region is the set K for the case $m = 2$



To illustrate the results that we have obtained in this section, we created Figs. 1 and 2. Figure 1 shows the space K of all possible vectors (p_1, p_2) in the $m = 2$ case, and Fig. 2 shows four different slices from the three-dimensional space of all possible vectors (p_1, p_2, p_3) in the $m = 3$ case. One important feature to notice in each figure is that some possibilities are ruled out. For example, in Fig. 1, we see that if p_1 is very small, then it is not possible for p_2 to be large. A second important feature is symmetry. Since the labels $1, \ldots, m$ are interchangeable, each region plotted in the two figures is symmetric with respect to the line $p_1 = p_2$.

3 The restricted CDF estimator

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

$$\hat{F}(t) = \frac{1}{m} \sum_{i=1}^{m} \hat{F}_i(t), \tag{3}$$

where $\hat{F}_i(t)$ is the empirical distribution function (EDF) for the data values that were given rank i. If any of the strata are not represented, as may occur in JPS, then $\hat{F}(t)$ is the average of $\hat{F}_i(t)$ over all the strata with nonzero sample sizes. The estimator $\hat{F}$ is unbiased for F, but the vector $(\hat{F}_1(t), \ldots, \hat{F}_m(t))$ of in-stratum CDF estimates need not be a possible value for $(F_1(t), \ldots, F_m(t))$, the true vector of in-stratum CDF values at t. In situations like this where an estimator may fall outside of the feasible region, one can typically obtain a better estimate using an estimator that is forced to fall inside the feasible region. In this section, we provide a method for obtaining such an estimator of $(F_1(t), \ldots, F_m(t))$.

Suppose that we wish to estimate the values $F_1(t), \ldots, F_m(t)$ for some fixed real number t. 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 t. If $\mathbf{p} \equiv (p_1, \ldots, p_m)$ is a candidate vector of in-stratum

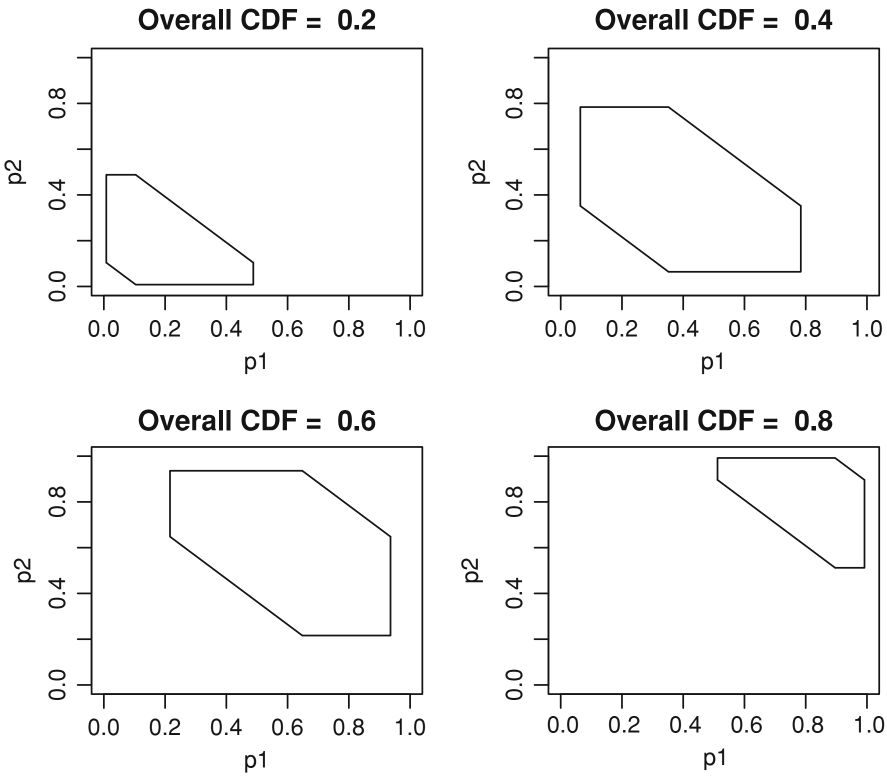


Fig. 2 Some slices from the space of possible vectors (p_1, p_2, p_3) when the set size is $m = 3$. Each *plot* shows the potential values for p_1 and p_2 when the overall CDF value $\bar{p}$ is fixed at some particular value. The value p_3 is determined by the constraint that $\bar{p} = \frac{1}{3}(p_1 + p_2 + p_3)$

CDF values, then the likelihood function is given by

$$L(\mathbf{p}|\mathbf{x}, \mathbf{n}) = \prod_{i=1}^{m} \binom{n_i}{x_i} p_i^{x_i}(1 - p_i)^{n_i - x_i} \quad \text{for } \ 0 \le p_i \le 1, \ i = 1, \ldots, m. \tag{4}$$

Thus, the log likelihood function can be written as

$$l(\mathbf{p}|\mathbf{x}, \mathbf{n}) = c + \sum_{i=1}^{m} \{x_i \log(p_i) + (n_i - x_i)\log(1 - p_i)\} \quad \text{for } \ 0 < p_i < 1, \\ i = 1, \ldots, m, \tag{5}$$

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 $(\hat{F}_1(t), \ldots, \hat{F}_m(t))$. However, if we maximize l only over the set K of vectors $\mathbf{p}$ that satisfy the constraints (1) given in Theorem 1, then we obtain the restricted vector of in-stratum estimates

$(\hat{F}_1^r(t), \ldots, \hat{F}_m^r(t))$. That is, we define the restricted estimator by

$$(\hat{F}_1^r(t), \ldots, \hat{F}_m^r(t)) \equiv \arg\max_{\mathbf{p}\in K} l(\mathbf{p}|\mathbf{x}, \mathbf{n}),$$

The corresponding restricted estimator for the overall CDF value is

$$\hat{F}^r(t) \equiv \frac{1}{m}\sum_{i=1}^{m} \hat{F}_i^r(t).$$

The restricted estimator $(\hat{F}_1^r(t), \ldots, \hat{F}_m^r(t))$ is not unbiased. One can show, however, that it is a consistent estimate of the true vector $(F_1(t), \ldots, F_m(t))$ in the sense that if $\min\{n_1, \ldots, n_m\}$ goes to infinity, then $(\hat{F}_1^r(t), \ldots, \hat{F}_m^r(t))$ converges in probability to $(F_1(t), \ldots, F_m(t))$.

We obtain the estimate $(\hat{F}_1^r(t), \ldots, \hat{F}_m^r(t))$ by maximizing the concave function $l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ over the convex set K of all possible vectors of in-stratum CDF values. This maximization can be done in a variety of ways. The algorithm that we use involves (i) taking advantage of Theorems 1 and 2 to find the value $\mathbf{p}$ that maximizes l when the overall CDF value $\bar{p}$ is fixed and (ii) doing a systematic search for the overall CDF value $\bar{p}$ for which the maximizing likelihood is highest. Details of the algorithm are described in the appendix, and an R function for implementing the algorithm is available from the authors.

To illustrate the restricted estimator $\hat{F}^r(t)$ and show how it differs from the standard EDF-based estimator $\hat{F}(t)$, we present a few examples. Since the two estimators coincide when the vector $(\hat{F}_1(t), \ldots, \hat{F}_m(t))$ satisfies the constraints (1), we focus on cases where these constraints are not met. Note that while $(\hat{F}_1(t), \ldots, \hat{F}_m(t))$ and $(\hat{F}_1^r(t), \ldots, \hat{F}_m^r(t))$ will often coincide for many values of t, it is impossible (in the absence of ties) for them to coincide for *all* values of t since the restricted estimator does not allow any of the values $\hat{F}_1^r(t), \ldots, \hat{F}_m^r(t)$ to be 0 or 1 without all of the values $\hat{F}_1^r(t), \ldots, \hat{F}_m^r(t)$ being the same.

Example 1 Suppose that $m = 2$, $\mathbf{n} = (10, 10)$, and $\mathbf{x} = (0, 10)$ when some particular value t is fixed. The standard estimate of the in-stratum CDF values is (0, 1), which does not meet the constraints (1). The restricted estimate is (0.25, 0.75), which does meet the constraints. Note that while the estimates of the in-stratum CDF values are different, the estimate of the overall CDF is 0.5 in either case.

Example 2 Suppose that $m = 2$, $\mathbf{n} = (10, 10)$, and $\mathbf{x} = (0, 5)$ when some particular value t is fixed. The standard estimate of the in-stratum CDF values is (0, 0.5), which does not meet the constraints (1). The restricted estimate is (0.0566, 0.4194), which does meet the constraints. In this example, unlike in Example 1, the estimates of the overall CDF differ slightly. The standard EDF-based estimate is 0.25, while the restricted estimate is 0.238. In the $m = 2$ equal-sample-size setting, the general pattern is that the restricted estimate of the overall CDF value is always farther from 0.5 than is the standard estimate of the overall CDF value. The difference can never be very large, however. In fact, in the $m = 2$ equal-sample-size setting, it is always true that $|\hat{F}^r(t) - \hat{F}(t)| < 0.014$.

Example 3 Suppose that $m = 2$, $\mathbf{n} = (10, 5)$, and $\mathbf{x} = (0, 5)$ when some particular value t is fixed. The standard estimate of the in-stratum CDF values is (0, 1), which does not meet the constraints (1). The restricted estimate is (0.158, 0.637), which does meet the constraints. The standard estimate of the overall CDF value is 0.5, while the restricted estimate is 0.398. Because of the smaller sample size in the second stratum, the sample proportion from the first stratum carries greater weight in determining the restricted estimate than does the sample proportion from the second stratum. The estimate of the in-stratum CDF value for the first stratum moved only from 0 (standard) to 0.158 (restricted), while the estimate for the second stratum moved from 1 (standard) all the way down to 0.637 (restricted). This example shows that in the JPS case, large differences $|\hat{F}^r(t) - \hat{F}(t)|$ can occur when the in-stratum sample sizes are very unbalanced.

To compare the performance of the restricted estimator $\hat{F}^r(t)$ to that of the standard EDF-based estimator $\hat{F}(t)$, we computed mean squared errors (MSEs) for each estimator under different types of rankings, different average in-stratum sample sizes, and different values $\bar{p}$ for the true overall CDF. We also compared the performance of the two estimators both under balanced RSS and under JPS. We modeled the rankings using a one-parameter model that extends from perfect rankings at one extreme to random rankings at the other. Specifically, we assumed that the true CDF $F_i(t)$ for the ith stratum satisfies

$$F_i(t) = \lambda F_{(i)}(t) + (1 - \lambda)F(t), \tag{6}$$

where $F_{(i)}(t)$ is the CDF for a true ith order statistic, $F(t)$ is the overall CDF, and $\lambda \in [0, 1]$ is the parameter. This model is equivalent to assuming that each set of units is either ranked perfectly or ranked at random, with the probability of perfect ranking being λ. Thus, $\lambda = 1$ gives perfect rankings, and $\lambda = 0$ gives random rankings. Alternate models for imperfect rankings are discussed by Ridout and Cobby (1987), Bohn and Wolfe (1994), and Fligner and MacEachern (2006).

Figure 3 shows some results for the case of JPS with $m = 2$ and $\lambda = 1$. The top two panels in Fig. 3 show the MSEs for each estimator as a function of the overall CDF value $\bar{p}$ in the cases where the total sample size is $N = 4$ and $N = 16$. The MSEs for estimating each parameter $F(t)$, $F_1(t)$, and $F_2(t)$ are plotted as separate curves, with dotted lines giving the MSEs for the standard estimator and solid lines giving the MSEs for the restricted estimator. We see that while the restricted estimator does not achieve uniformly smaller MSEs for estimating $F_1(t)$ and $F_2(t)$, the areas beneath the MSE curves are much smaller for the restricted estimator than for the standard estimator. We also see from comparing the top two panels of Fig. 3 that the advantage of the restricted estimator over the standard estimator decreases as the total sample size N increases. The bottom two panels in Fig. 3 show the same information as the top two panels, but in a slightly different form. The plotted curves give the MSEs for the two estimators *relative* to the MSEs for the restricted estimator. We see that while the standard estimator slightly outperforms the restricted estimator in estimating $F(t)$ when the true CDF value is close to 0.5, the restricted estimator far outperforms the standard estimator when the true CDF value lies in either tail of the distribution.

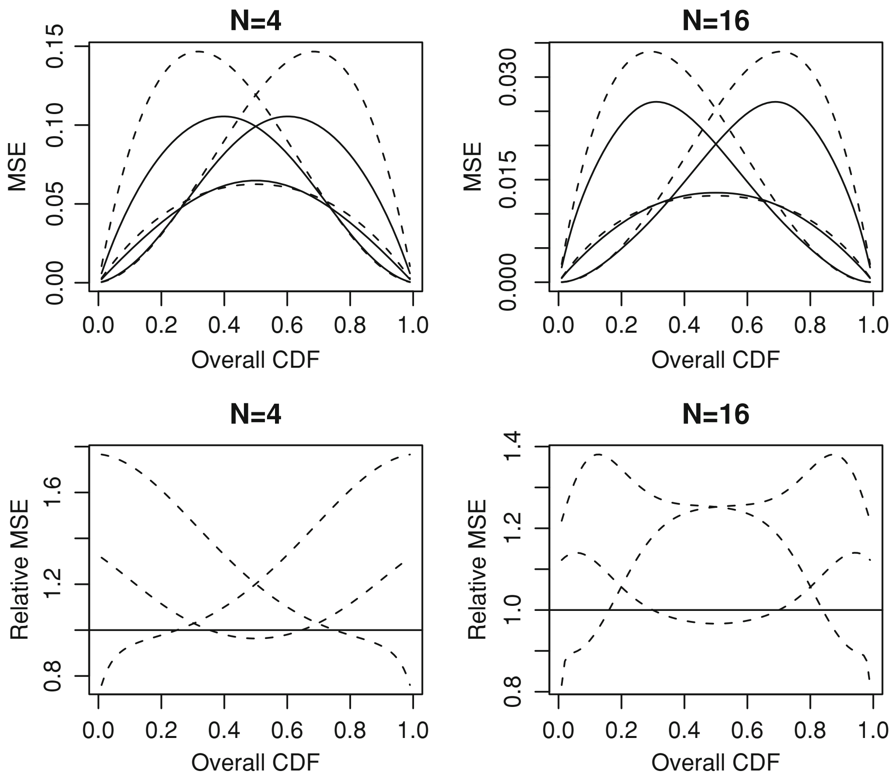


Fig. 3 MSEs and relative MSEs for the restricted CDF estimator (*solid lines*) and the standard CDF estimator (*dashed lines*) in the JPS setting with $m = 2$, total sample sizes $N = 4$ and $N = 16$, and overall CDF values $\bar{p}$ ranging from 0 to 1. The *top two plots* show MSEs, while the *bottom two plots* show MSEs relative to the MSE for the restricted estimator. In the top two plots, the *curves* corresponding to estimates of $F(t)$, $F_1(t)$, and $F_2(t)$ are those with modes in the *center*, *left*, and *right*, respectively. In the bottom two plots, the *dashed curves* can be distinguished by noting that when $\bar{p} = 0.1$, the *top* and *bottom curves* are those corresponding to $F_1(t)$ and $F_2(t)$, respectively

If we compute the areas under curves like those plotted in the top two panels of Fig. 3, then we obtain integrated MSEs (IMSEs). A natural way to compare the performance of the two estimators is by computing the ratio of IMSEs

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

Relative efficiencies larger than 1 indicate that the restricted estimator is outperforming the standard estimator, while relative efficiencies less than 1 indicate the opposite. Table 1 contrasts the relative efficiencies under JPS with the relative efficiencies under balanced RSS for a variety of choices for the parameter λ and the average sample size N/m (or n in the RSS case). The upper half of Table 1 shows results for the RSS case, while the lower half of Table 1 shows results for the JPS case. We see that under either RSS or JPS, the restricted estimator provides better small-sample estimates of

Table 1 Calculated efficiencies for the restricted estimator relative to the standard EDF-based estimator for $m = 2$, RSS and JPS, and different levels of imperfect rankings

Type	Parameter	λ	Average sample size				
			2	4	6	8	10
RSS	$F(t)$	0	0.98	0.98	0.99	0.99	1.00
		1/3	0.98	0.98	0.99	0.99	0.99
		2/3	0.98	0.98	0.99	0.99	0.99
		1	0.98	0.98	0.98	0.98	0.99
	$F_1(t)$	0	1.39	1.20	1.12	1.07	1.05
		1/3	1.39	1.21	1.14	1.10	1.08
		2/3	1.37	1.25	1.21	1.19	1.17
		1	1.32	1.26	1.25	1.24	1.24
JPS	$F(t)$	0	1.06	1.05	1.02	1.01	1.00
		1/3	1.06	1.05	1.02	1.01	1.00
		2/3	1.05	1.05	1.03	1.01	1.01
		1	1.04	1.05	1.03	1.01	1.01
	$F_1(t)$	0	1.45	1.28	1.16	1.09	1.06
		1/3	1.44	1.30	1.18	1.12	1.09
		2/3	1.40	1.33	1.25	1.21	1.18
		1	1.32	1.32	1.29	1.27	1.26

By symmetry, the relative efficiencies for estimating $F_2(t)$ are the same as the relative efficiencies for estimating $F_1(t)$

the in-stratum CDFs $F_1(t)$ and $F_2(t)$ than does the standard EDF-based estimator. When we estimate $F(t)$, however, the picture is somewhat different. Under RSS, the standard estimator slightly outperforms the restricted estimator for estimating $F(t)$. Under JPS, however, the restricted estimator is more efficient than the standard estimator. A comparison of the top half of Table 1 with the bottom half of the table shows that the relative advantage in using the restricted estimator rather than the standard estimator tends to be larger for JPS than for balanced RSS.

The results presented in Fig. 3 and Table 1 show the good performance of the restricted estimator in terms of MSE. However, it is also of interest to separately examine the bias. Figure 4 gives some results for the JPS case when $m = 2$ and $\lambda = 1$. The two plots in Fig. 4 show the bias of the restricted estimators $F^r(t)$, $F_1^r(t)$, and $F_2^r(t)$ as a function of the overall CDF value $\bar{p}$ for total sample sizes $N = 4$ and $N = 16$. We see that under perfect rankings, $F_1^r(t)$ is positively biased, $F_2^r(t)$ is negatively biased, and the overall CDF estimator $F^r(t)$ is hardly biased at all. The magnitude of the bias decreases as the sample size increases.

4 Estimating the population mean

The standard RSS or JPS estimator for the population mean μ can be obtained as

$$\hat{\mu} = \int_t t \, \mathrm{d}\hat{F}(t),$$

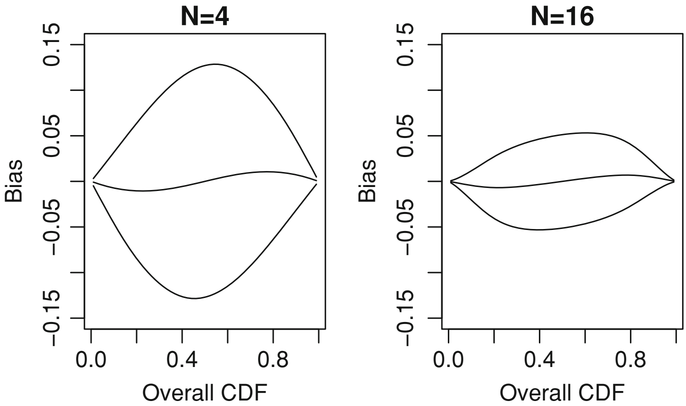


Fig. 4 Bias of the restricted CDF estimators $F^r(t)$, $F_1^r(t)$, and $F_2^r(t)$ in the JPS setting with $m = 2$, total sample sizes $N = 4$ and $N = 16$, and overall CDF values $\bar{p}$ ranging from 0 to 1. From top to bottom, the *curves* correspond to $F_1^r(t)$, $F^r(t)$, and $F_2^r(t)$. All values were computed under an assumption of perfect rankings

where $\hat{F}(t)$ is the standard EDF-based estimator (3) of the overall CDF value that was discussed in Sect. 3. If we replace $\hat{F}(t)$ with $\hat{F}^r(t)$, we obtain the restricted mean estimator

$$\hat{\mu}^r = \int_t t \, \mathrm{d}\hat{F}^r(t).$$

In the balanced RSS case, the restricted estimator is typically not more efficient than the standard estimator $\hat{\mu}$. This is not surprising given our previous finding that in the balanced RSS case, $\hat{F}^r(t)$ is outperformed by $\hat{F}(t)$ in terms of IMSE. However, in the JPS setting, the restricted estimator does tend to be more efficient than the standard estimator. We demonstrate this increased efficiency by presenting results from two related simulation studies. The first study focused on assessing the performance of the restricted mean estimator relative to that of the standard JPS estimator, and the second focused on comparing the performance of the restricted estimator to that of the isotonic JPS mean estimator developed by Wang et al. (2008).

The first simulation study considered different parent distributions, different types of rankings, and different average sample sizes N/m. The parent distributions, which were chosen to represent a range of possibilities in terms of how heavy the tails of the distribution are, were (i) the normal distribution, (ii) the t distribution with $3\,df$, (iii) the $Gamma(5, 1)$ distribution, (iv) the uniform distribution, and (v) the $Beta(1/2, 1/2)$ distribution. The rankings used were those given by the model (6) from Section 3, and the average sample sizes considered were $1, 2, \ldots, 10$. The set size was fixed at $m = 2$, and 100,000 samples were simulated for each combination of parent distribution, type of ranking ($\lambda = 0, 1/3, 2/3, 1$), and average sample size. For each combination of factor levels, a relative efficiency value was computed as

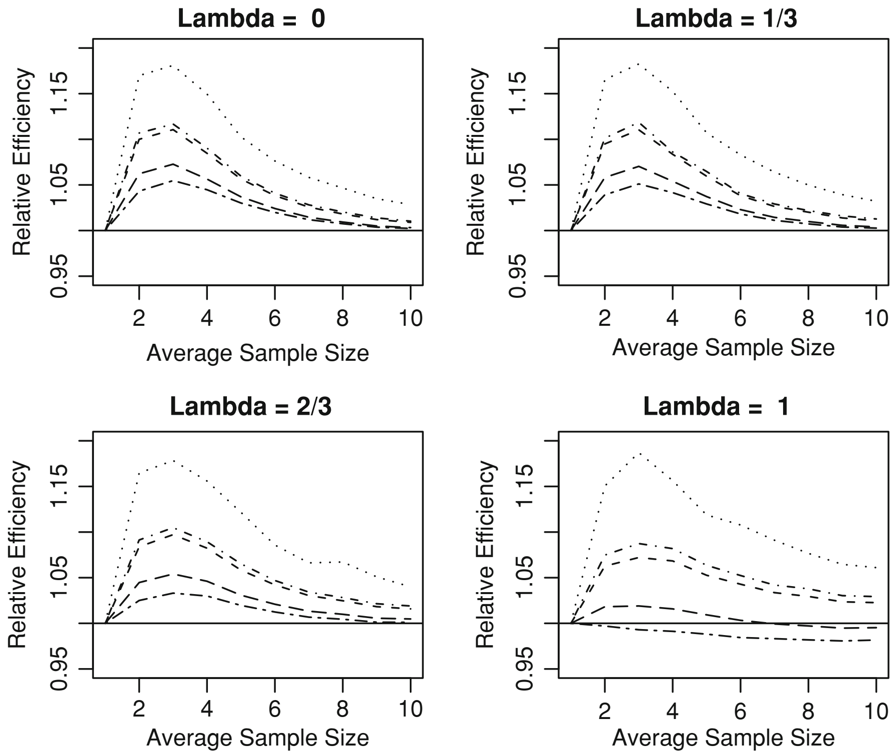


Fig. 5 Simulated efficiencies for the restricted JPS mean estimator relative to the standard JPS mean estimator for different types of imperfect rankings, different underlying distributions, and different average sample sizes when $m = 2$. The parameter λ is the proportion of all sets that are ranked perfectly. The distributions are normal (*dashed*), t with $3\,df$ (*dotted*), $Gamma(5, 1)$ (*dotdash*), uniform (*longdash*), and $Beta(1/2, 1/2)$ (*twodash*). Note that higher simulated efficiencies and heavier tails seem to go together

$$\text{Relative efficiency} = \frac{\text{MSE}(\hat{\mu})}{\text{MSE}(\hat{\mu}^r)}.$$

The results are presented in Figs. 5 and 6.

Figure 5, which summarizes the results by the type of ranking, shows that except for certain cases involving the very light-tailed uniform and $Beta(1/2, 1/2)$ distributions and perfect rankings, the restricted mean estimator outperforms the standard EDF-based mean estimator. The relative efficiencies are highest when the average sample size is around three, and they decrease as the average sample size increases beyond that value. The relative efficiencies are highest for the heavy-tailed distributions (t with $3\,df$ and $Gamma(5, 1)$) and lowest for the light-tailed distributions (uniform and $Beta(1/2, 1/2)$). This finding is not surprising given that $\hat{F}^r(t)$ outperforms $\hat{F}(t)$ for estimating overall CDF values near the tails, but not for estimating overall CDF values near the center of the distribution.

Figure 6 depicts essentially the same results shown in Fig. 5, but it groups the results by parent distribution rather than by the type of ranking. We see from the bottom two

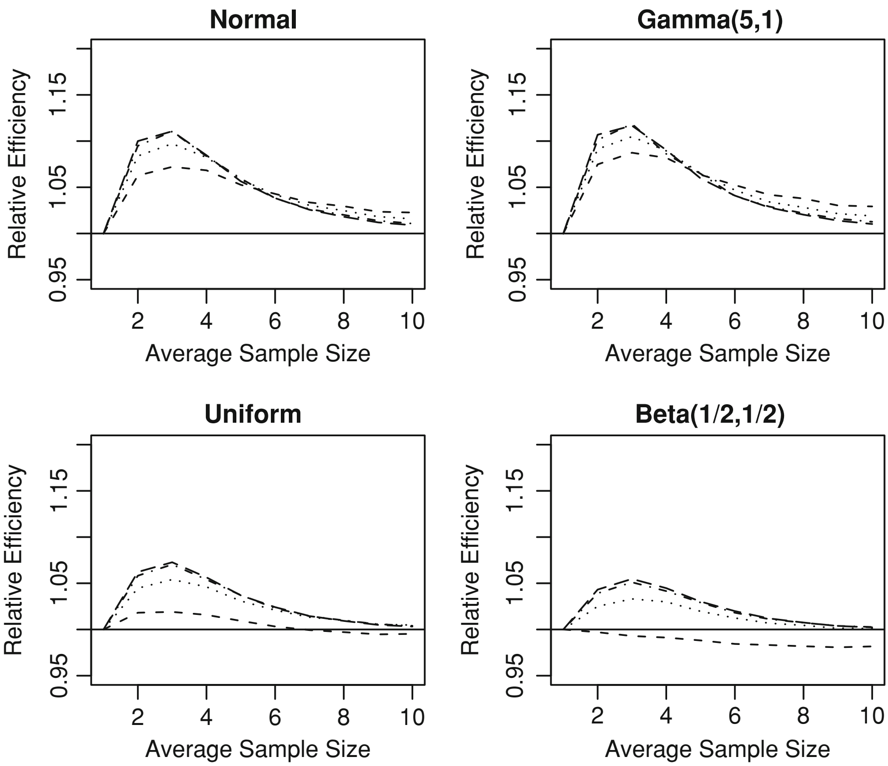


Fig. 6 Simulated efficiencies for the restricted JPS mean estimator relative to the standard JPS mean estimator for different types of imperfect rankings, different underlying distributions, and different average sample sizes when $m = 2$. The parameter λ is the proportion of all sets that are ranked perfectly. The values of the parameter λ are 1 (*dashed*), 2/3 (*dotted*), 1/3 (*dotdash*), and 0 (*longdash*)

plots in Fig. 6 that when the distribution is light-tailed, the relative efficiencies tend to become smaller as the rankings improve. However, we see from the top two plots that for heavy-tailed distributions, the effect of the rankings depends on the average sample size. For small average sample sizes, the relative efficiency is highest when the rankings are random, but for larger average sample sizes, the relative efficiency is highest when the rankings are perfect. Considering Figs. 5 and 6 as a whole, we see strong evidence that the restricted mean estimator is preferable to the standard mean estimator in the small-sample JPS setting.

The second simulation study also considered different parent distributions, different types of rankings, and different average sample sizes N/m. The parent distributions used were (i) the normal distribution, (ii) the $Gamma(5, 1)$ distribution, (iii) the uniform distribution, and (iv) the $Beta(1/2, 1/2)$ distribution. The types of rankings used were (i) perfect rankings, (ii) random rankings, and (iii) perfectly wrong rankings obtained by ranking the units in each set in an order exactly opposite to the true ordering. For the first two types of rankings, the stochastic ordering assumption used by Ozturk (2007) and Wang et al. (2008) holds, but for the third ranking type, that

Table 2 Simulated efficiencies for the restricted JPS mean estimator and the isotonic mean estimator of Wang et al. (2008) (WLS) relative to the standard JPS estimator for different types of rankings, different underlying distributions, and different average sample sizes when $m = 2$

Type of rankings	Distribution	Est.	Average sample size				Est.	Average sample size			
			2	3	4	5		2	3	4	5
Perfect	Normal	New	1.06	1.07	1.07	1.05	WLS	1.02	1.02	1.01	1.01
	$Gamma(5, 1)$	New	1.07	1.09	1.08	1.06	WLS	1.02	1.02	1.01	1.01
	Uniform	New	1.02	1.02	1.02	1.01	WLS	1.03	1.03	1.02	1.01
	$Beta(1/2, 1/2)$	New	1.00	0.99	0.99	0.99	WLS	1.03	1.04	1.03	1.02
Random	Normal	New	1.10	1.11	1.08	1.06	WLS	1.08	1.10	1.08	1.07
	$Gamma(5, 1)$	New	1.11	1.12	1.09	1.06	WLS	1.08	1.10	1.09	1.07
	Uniform	New	1.06	1.07	1.06	1.04	WLS	1.08	1.10	1.08	1.07
	$Beta(1/2, 1/2)$	New	1.04	1.05	1.04	1.03	WLS	1.08	1.10	1.08	1.07
Wrong	Normal	New	1.06	1.07	1.07	1.05	WLS	0.94	0.87	0.81	0.79
	$Gamma(5, 1)$	New	1.07	1.09	1.08	1.06	WLS	0.95	0.89	0.83	0.81
	Uniform	New	1.02	1.02	1.02	1.01	WLS	0.92	0.85	0.79	0.77
	$Beta(1/2, 1/2)$	New	1.00	0.99	0.99	0.99	WLS	0.92	0.85	0.80	0.77

The three types of rankings are perfect rankings, random rankings, and perfectly wrong rankings

assumption fails. The set size was fixed at $m = 2$, and 100,000 samples were simulated for each combination of parent distribution, type of ranking, and average sample size. For each combination of factor levels, a relative efficiency value was computed just as in the first simulation study. The results are presented in Table 2, which gives results for the restricted estimator on the left side and results for the isotonic mean estimator on the right side.

We see from Table 2 that when the stochastic ordering assumption fails, as it does when the rankings are perfectly wrong, then the restricted estimator may significantly outperform the isotonic mean estimator. Even when the stochastic ordering assumption holds, however, the restricted estimator may be every bit as good as the isotonic mean estimator. Under perfect rankings, the restricted estimator outperforms the isotonic mean estimator when the data come from either of the two relatively heavy-tailed distributions (normal and $Gamma(5, 1)$), but the isotonic mean estimator outperforms the restricted estimator when the distribution is uniform or $Beta(1/2, 1/2)$. A similar pattern holds when the rankings are random, with the restricted estimator slightly outperforming the isotonic estimator for normal and $Gamma(5, 1)$ data, but not for uniform and $Beta(1/2, 1/2)$ data. Overall, the restricted JPS mean estimator seems to be at least as effective as the isotonic mean estimator in this $m = 2$ setting.

5 An example

As an illustration of the restricted mean estimator developed in Sect. 4, we use the estimator to estimate the mean body fat percentage for a certain population of men. Our main data set, taken from the StatLib Datasets Archive (2009), consists of

Table 3 Standard and restricted JPS estimates of the overall CDF for the body fat percentage data

Body fat percentage	Rank	Value of $\hat{F}(\cdot)$	Value of $\hat{F}^r(\cdot)$
8.3	2	0.125⋆	0.127
14.7	1	0.375⋆	0.320
18.7	3	0.458⋆	0.432
19.2	4	0.521⋆	0.516
20.4	4	0.583⋆	0.581
21.8	2	0.708⋆	0.696
29.0	3	0.792⋆	0.789
31.6	4	0.854⋆	0.855
32.9	4	0.917⋆	0.912
33.6	3	1.000	1.000

Stars indicate points where the in-stratum CDF values do not satisfy the constraints (1)

measurements of body fat percentage and various body circumference measurements for 252 men. We think of these 252 men as a random sample from a much larger population, and we apply JPS to the set of 252 men to obtain an estimate. The gold standard measurement of body fat percentage requires underwater weighing, which is much more expensive than making body circumference measurements. Thus, it is natural to apply RSS or JPS in this setting. We chose to use hip circumference as the variable for ranking. Using set size $m = 4$, we drew a JPS of size $N = 10$. The ordered measured data values and the corresponding ranks are given in the left-most two columns of Table 3.

Using the measured data values and the corresponding ranks, we computed both the standard JPS CDF estimate $\hat{F}(\cdot)$ and the restricted CDF estimate $\hat{F}^r(\cdot)$. These values are given in the right-most two columns of Table 3. We see from the table that apart from the bottom line, where both CDF estimates are necessarily 1, the two estimates never coincide. Thus, the standard in-stratum CDF estimates violate the constraints (1) at every point where it is possible for those constraints to be violated. One of the biggest deviations occurs at the third ordered value 18.7, where $\mathbf{n} = (1, 2, 3, 4)$ and $\mathbf{x} = (1, 1, 1, 0)$. At that point, the standard in-stratum CDF estimates are $(\hat{F}_1(18.7), \ldots, \hat{F}_4(18.7)) = (1.000, 0.500, 0.333, 0.000)$, and the restricted estimates are $(\hat{F}_1^r(18.7), \ldots, \hat{F}_4^r(18.7)) = (0.896, 0.478, 0.320, 0.035)$, which satisfy the constraints (1). Integrating, we find that the estimates for the population mean are $\hat{\mu}^r = 20.72$ and $\hat{\mu} = 21.04$.

6 Conclusions and possible extensions

In this paper, we have shown that strata arising from ranking information must satisfy additional constraints that need not hold for strata that arise in other ways. We have also shown that by taking advantage of these additional constraints, we can obtain better small-sample estimates of the in-stratum CDFs using either RSS or JPS. In the JPS case, we can also obtain better small-sample estimates of the overall CDF and the population mean. More efficient estimation of the overall CDF and the population mean is of obvious importance, and better estimation of in-stratum CDFs is also valuable. For example, better estimates of the in-stratum CDFs give better information about the ranking process, and they can be used to create statistical procedures that are calibrated to account for the effect of inaccuracies in the rankings (see Ozturk 2007). In comparing the restricted estimator to the standard EDF-based estimator, we have focused on the case $m = 2$ for computational convenience. However, our work with larger set sizes ($m > 2$) suggests that gains of comparable or larger magnitude are available for larger values of m.

While Ozturk (2007) and Wang et al. (2008) used a stochastic-ordering assumption to obtain more efficient inference, the gains offered by our methods do not require any additional assumptions. Nonetheless, our new methods are sometimes every bit as effective as methods that do make a stochastic-ordering assumption. For example, we saw in Sect. 4 that when $m = 2$, our new JPS mean estimator is just as good as the isotonic mean estimator due to Wang et al. (2008) when the rankings are random or perfect, and significantly better when the stochastic ordering assumption is violated. Another approach to incorporating stochastic ordering constraints might involve imposing those constraints on top of the constraints (1) that we have used in this paper. For example, if the constraint $p_1 \geq p_2$ is believed to hold in the $m = 2$ setting depicted in Fig. 1, then the estimate of (p_1, p_2) could be required to fall in that portion of the pictured region that lies below the line $p_1 = p_2$. Maximizing the log likelihood (5) over all vectors $\mathbf{p}$ of in-stratum CDF values that satisfy all constraints might then lead to even better estimates of the in-stratum CDF values and the overall CDF value.

Previous work on JPS has shown that JPS is more flexible than RSS in that rankers can be permitted to declare ties. This continues to be true for the methods presented in this paper. When estimating the CDF in a situation where ties have been declared, one simply allows the in-stratum sample sizes $\mathbf{n} \equiv (n_1, \ldots, n_m)$ and the counts $\mathbf{x} \equiv (x_1, \ldots, x_m)$ defined in Sect. 3 to be noninteger values. One then defines the likelihood $L(\mathbf{p}|\mathbf{n}, \mathbf{x})$ using exactly the formula (4) that was used earlier, and one maximizes the log likelihood over the same convex set K to obtain estimates of the in-stratum CDF values and the overall CDF value.

7 Appendix: The optimization algorithm

Our algorithm for maximizing the log likelihood (5) over the space K consists of two parts. The first part of the algorithm is a procedure, motivated by Theorems 1 and 2, for maximizing the log likelihood $l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ over the space K_r where the overall CDF value is fixed at r. The second part of the algorithm is a procedure for finding the value of r where the maximizing log likelihood value is largest. We describe these two parts below.

Suppose that r is fixed. Our procedure for maximizing the log likelihood (5) over the space K_r consists of starting at a point in K_r, making a series of moves, each of which increases the value of $l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ while keeping $\mathbf{p}$ inside the space K_r, and stopping when no further increases of significant size are possible. Implementing the algorithm requires computing quantities that are related to the directional derivatives of $l(\mathbf{p}|\mathbf{x}, \mathbf{n})$. Since $l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ is given by (5), we have that

$$\frac{\mathrm{d}}{\mathrm{d}p_i} l(\mathbf{p}|\mathbf{x}, \mathbf{n}) = \frac{x_i}{p_i} - \frac{(n_i - x_i)}{1 - p_i}.$$

It thus follows that the derivative of $l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ at $\mathbf{p} \equiv (p_1, \ldots, p_m)$ in the direction of $\mathbf{q} \equiv (q_1, \ldots, q_m) \in K_r$ is given by

$$D(\mathbf{p}, \mathbf{q}) \equiv \sum_{i=1}^{m} \left(\frac{x_i}{p_i} - \frac{(n_i - x_i)}{1 - p_i} \right) (q_i - p_i).$$

The value $D(\mathbf{p}, \mathbf{q})$ is positive if and only if moving from $\mathbf{p}$ in the direction of $\mathbf{q}$ would increase the value of l, and $D(\mathbf{p}, \mathbf{v}) \leq 0$ for all vertices $\mathbf{v}$ of K_r if and only if $\mathbf{p}$ maximizes $l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ over K_r. Our maximization procedure is as follows.

Procedure 1: Maximize $l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ over K_r using the following steps.

Step I Set $p_i = r$ for $i = 1, \ldots, m$. Then $\mathbf{p} \in K_r$.

Step II For each vertex $\mathbf{v_j}$ of K_r, find the directional derivative $D(\mathbf{p}, \mathbf{v_j})$. Note that by Theorem 2, the space K_r has $m!$ distinct vertices.

Step III If none of the derivatives $D(\mathbf{p}, \mathbf{v_j})$ is bigger than some cut-off value, say 0.00001, then stop, taking the current value of $\mathbf{p}$ as the maximizer. Otherwise, compute the weighted average vector

$$\mathbf{v_{move}} = \frac{\sum_{j=1}^{m!} D(\mathbf{p}, \mathbf{v_j}) I(D(\mathbf{p}, \mathbf{v_j}) > 0)\, \mathbf{v_j}}{\sum_{j=1}^{m!} D(\mathbf{p}, \mathbf{v_j}) I(D(\mathbf{p}, \mathbf{v_j}) > 0)},$$

where $I(A)$ is the indicator for the event A.

Step IV Replace $\mathbf{p}$ with the convex combination $\mathbf{p}' \equiv (1 - \lambda)\mathbf{p} + \lambda \mathbf{v_{move}}$ that maximizes $l(\cdot|\mathbf{x}, \mathbf{n})$. This convex combination can be found via bisection by determining the value $\lambda \in (0, 1]$ such that $D((1 - \lambda)\mathbf{p} + \lambda \mathbf{v_{move}}, \mathbf{v_{move}}) = 0$.

Step V If the distance from the new value for $\mathbf{p}$ to the previous value for $\mathbf{p}$ is sufficiently small, then stop, taking the new $\mathbf{p}$ as the maximizer. Otherwise, return to Step II.

Procedure 1 is used inside of Procedure 2, which is a method for finding the overall CDF value r for which $\max_{\mathbf{p} \in K_r} l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ is largest. Procedure 2 takes advantage of the fact that since $l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ is a concave function of $\mathbf{p}$, the function $r \to \max_{\mathbf{p} \in K_r} l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ is a univariate concave function of r. Once we obtain the value r for which $\max_{\mathbf{p} \in K_r} l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ is maximal, we find the restricted estimate $\arg\max_{\mathbf{p} \in K} l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ as $\arg\max_{\mathbf{p} \in K_r} l(\mathbf{p}|\mathbf{x}, \mathbf{n})$.

Procedure 2: Find the value of r that maximizes $\max_{\mathbf{p} \in K_r} l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ using the following steps.

Step I Set $c_1 = 0$, $c_5 = 1$, and $c_3 = 1/2$. Compute $l_3 = \max_{\mathbf{p} \in K_{c_3}} l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ using Procedure 1.

Step II Set $c_2 = (c_1 + c_3)/2$ and $c_4 = (c_3 + c_5)/2$. Compute $l_2 = \max_{\mathbf{p} \in K_{c_2}} l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ and $l_4 = \max_{\mathbf{p} \in K_{c_4}} l(\mathbf{p}|\mathbf{x}, \mathbf{n})$ using Procedure 1.

Step III Let i be the index such that l_i is the maximum of $\{l_2, l_3, l_4\}$. Determine new values for c_1, c_3, c_5, and l_3 by setting $(c_1', c_3', c_5') = (c_{i-1}, c_i, c_{i+1})$ and $l_3' = l_i$.

Step IV If the distance between the new c_5 and the new c_1 is smaller than some cut-off value, say 0.0005, then stop, concluding that $r = c_3$ leads to the maximum value for $\max_{\mathbf{p} \in K_r} l(\mathbf{p}|\mathbf{x}, \mathbf{n})$. Otherwise, return to Step II.

Acknowledgments The second author's research is partially supported by National Science Foundation Grant DMS 06-050141. The authors thank two referees and an associate editor for helpful comments that have led to an improved manuscript.

References

Barnett, V., Moore, K. (1997). Best linear unbiased estimates in ranked-set sampling with particular reference to imperfect ordering. *Journal of Applied Statistics, 24*, 697–710.

Bohn, L. L., Wolfe, D. A. (1994). The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analogue of the Mann–Whitney–Wilcoxon statistic. *Journal of the American Statistical Association, 89*, 168–176.

Du, J., MacEachern, S. (2008). Judgement post-stratification for designed experiments. *Biometrics, 64*, 345–354.

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.

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

Ridout, M. S., Cobby, J. M. (1987). Ranked set sampling with non-random selection of sets and errors in ranking. *Applied Statistics, 36*, 145–152.

StatLib Datasets Archive. (2009). Bodyfat dataset. Available at http://lib.stat.cmu.edu/datasets

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.

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

Ziegler, G. M. (1995). *Lectures on polytopes*. New York: Springer.