Isotonized CDF Estimation from Judgment Post-Stratification Data with Empty Strata

Xinlei Wang, Ke Wang and Johan Lim*

Abstract

In applications that require cost efficiency, sample sizes are typically small so that the problem of empty strata may often occur in judgment post-stratification (JPS), an important variant of balanced ranked set sampling. In this paper, we consider estimation of population cumulative distribution functions (CDF) from JPS samples with empty strata. In the literature, the standard and restricted CDF estimators (Stokes and Sager 1988, Frey and Ozturk 2011) do not perform well when simply ignoring empty strata. In this paper, we show that the original isotonized estimator (Ozturk 2007) can handle empty strata automatically through two methods, MinMax and MaxMin. However, blindly using them can result in undesirable results in either tail of the CDF. We thoroughly examine MinMax and MaxMin and find interesting results about their behaviors and performance in the presence of empty strata. Motivated by these results, we propose modified isotonized estimators to improve estimation efficiency. Through simulation and empirical studies, we show that our estimators work well in different

*Xinlei Wang is Associate Professor, and Ke Wang is a Ph.D. candidate, Department of Statistical Science, Southern Methodist University, 3225 Daniel Avenue, P O Box 750332, Dallas, Texas 75275-0332, swang@smu.edu and kwang@smu.edu. Johan Lim is Associate Professor, Department of Statistics, Seoul National University, Seoul, Korea, johanlim@snu.ac.kr.

regions of the CDF, and also improve the overall performance of estimating the whole function.

Keywords: Cumulative Distribution Function; Distribution Free; Imperfect Ranking; Isotonic Regression; MinMax; MaxMin; Nonparametric; Ranked Set Sampling; Relative Efficiency; Stochastic Ordering.

1 Introduction

Judgment post-stratification (JPS) is a sampling method proposed by MacEachern, Stasny and Wolfe (2004), which can be typically described by the following procedure. First, a random sample of size n is selected, and the characteristic of interest is measured for all the n units, say $Y_1, \ldots Y_n$. Then for each unit i ($i = 1, \ldots n$), a random sample of $H - 1$ additional units is selected and compared with the unit, and a rank O_i (or ranks, if there is more than one ranker) is assigned to it by eye or some other relatively inexpensive ranking method (without actual measurement of the $H - 1$ units). Thus, a JPS sample of size n can be expressed by $\{(Y_i, O_i)\}$, where $i = 1, ..., n$ and $O_i \in \{1, ..., H\}$.

JPS is similar to ranked set sampling (RSS, Chen et al. 2006, Wolfe 2004). Both are useful in situations where Y is expensive to measure, but sampling units can be easily recruited and ranked by some means not requiring quantification. Both utilize the assigned ranks to provide auxiliary information about the measured units. Thus, they can provide improved estimators of the mean, variance and distribution functions over simple random sampling (SRS) of the same size. The difference is that the number of measured units, say n_h, in the set (or stratum) with rank h ($h = 1, \ldots H$) is random in a JPS sample, following a multinomial distribution with parameters $(n, 1/H, \ldots 1/H)$, while each n_h is typically fixed in advance in a RSS sample. When the sample size n is large, JPS (with one ranker) is asymptotically equivalent to a balanced RSS. Recent developments in this area include MacEachern et al. (2002), MacEachern et al. (2004), Wang et al. (2006), Fligner and MacEachern (2006), Balakrishnan and Li (2006), Deshpande, Frey and Ozturk (2006), Stokes, Wang and Chen (2007), Frey (2007a), Frey (2007b), Frey, Ozturk and Deshpande (2007), Ozturk (2007), Samawi and Muttlak (2007), Ozturk (2008), Li and Balakrishnan (2008), Du and MacEachern (2008), Wang, Lim and Stokes (2008), Ghosh and Tiwari (2008),Balakrishnan and Li (2008), Ghosh and Tiwari (2009), Ozturk and Balakrishnan (2009), Frey and Ozturk (2011), Chen and Lim

(2011), etc.

JPS offers several advantages over RSS. First, if the ranking information is ignored, the measured observations can be analyzed using conventional statistical methods, as they are a standard random sample. This is very useful when data are collected for multiple analysis purposes. In such cases, one could anticipate using some advanced analysis methods not yet developed for RSS. Secondly, it is possible to allow more than one ranker to provide ranking information on the same measured unit, while it is difficult to do so in RSS. Thirdly, rankers can be allowed to express uncertainty about ranks (referred to as imprecise ranking), rather than being forced into stating an exact ordering. Last, JPS might allow for a large number of ranking classes (i.e., H) in some applications where perfect ranking is realistic. Since we only need to determine the rank of each fully measured unit among its comparison group, pairwise comparison is sufficient if ranking can be done perfectly, which is easy to do even if H is large. In RSS, by contrast, we need to determine, within each set which is the one with a given rank, which could be much more difficult when H is large.

In this paper, we will consider the problem of estimating population cumulative distribution functions (CDF), denoted by $F(\cdot)$, from JPS data with empty strata. Let $F_{[h]}(y)$ be the CDF of the hth stratum. The relationship $F(y) = \sum_{h=1}^{H} F_{[h]}(y)/H$ holds (Dell and Clutter 1972) for RSS or JPS samples. So $F(y)$ can be estimated by

$$\hat{F}(y) = \frac{1}{H}\sum_{h=1}^{H} \hat{F}_{[h]}(y), \tag{1}$$

where $\hat{F}_{[h]}$ is an estimator of the in-stratum CDF $F_{[h]}$. Existing estimators of $F(y)$ differ in ways of estimating $F_{[h]}$s. Stokes and Sager (1988) considered an unbiased estimator of $F(y)$ for RSS samples, say $\hat{F}^e$, where in (1) the empirical distribution function (EDF) for the hth stratum, $\hat{F}^e_{[h]}$ is used to estimate $F_{[h]}$; that is, $\hat{F}^e_{[h]}(y) = \sum_{i=1}^{n} I(Y_i \leq y, O_i = h)/n_h$, where $I(\cdot)$ is the indicator function. Ozturk (2007) considered the stochastic ordering constraint among the in-stratum CDFs from RSS samples,

$$F_{[1]}(y) \geq \cdots \geq F_{[H]}(y), \tag{2}$$

and used the isotonized estimator $\hat{F}^i_{[h]}$ satisfying constraint (2) in equation (1), which can be easily computed using isotonic regression methods (see Section 2). The corresponding estimator of $F(y)$ is denoted by $\hat{F}^i$. Both estimators can be extended to JPS easily. More recently, Frey and Ozturk (2011) considered constrained estimators of $F_{[h]}$, say $\hat{F}^r_{[h]}$, in (1) for JPS samples, where, in a certain sense, the in-stratum CDFs cannot be more extreme than the CDFs for order statistics from the overall distribution. Their estimator of $F(y)$ is denoted by $\hat{F}^r$.

For RSS or JPS applications, the sample size n can be very small because Y is expensive to measure. Since n_hs are random in JPS samples, empty strata can often occur when n is small and/or H is large so that $F_{[h]}$ cannot be directly estimated. Such problems do not arise in RSS. In fact, the probability that a JPS sample has at least one empty stratum is a function of H and n, given by

$$\sum_{i=1}^{H-1} (-1)^{i-1} \begin{pmatrix} H \\ i \end{pmatrix} \left(\frac{H-i}{H} \right)^n .$$

For example, for $H = 3$ and $n = 6$, the probability is about 0.26 (see Web Table 1 for more cases). When used with JPS data, both $\hat{F}^e$ and $\hat{F}^r$ ignore empty strata if any and the average is taken over all the non-empty ones, namely,

$$\hat{F}(y) = \frac{1}{\sum_{h=1}^{H} I(n_h > 0)} \sum_{n_h > 0} \hat{F}_{[h]}(y). \tag{3}$$

By contrast, $\hat{F}^i$ can handle empty cells by pooling over adjacent cells, as will be discussed in Section 2.

In this paper, we focus on JPS data with empty strata. In Section 2, methods for computing $\hat{F}^i$ (i.e., MinMax and MaxMin) are discussed; and in the presence of empty strata, the behaviors and performance of the two methods are examined in detail. In Section 3, we propose three modified isotonized estimators of $F(y)$ to deal with empty strata; the first one is to improve estimation efficiency on the two tails of $F(y)$, and the others are to improve efficiency in the middle part of $F(y)$. Section 4 compares the performance of the proposed estimators with the existing ones, $\hat{F}^e$ and $\hat{F}^r$, for data with different numbers of rank classes and data with ranking errors. The overall performance of the CDF estimators is also examined and compared. We provide a data example in Section 5, in which information from two rankers that used unknown "natural" ranking mechanisms is available. A discussion follows in Section 6.

2 Behaviors of MinMax and MaxMin Methods

The isotonized estimator $\hat{F}^i(y)$ is motivated by the observation that the distributions from different (judgment) rank classes are often stochastically ordered, even if ranking is imperfect. As mentioned in Wang et al. (2008), such ordering holds for a wide range of practical ranking mechanisms. Because of the inherent variability of the observations, (2) can be violated by the in-stratum EDFs, especially when the total sample size is small. Reduction in mean square errors (MSE) can be obtained by imposing the order constraint on the EDFs (Ozturk 2007).

To construct $\hat{F}^i(y)$ from (1), we first need to compute the in-stratum estimates $\{\hat{F}^i_{[h]}(y)\}$ for $h = 1, \ldots, H$, i.e., the isotonic regression estimator of $\{\hat{F}^e_{[h]}(y)\}$ with weights $(n_h)_{h=1}^H$. That is, given y, $\{\hat{F}^i_{[h]}\}$ minimizes the weighted least square $\sum_{h=1}^H \left\{\hat{F}^e_{[h]}(y) - F_{[h]}(y)\right\}^2 n_h$ under the constraint (2). Further, there exist two analytical forms available for solving the optimization problem,

$$\hat{F}^{i-}_{[h]}(y) = \min_{r \leq h} \max_{s \geq h} \sum_{g=r}^{s} \frac{n_g \hat{F}^e_{[g]}(y)}{n_{rs}} \tag{4}$$

or

$$\hat{F}^{i+}_{[h]}(y) = \max_{s\geq h} \min_{r\leq h} \sum_{g=r}^{s} \frac{n_g \hat{F}^{e}_{[g]}(y)}{n_{rs}} \tag{5}$$

where $n_{rs} = \sum_{g=r}^{s} n_g$. When there is no empty stratum, the above two formulas are equivalent (Robertson and Waltman 1968), which can be implemented by the pool adjacent violators algorithm (PAVA).

For data with empty cells, expressions (4) and (5) need to deal with the case that a (pooled) stratum could be empty (e.g., if the hth stratum is empty, then $n_{rs} = 0$ if $r = s = h$). The default choice is to ignore the empty ones whenever taking max or min. To be mathematically rigorous, we make it explicit by first introducing the index set

$$\mathcal{I}(u) = \left\{ v \mid \sum_{k=\min(u,v)}^{\max(u,v)} n_k > 0, \quad 1 \leq v \leq H \right\}$$

and then re-writing (4) and (5) into

$$\hat{F}^{i-}_{[h]}(y) = \min_{r\leq h, r\in\mathcal{I}(H)} \max_{s\geq h, s\in\mathcal{I}(r)} \sum_{g=r}^{s} \frac{n_g \hat{F}^{e}_{[g]}(y)}{n_{rs}} \tag{6}$$

and

$$\hat{F}^{i+}_{[h]}(y) = \max_{s\geq h, s\in\mathcal{I}(1)} \min_{r\leq h, r\in\mathcal{I}(s)} \sum_{g=r}^{s} \frac{n_g \hat{F}^{e}_{[g]}(y)}{n_{rs}}, \tag{7}$$

which reduce to (4) and (5), respectively, if data do not have empty cells. In this paper, we refer to (6) or $\hat{F}^{i}(y)$ constructed from (6) as MinMax, and (7) or $\hat{F}^{i}(y)$ constructed from (7) as MaxMin. In the presence of empty cells, they are no longer equivalent and perform very differently. The following two propositions explain their behaviors when they handle (non)empty cells (see Web Appendices A & B for proof).

Proposition 1. *For an empty stratum not located at the boundary (i.e., a stratum with at least one non-empty stratum at each side), MaxMin uses the estimate of the nearest non-*

empty one on its left to impute it while MinMax uses that of the nearest non-empty one on its right to do so.

Proposition 2. *MinMax and MaxMin give the same estimate for any non-empty stratum or an empty stratum located at the boundary (i.e., a stratum to whose left/right strata are all empty).*

Note that Propositions 1 and 2 provide an efficient algorithm for computing (6) and (7). That is, first compute the in-stratum CDF estimates for all non-empty cells using PAVA, and fill in the empty cells at the boundary using the nearest value available; to fill in those empty cells not located at the boundary, we use the nearest value to the right for MinMax while we use the nearest value to the left for MaxMin. Based on the above, we can conjecture that MinMax works well on the left tail of $F(y)$. The reason is that MinMax imputes the CDF of a non-boundary empty stratum $F_{[h]}$ by that of the neighbor on its right, which is always smaller than or equal to $F_{[h]}$. This may enhance the performance when $F(y)$ is close to 0, but deteriorate it when $F(y)$ is close to 1. Similarly, MaxMin is expected to work well on the right tail of $F(y)$ but poorly on the left tail.

To confirm the conjecture, we conducted a simulation study, in which JPS samples (of different sizes) were generated from the standard normal distribution. To generate a JPS sample of size n with H rank classes, we first simulated an $n \times H$ matrix with all values from $N(0, 1)$ independently; for each row, we randomly selected one entry with probability $1/H$ to enter the JPS sample, along with its rank among the H values of the row. The number of rank classes H was fixed at 5 and the average sample size $\bar{n}$ (i.e., $\bar{n} \equiv n/H$) was set to 2, 3, 4, 5 and 6, respectively.

Let $y = F^{-1}(p)$, where $p \in [0, 1]$. The relative efficiency (RE) of the isotonized estimator

$\hat{F}^i$ over the standard estimator $\hat{F}^e$ in estimating $F(y)$ is defined by

$$RE(p) = \frac{MSE\left\{\hat{F}^e\left(F^{-1}(p)\right)\right\}}{MSE\left\{\hat{F}^i\left(F^{-1}(p)\right)\right\}},$$

where $\hat{F}^i$ can be $\hat{F}^{i-}$ constructed from $\hat{F}^{i-}_{[h]}$ (MinMax), or $\hat{F}^{i+}$ constructed from $\hat{F}^{i+}_{[h]}$ (MaxMin).

Proposition 3. *Given p, both $MSE\left\{\hat{F}^{i-}\left(F^{-1}(p)\right)\right\}$ and $MSE\left\{\hat{F}^{i+}\left(F^{-1}(p)\right)\right\}$ are distribution-free for JPS samples under perfect ranking.*

Proposition 3 (see Appendix for proof), combined with that $MSE\left\{\hat{F}^e\left(F^{-1}(p)\right)\right\}$ is distribution-free (Stokes and Sager 1988), leads to the conclusion that $RE(p)$ is distribution-free for both MinMax and MaxMin under perfect ranking.

We compare the performance of $\hat{F}^{i-}$ and $\hat{F}^{i+}$ based on $RE(p)$, reported in Figure 1 for $p = 0.2$, 0.5 and 0.8 and in Web Figure 1 for $p = 0.1, 0.2, \ldots 0.9$, where "-" represents $\hat{F}^{i-}$ (MinMax) and "+" represents $\hat{F}^{i+}$ (MaxMin). For each setting, RE is estimated from 10,000 replicates. The figures show clear patterns as expected. When p is close to zero, MinMax preforms very well; and MaxMin performs poorly and can be even worse than the standard one $\hat{F}^e$. The opposite occurs when p is close to 1. In the middle part (p around 0.5), they perform equally well and is more efficient than $\hat{F}^e$. Similar observations can be made for different values of H based on simulation results not reported here.

The above results were based on all the JPS samples generated, including many without any empty strata in which MinMax and MaxMin are strictly equivalent. The observed difference in performance comes from the samples with at least one empty stratum. To see the difference more clearly, Figure 2 reports results for $p = 0.2$, 0.5 and 0.8 (and Web Figure 2 for $p = 0.1, 0.2, \cdots, 0.9$), based on data with at least one empty stratum only; that is, when we generated a JPS sample and found that it had no empty stratum, we discarded it and regenerated the sample until we got a sample with at least one empty stratum. A better way of generating such data will be discussed in Section 6. (Web) Figure 2 shows similar patterns as in (Web) Figure 1 (i.e., MinMax/MaxMin performs well on the left/right tail and poorly on the other end), but with much larger average improvement over the standard estimator $\hat{F}^e$.

In practice, once a JPS sample is collected, the information about whether there exist any empty strata becomes available so that one knows whether he needs to handle the problem or not. Thus, in any specific application of our methods introduced in Section 3, the performance on JPS data without empty strata is irrelevant. Throughout the rest of this paper, we report numerical results based on JPS data with at least one empty stratum; and ranking is perfect unless specified. Due to the space limit, we only report RE for $p = 0.2$, 0.5 and 0.8 in all the figures of the paper but report RE for $p = 0.1, 0.2, \cdots, 0.9$ in web figures of the supplemental materials.

In summary, all the results above show that when empty strata occur, blindly using MinMax or MaxMin can result in undesirable performance on one of the tails of $F(y)$.

3 Modified Isotonized Estimators

3.1 The median threshold estimator

We have shown by simulation that if $y < F^{-1}(0.5)$, $\hat{F}^{i-}(y)$ is more efficient, and for $y > F^{-1}(0.5)$, $\hat{F}^{i+}(y)$ is more efficient. Motivated by this interesting observation, we consider the following median threshold estimator to combine MinMax and MaxMin to improve the efficiency on both tails, namely

$$\hat{F}_1(y) = \begin{cases} \hat{F}^{i-}(y), & if\ y \leq F^{-1}(0.5), \\ \hat{F}^{i+}(y), & if\ y > F^{-1}(0.5). \end{cases} \tag{8}$$

Obviously, $\hat{F}_1(y)$ is equivalent to $\hat{F}^i(y)$ for JPS data without any empty stratum. For data with empty strata, (8) suggests that when estimating the in-stratum CDFs, we use $\hat{F}^{i-}_{[h]}(y)$ in (6) if $y \leq F^{-1}(0.5)$, which is equivalent to using the nearest value to the right (where possible) for imputation of an empty cell; and we use $\hat{F}^{i+}_{[h]}(y)$ in (7) if $y > F^{-1}(0.5)$, which is equivalent to using the nearest value to the left (where possible) for imputation.

Proposition 4. *Given p, $MSE\left\{\hat{F}_1\left(F^{-1}(p)\right)\right\}$ is distribution-free for JPS samples under perfect ranking.*

The proof of the above result can be found again in Appendix. Since the true median $F^{-1}(0.5)$ is often unknown in practice, it needs to be estimated from data. Here, we use the regular sample median $\hat{y}_{med}$ for simplicity. We also tried the JPS median estimator that is given by $\inf\{y : \hat{F}^e(y) \geq 0.5\}$ and found the performance was similar to (or even slightly worse than) that using $\hat{y}_{med}$.

Figure 2 and Web Figure 2 further compare simulated $RE(p)$ of the above combined estimator $\hat{F}_1$ ("•") with those of MinMax ("-"), and MaxMin ("+") under the same simulation settings in Section 2. Clearly, $\hat{F}_1$ mimics the good performance of MinMax on the left tail and that of MaxMin on the right tail. As will be shown later, by combining MinMax and MaxMin through $\hat{F}_1$, the overall estimation efficiency of the whole CDF is improved, too.

3.2 Alternative isotonized estimators

Though it has excellent performance in both ends, the median threshold estimator $\hat{F}_1$ is worse than both MinMax and MaxMin in the middle part. This is because the sample median, rather than the true median, is used as the threshold, according to an unreported simulation study. The performance is the worst at $p = 0.5$. The following estimators are constructed to improve the performance around $p = 0.5$.

Suppose the hth stratum is empty and not located at the boundary. Here, the basic idea is to find an in-stratum estimator $\hat{F}^*_{[h]}(y)$ to fill the "gap" between $\hat{F}^i_{[h-]}(y)$ and $\hat{F}^i_{[h+]}$, where the stratum h_-/h_+ is the nearest nonempty one on its left/right. Let t index the nonempty strata. We want the "filler" to be close to the isotonized estimates $\hat{F}^i_{[t]}$'s from the nonempty strata. Thus, we set

$$\hat{F}^*_{[h]}(y) = \underset{x}{arg\min} \sum_t \left\{ \hat{F}^i_{[t]}(y) - x \right\}^2$$

satisfying the constraint $\hat{F}^i_{[h-]}(y) \geq x \geq \hat{F}^i_{[h+]}(y)$. This leads to

$$\hat{F}^*_{[h]}(y) = \begin{cases} \bar{F}^i(y), & \text{if } \hat{F}^i_{[h-]}(y) \geq \bar{F}^i(y) \geq \hat{F}^i_{[h+]}(y) \\ \hat{F}^i_{[h-]}(y), & \text{if } \hat{F}^i_{[h-]}(y) \leq \bar{F}^i(y) \\ \hat{F}^i_{[h+]}(y) & \text{if } \bar{F}^i(y) \leq \hat{F}^i_{[h+]}(y) \end{cases} \tag{9}$$

where $\bar{F}^i(y)$ is the average of $\hat{F}^i_{[t]}(y)$'s over the nonempty strata. The estimator (1) using $\hat{F}^*_{[h]}$ to estimate $F_{[h]}$ is denoted by $\hat{F}_2$.

Another estimator that would be natural to consider for better performance in the middle is to use the average of MinMax and MaxMin, namely,

$$\hat{F}_3(y) = \frac{\hat{F}^{i-}(y) + \hat{F}^{i+}(y)}{2}. \tag{10}$$

This is equivalent to estimating $F_{[h]}$ by $(\hat{F}^i_{[h-]} + \hat{F}^i_{[h+]})/2$ for any empty stratum not located at the boundary, while for a stratum at the boundary, it is equivalent to estimating $F_{[h]}$ by that of the nearest nonempty one.

Proposition 5. *Given p, both $MSE\left\{\hat{F}_2\left(F^{-1}(p)\right)\right\}$ and $MSE\left\{\hat{F}_3\left(F^{-1}(p)\right)\right\}$ are distribution-free for JPS samples under perfect ranking.*

See Appendix for proof. Again, we can use (Web) Figure 2 to compare simulated $RE(p)$ of the three modified isotonized estimators $\hat{F}_1$ (represented by "•"), $\hat{F}_2$ (represented by "*"), and $\hat{F}_3$ (represented by "△"), as well as $\hat{F}^r$ (represented by "#", Frey and Ozturk 2011), over the standard estimator $\hat{F}^e$ under perfect ranking. When p is close to 0.5, both $\hat{F}_2$ and $\hat{F}_3$ indeed improve a lot over $\hat{F}_1$, and $\hat{F}_3$ is clearly the best among all. When p is close to 0 or 1, $\hat{F}_1$ is clearly the best and $\hat{F}_2$ is better than $\hat{F}_3$. All the new estimators $\hat{F}_1$, $\hat{F}_2$ and $\hat{F}_3$ greatly outperform $\hat{F}^r$ and $\hat{F}^e$ everywhere. Note that $\hat{F}^r$ and $\hat{F}^e$ do not assume the stochastic ordering (2) so that they are not expected to do well under perfect ranking. They might do somewhat better under incorrect ranking that violates the ordering. Also, as mentioned in the introduction, both $\hat{F}^r$ and $\hat{F}^e$ take the average over non-empty strata, which is equivalent to imputing any empty stratum by the average of non-empty strata. Doing so might make the in-stratum CDF estimates violate the stochastic ordering or the constraint originally imposed, which could degrade their performance.

4 Simulation

4.1 Comparison of $RE(p)$ for different H

We examined the performance of the proposed estimators based on $RE(p)$ for different numbers of rank classes. For $\hat{F}_1$, $\hat{F}_2$ and $\hat{F}_3$, we considered $H = 2, 4, 6, 8, 10$ and fixed the average sample size $\bar{n}$ at 3. We estimated $RE(p)$ based on 10,000 replicates under each setting. For $\hat{F}^r$, since the algorithm is slow, we only considered $H = 2, 4, 6$ with $\bar{n} = 3$; and for each, we used 2500 replicates instead. (Web) Figure 3 shows that among the three new estimators, $\hat{F}_1$ is the best on the tails, $\hat{F}_3$ is the best in the middle part, and $\hat{F}_2$ is somewhere between $\hat{F}_1$ and $\hat{F}_3$. They all perform much better than $\hat{F}^r$ for every p, which is slightly better than $\hat{F}^e$. From Web Figure 3, it appears that when p is around 0.5, REs of $\hat{F}_1$, $\hat{F}_2$ and $\hat{F}_3$, increase as H increases for the fixed average sample size. However, for all the estimators, RE is not a monotone function of H for small or large values of p. In addition, it is interesting to observe that for $H = 2$, all the estimators are the same as the standard one $\hat{F}^e$ when there is an empty stratum. This is because there is only one choice available for imputation.

4.2 Comparison of $MISE$ for different distributions

To compare the overall performance of the CDF estimators, we considered the mean integrated squared error ($MISE$), which is defined as $E\left[\int_{-\infty}^{\infty}\left\{\hat{F}(y)-F(y)\right\}^2 dy\right]$. In each iteration, we compute $\int_{-\infty}^{\infty}\left\{\hat{F}(y)-F(y)\right\}^2 dy$. Then by taking the average over multiple iterations, we obtain simulated $MISE$. The relative efficiency of $\hat{F}$ over $\hat{F}^e$ is now defined as the ratio of $MISE$s, i.e., $MISE(\hat{F}^e)/MISE(\hat{F})$. Although the performance of the estimators at any single point $F^{-1}(p)$ is distribution free, $MISE$ does depend on the population distribution. This is because

$$\int_{-\infty}^{\infty}\left\{\hat{F}(y)-F(y)\right\}^2 dy = \int_0^1\left\{\hat{F}\left(F^{-1}(p)\right)-p\right\}^2 dF^{-1}(p) = \int_0^1\left\{\hat{F}\left(F^{-1}(p)\right)-p\right\}^2 \frac{1}{f(F^{-1}(p))}dp.$$

We generated JPS data with at least one empty stratum from different distributions including $Unif(0,1)$, $N(0,1)$, $Exp(1)$, $Beta(0.5,0.5)$. Here, $H = 5$ and $\bar{n} = 3$. The number of iterations was 10,000 for $\hat{F}_1$, $\hat{F}_2$ and $\hat{F}_3$, and 2,500 for $\hat{F}^r$. The left panel of Table 1 shows that in terms of the overall performance in estimating $F(y)$, $\hat{F}_3 > \hat{F}_2 > \hat{F}_1 > \hat{F}^r > \hat{F}^e$ for all the four distributions considered.

4.3 Comparison of $RE(p)$ under imperfect ranking

In practice, ranking errors might often occur. To study the robustness of the CDF estimators under imperfect ranking, we conducted two simulation studies. In the first study, we considered Dell and Clutter's ranking error model, $X_i = Y_i + \epsilon_i$, where ϵ_i is the error term with mean zero; and actual measurements are taken on Y_is, but ranking is done based on X_is. We generated $Y_i \overset{iid}{\sim} N(0,1)$ and $\epsilon_i \overset{iid}{\sim} N(0,0.78)$ so that the correlation between X and Y was about 0.75. Again, we set $H = 5$ and $\bar{n} = 2,3,4,5,6$, and generated JPS samples with at least one empty stratum. In (Web) Figure 4, all the estimators demonstrate similar patterns to those observed under perfect ranking, but the differences become smaller here. The modified estimators $\hat{F}_1$, $\hat{F}_2$ and $\hat{F}_3$ are still consistently better than $\hat{F}^r$ and $\hat{F}^e$ for all the p values, though the improvement is not as big as that without ranking errors.

In our second study, we considered JPS data with purely random ranking, an extreme case of imperfect ranking, in which the "$<$" does not hold in (2) and the in-stratum CDFs are all equal to the population CDF. As in Section 4.2, we set $H = 5$ and $\bar{n} = 3$, and report $MISE$ for $Unif(0,1)$, $N(0,1)$, $Exp(1)$, $Beta(0.5,0.5)$ in the right panel of Table 1. We can see that the overall performance follows the order $\hat{F}_2 > \hat{F}_3 \geq \hat{F}_1 > \hat{F}^r > \hat{F}^e$ for all the distributions considered, and the differences in performance become even smaller, compared to the case $\rho = 0.75$. The only difference from the order reported for perfect ranking in Section 4.2 is that here, $\hat{F}_2$ outperforms $\hat{F}_3$ and becomes the best, suggesting $\hat{F}_2$ might be more robust to ranking errors than $\hat{F}_3$.

5 An Empirical Study

This section provides an empirical comparison of the five estimators $\hat{F}_1$, $\hat{F}_2$, $\hat{F}_3$, $\hat{F}^r$ and $\hat{F}^e$ under practical situations including imperfect ranking and multiple rankers, using data in Table 2 of Wang et al. (2008) (now reproduced as Web Table 2 for easy access). The table was derived from a data set in Weisberg (1985) that consists of allometric measurements for 62 species of mammals. The variable of interest is the log of adjusted brain weight, defined as $Y = \log\{\text{brain weight}/(\text{body weight})^{2/3}\}$; and the goal of our analysis is to estimate the population CDF of Y. As mentioned in Wang et al. (2008), to produce the table, the species were randomly grouped into 20 sets, three species each (two were randomly discarded for this purpose). Within each set, ranks of Y are available from two different rankers. The rankers did not know the value of Y for each species before they assigned ranks. So they made judgments based on the conjecture that a "clever" species tends to have a large adjusted brain weight. Here, ranking is not perfect and the data do not follow any known ranking error model.

In our experiment, selection of JPS samples was simulated, where we set $H = 3$ and $\bar{n} = 2, 3, 4, 5, 6$; and for each sample, the estimates of the CDF were calculated. To generate a JPS sample of size $n = H \times \bar{n}$, n sets were randomly drawn with replacement from the total 20 sets and then one species was randomly drawn from each selected set to enter the JPS sample. In total, 10,000 samples for each setting, all with empty strata, were generated. For the purpose of evaluation, the 60 mammals were treated as the "true" population, and the EDF computed from their data was taken as the true population CDF when computing MSEs or $MISE$s, which is a step function itself.

We first considered estimation using only one ranker. Web Figure 5 reports $RE(p)$ of the three new estimators $\hat{F}_1$, $\hat{F}_2$, $\hat{F}_3$ and $\hat{F}^r$ over $\hat{F}^e$ at different p based on ranks from the second ranker. The left panel of Table 2 reports RE based on $MISE$ to examine the overall performance. We can observe that again, $\hat{F}_1$ is the best when p is close to 0 or 1; $\hat{F}_2$ and $\hat{F}_3$ are about the same, which are better than the others when p is close to 0.5. Also, note that REs for $p = 0.9$ are greater than those for all the other p values, especially for $\hat{F}_1$. This indicates that under a natural ranking process, the performance can be very different on the two tails. In terms of the overall performance, we have $\hat{F}_2 \approx \hat{F}_3 > \hat{F}_1 > \hat{F}^r > \hat{F}^e$ from Table 2. Note that when $H = 3$, $\hat{F}_2$ and $\hat{F}_3$ are identical. The slight difference we observe here is due to drawing separate samples for the two estimators.

In spirit of Wang et al. (2008), we considered estimation using two rankers. Let O_{ij} be the judgment order of Y_i assigned by ranker j among its own set of unmeasured units, for

$j = 1, 2$. We first transposed a JPS sample, denoted by $\mathcal{D} = [(Y_{i,} O_{i1}, O_{i2})]_{i=1}^{n}$, to

$$\tilde{\mathcal{D}} = [(Y_1, O_{11}), (Y_1, O_{12}), (Y_2, O_{21}), (Y_2, O_{22}), ..., (Y_n, O_{n1}), (Y_n, O_{n2})]^T,$$

where each Y_i value was replicated twice; then we used the CDF estimators with $\tilde{\mathcal{D}}$ as if there were only one ranker but $2n$ observations. Web Figure 6 and the right panel of Table 2 give results. Obviously, conclusions are consistent with those from the case of one ranker. In addition, when compared with the case of one ranker, it appears the improvement from $\hat{F}_2$ or $\hat{F}_3$ increases while that from $\hat{F}^r$ decreases.

To summarize, this example establishes that the new estimators for JPS data with empty strata can provide an advantage even when the error-prone ranking process is a "natural" one that does not follow a known model.

6 Discussion

Like ranked set sampling, judgment post-stratification is useful in applications requiring cost efficiency. Such applications typically have small sample sizes. Dealing with empty strata when n is small or H is relatively large is an important issue. In this paper, the behaviors and performance of the two methods, MinMax and MaxMin, associated with the isotonized CDF estimator $\hat{F}^i$ proposed by Ozturk (2007), have been carefully examined in the presence of empty strata. It has been found that MinMax works well on the left tail, MaxMin works well on the right tail and they are about the same in the middle part of $F(y)$. Motivated by this observation, three modified isotonized estimators $\hat{F}_1$, $\hat{F}_2$ and $\hat{F}_3$, have been proposed to combine the strength of MinMax and MaxMin. Through simulation studies and a data example, we have shown that all the three estimators can achieve better estimation performance over the existing estimators $\hat{F}^r$ and $\hat{F}^e$ in different regions of $F(y)$, which also improve the overall performance in estimating the whole function. Among the three, it has been consistently observed that $\hat{F}_1$ has the best performance on both tails of the CDF, and $\hat{F}_3$ is the best in the middle part under perfect ranking. In the presence of ranking errors, $\hat{F}_2$ appears to be more robust than $\hat{F}_3$.

In our numerical studies, we generated JPS samples with at least one empty stratum by simply discarding samples without any empty cell. This can be very inefficient since the probability of having at least one empty cell is very small sometimes. As one of the referees suggested, the following procedure is much faster, given the sample size n and the number of rank classes H: (1) select a random number h from the uniform distribution on $\{1, \ldots, , H\}$ and let $n_h = 0$; (2) generate $(n_1, \ldots, n_{h-1}, n_{h+1}, \ldots n_H)$ from the multinomial distribution with parameters n and $(1/(H-1), \ldots, 1/(H-1))$; (3) accept the vector $(n_1, \ldots, n_H)$ with probability $1/c$, where c is the number of zeros in the vector $(n_1, \ldots, n_H)$; if accepted, go to step (4), otherwise go to step (1); (4) generate an $n \times H$ matrix with all values from $N(0, 1)$ (or any other distribution) independently; in each row, put all the H values in an increasing order; then we select n_h values from the hth column for $h = 1, \ldots, H$, one value from each row, to enter the JPS sample.

Our paper presents the first but important attempt at dealing with the issue of empty strata for JPS samples by providing three easy-to-construct CDF estimators. There is ample space in estimating population CDFs for future research. For example, there are various potential ways to build in directional sensitivity of the two tails, such as using some smooth transition from MinMax to MaxMin rather than the jumpy transition used in $\hat{F}_1$. One could also use the Rao-Blackwell Theorem with $\hat{F}_1$ for possible improvement through averaging. In the case that a sample is highly unbalanced, one might want to try other versions of median estimators in $\hat{F}_1$ such as weighted ones, to avoid bias. Further, instead of assigning equal weights to MinMax and MaxMin as in $\hat{F}_3$, one could create weights for the order statistics of the sample, and then use these weights to construct a weighted average of MinMax and

MaxMin.

Finally, we note that under the context of JPS data with empty strata, the estimation of the in-stratum CDFs is important itself and would be useful to reduce the impact of ranking errors on statistical procedures. It would be interesting to investigate it in depth in the future. In addition, quantile estimation could be a natural topic to consider next, too.

7 Supplementary Materials

Web Appendices, Tables and Figures referenced in Sections 1-5 are available under the Paper Information link at the Biometrics website http://www.biometrics.tibs.org.

Acknowledgement

We are grateful to the reviewers for their thoughtful and constructive suggestions, which have greatly improved the manuscript.

Appendix: Proof of Propositions 3-5

1. We first show that for JPS data, under perfect ranking, the mean squared error $MSE\left\{\hat{F}\left(F^{-1}(p)\right)\right\}$ is distribution free if $\hat{F}$ is in the form of $g(\{\hat{F}^e_{[h]}\}_{h=1}^H, \{n_h\}_{h=1}^H)$, where g is a certain function.

   Let $t = F^{-1}(p)$, and $Y_{[h]i}$ be the ith observation in the hth stratum, $i = 1, \ldots, n_h$. The EDF of the non-empty hth stratum is

$$\hat{F}^e_{[h]}(t) = \frac{1}{n_h} \sum_{i=1}^{n_h} I(Y_{[h]i} \leq t).$$

Note that $I(Y_{[h]i} \le t) = I(Y_{[h]i} \le F^{-1}(p)) \sim \text{Bernoulli}(p_h)$, where

$$p_h = \sum_{j=h}^{H} \begin{pmatrix} H \\ j \end{pmatrix} p^j (1-p)^{H-j}$$

under the perfect ranking. Thus, $n_h \hat{F}^e_{[h]}(t) \sim \text{binomial}(n_h, p_h)$.

If $\hat{F} = g(\{\hat{F}^e_{[h]}\}_{h=1}^H, \{n_h\}_{h=1}^H)$, then

$$\begin{aligned} MSE\left\{\hat{F}(t)\right\} &= E\left\{\hat{F}(t) - p\right\}^2 = E\left\{g\left(\{\hat{F}^e_{[h]}(t)\}_{h=1}^H, \{n_h\}_{h=1}^H\right) - p\right\}^2 \\ &= E\left[E\left\{g\left(\{\hat{F}^e_{[h]}(t)\}_{h=1}^H, \{n_h\}_{h=1}^H\right) - p\right\}^2 \mid \{n_h\}_{h=1}^H\right] \end{aligned}$$

Given $\{n_h\}_{h=1}^H$, $\hat{F}^e_{[h]}(t)$'s are independent. So the inner expectation given $\{n_h\}_{h=1}^H$ is a function of $\{p_h\}_{i=1}^h$, $\{n_h\}_{h=1}^H$ and p . Since all p_hs are functions of p, we have

$$MSE\left\{\hat{F}(t)\right\} = E\left\{g_0\left(p, \{n_h\}_{h=1}^H\right) \mid \{n_h\}_{h=1}^H\right\}.$$

Further, under perfect ranking, $(n_1, \cdots, n_H) \sim \text{multinomial}(n, 1/H, \cdots, 1/H)$. Thus $MSE\left\{\hat{F}(t)\right\}$ is a function of p and n for a given H, which is distribution free.

2. It is easy to see from equations (6)-(10), MaxMin, MinMax, $\hat{F}_1$ (assuming that the true median is known) $\hat{F}_2$ and $\hat{F}_3$ are all in the form of $G(\{\hat{F}_{[h]}\}_{h=1}^H, \{n_h\}_{h=1}^H)$. Thus, their$MSE$s are all distribution free under perfect ranking.

References

Balakrishnan, N. and Li, T. (2006). Confidence intervals for quantiles and tolerance intervals based on ordered ranked set samples. *Annals of the Institute of Statistical Mathematics*, 58:757–777.

Balakrishnan, N. and Li, T. (2008). Ordered ranked set samples and application to inference. *Journal of Statistical Planning and Inference*, 138:3512–3524.

Chen, M. and Lim, J. (2011). Estimating variances of strata in ranked set sampling. *Journal of Statistical Planning and Inference*, In Print.

Chen, Z., Bai, Z., and Sinha, B. K. (2006). *Ranked Set Sampling: Theory and Applications.* Springer.

Dell, T. and Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 28(2):545–555.

Deshpande, J. V., Frey, J., and Ozturk, O. (2006). Nonparametric ranked-set sampling confidence intervals for quantiles of a finite population. *Environmental and Ecological Statistics*, 13:25–40.

Du, J. and MacEachern, S. N. (2008). Judgement post-stratification for designed experiments. *Biometrics*, 64(2):345–354.

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

Frey, J. (2007a). Distribution-free statistical intervals via ranked-set sampling. *Canadian Journal of Statistics*, 35:585–596.

Frey, J. (2007b). New imperfect rankings models for ranked set sampling. *Journal of Statistical Planning and Inference*, 137:1433–1445.

Frey, J. and Ozturk, O. (2011). Constrained estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics*, to appear.

Frey, J., Ozturk, O., and Deshpande, J. V. (2007). Nonparametric tests for perfect judgment rankings. *Journal of the American Statistical Association*, 102:708–717.

Ghosh, K. and Tiwari, R. C. (2008). Estimating the distribution function using k-tuple ranked set samples. *Journal of Statistical Planning and Inference*, 138:929–949.

Ghosh, K. and Tiwari, R. C. (2009). A unified approach to variations of ranked set sampling with applications. *Journal of Nonparametric Statistics*, 21:471–485.

Li, T. and Balakrishnan, N. (2008). Some simple nonparametric methods to test for perfect ranking in ranked set sampling. *Journal of Statistical Planning and Inference*, 138:1325–1338.

MacEachern, S. N., Ozturk, O., Wolfe, D. A., and Stark, G. (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., and Wolfe, D. A. (2004). Judgment post-stratification with imprecise rankings. *Biometrics*, 60:207–215.

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.

Ozturk, O. and Balakrishnan, N. (2009). An exact control-versus-treatment comparison test based on ranked set samples. *Biometrics*, 65:1213–1222.

Robertson, T. and Waltman, P. (1968). On estimating monotone parameters. *The Annals of Mathematical Statistics*, 39:1030–1039.

Samawi, H. M. and Muttlak, H. A. (2007). Estimation of ratio using rank set sampling. *Biometrical Journal*, 38:753–764.

Stokes, S. L. and 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.

Stokes, S. L., Wang, X., and Chen, M. (2007). Judgment post-stratification with multiple rankers. *Journal of Statistical Theory and Applications*, 6:344–359.

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

Wang, X., Stokes, L., Lim, J., and Chen, M. (2006). Concomitants of multivariate order statistics with application to judgment poststratification. *Journal of the American Statistical Association*, 101(476):1693–1704.

Weisberg, S. (1985). *Applied Linear Regression*. Wiley, New York.

Wolfe, D. A. (2004). Ranked set sampling: An approach to more efficient data collection. *Statistical Science*, 19:636–643.

Tables and Figures

	Perfect Ranking				Random Ranking			
Est.	$N(0,1)$	$Unif(0,1)$	$Exp(1)$	$Beta(0.5,0.5)$	$N(0,1)$	$Unif(0,1)$	$Exp(1)$	$Beta(0.5,0.5)$
$\hat{F}_1$	1.31	1.28	1.31	1.26	1.11	1.09	1.12	1.08
$\hat{F}_2$	1.34	1.36	1.33	1.37	1.12	1.12	1.13	1.12
$\hat{F}_3$	1.39	1.43	1.35	1.49	1.11	1.11	1.12	1.11
$\hat{F}^r$	1.06	1.04	1.05	1.03	1.06	1.04	1.08	1.02

Table 1: Simulated relative efficiencies (defined as ratio of $MISE$s) of the three modified isotonized estimators $\hat{F}_1$, $\hat{F}_2$, $\hat{F}_3$ and $\hat{F}^r$ (Frey and Ozturk 2011) over the standard estimator $\hat{F}^e$, for JPS data ($H = 5$, $\bar{n} = 3$) with at least one empty stratum from different distributions under perfect ranking and random ranking, respectively.

Case	2nd Ranker				Both Rankers			
AVG. Size $\bar{n}$	1	2	3	4	1	2	3	4
$\hat{F}_1$	1	1.14	1.14	1.15	1	1.12	1.13	1.20
$\hat{F}_2$	1.08	1.16	1.18	1.20	1.10	1.17	1.21	1.25
$\hat{F}_3$	1.08	1.16	1.17	1.20	1.11	1.16	1.21	1.24
$\hat{F}^r$	1.05	1.08	1.06	1.03	1.03	1.02	1.01	1.01

Table 2: An empirical study of adjusted brain weights of mammals: simulated relative efficiencies (defined as ratio of $MISE$s) of $\hat{F}_1$, $\hat{F}_2$, $\hat{F}_3$ and $\hat{F}^r$ over the standard estimator $\hat{F}^e$ are reported for each $\bar{n}$ under two cases: (1) one ranker only; and (2) two rankers, based on JPS data ($H = 3$) with at least one empty strata.

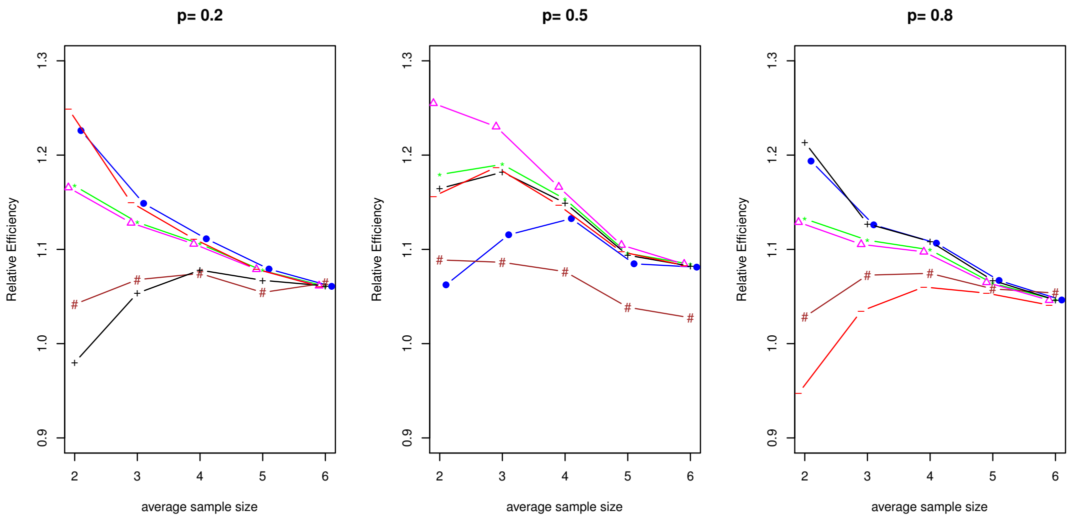

Figure 1: Simulated relative efficiencies of $\hat{F}^{i-}$ (MinMax, represented by "-"), $\hat{F}^{i+}$ (MaxMin, represented by "+"), $\hat{F}_1$ (represented by "●"), $\hat{F}_2$ (represented by "*"), $\hat{F}_3$(represented by "△"), and $\hat{F}^r$ (represented by "#", Frey and Ozturk 2011) over the standard estimator $\hat{F}^e$, as a function of average sample size (the number of rank classes H is fixed at 5), based on JPS data from perfect ranking (including those without empty strata).

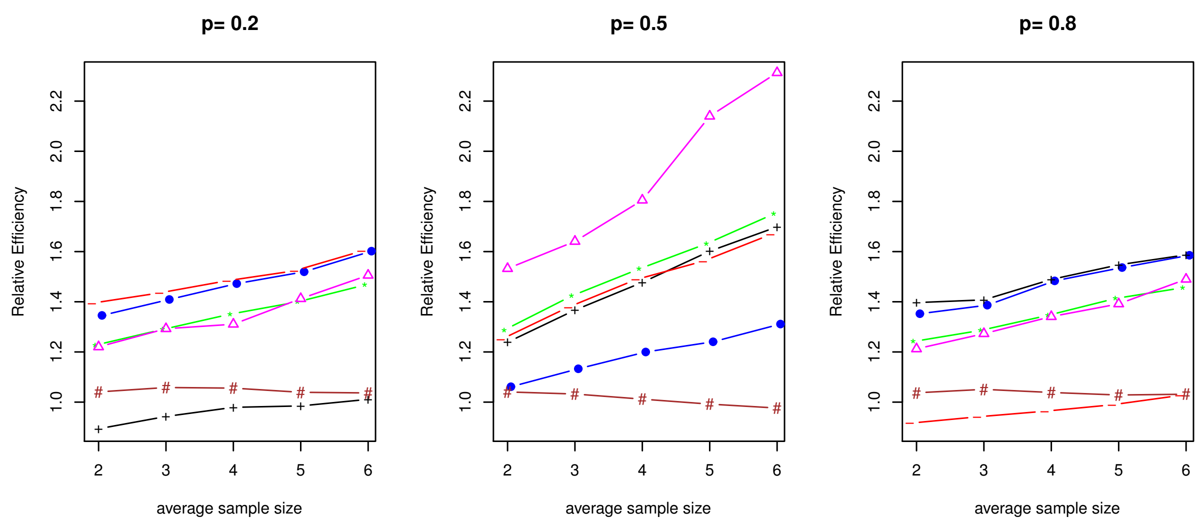

Figure 2: Simulated relative efficiencies of $\hat{F}^{i-}$ (MinMax, represented by "-"), $\hat{F}^{i+}$ (MaxMin, represented by "+"), $\hat{F}_1$ (represented by "●"), $\hat{F}_2$ (represented by "*"), $\hat{F}_3$(represented by "△"), and $\hat{F}^r$ (represented by "#", Frey and Ozturk 2011) over the standard estimator $\hat{F}^e$, as a function of average sample size (the number of rank classes H is fixed at 5), based on JPS data from perfect ranking with at least one empty stratum.

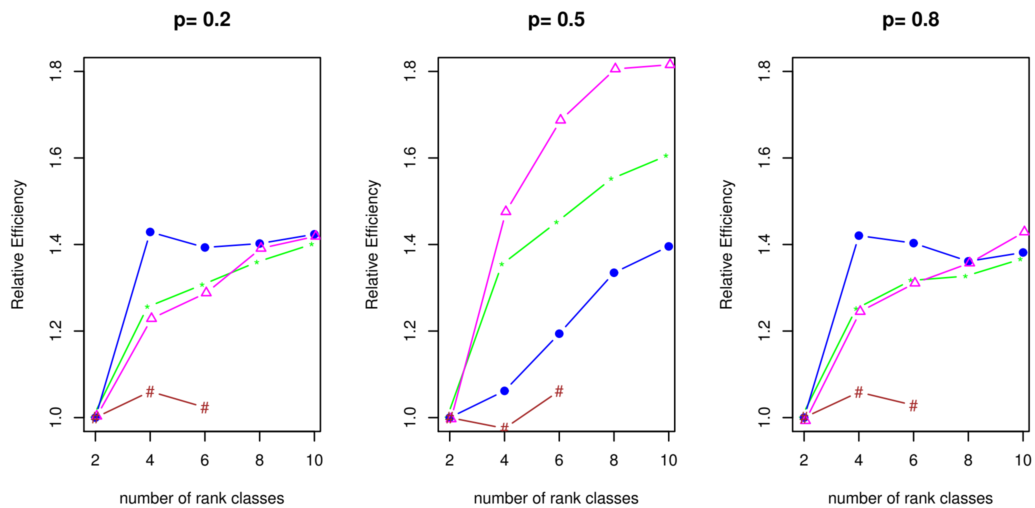

Figure 3: Simulated relative efficiencies of the three modified isotonized estimators $\hat{F}_1$ (represented by "●"), $\hat{F}_2$ (represented by "*"), $\hat{F}_3$(represented by "△"), and $\hat{F}^r$ (represented by "#", Frey and Ozturk 2011) over the standard estimator $\hat{F}^e$, as a function of number of rank classes (the average sample size $\bar{n}$ is fixed at 3), based on JPS data from perfect ranking with at least one empty stratum.

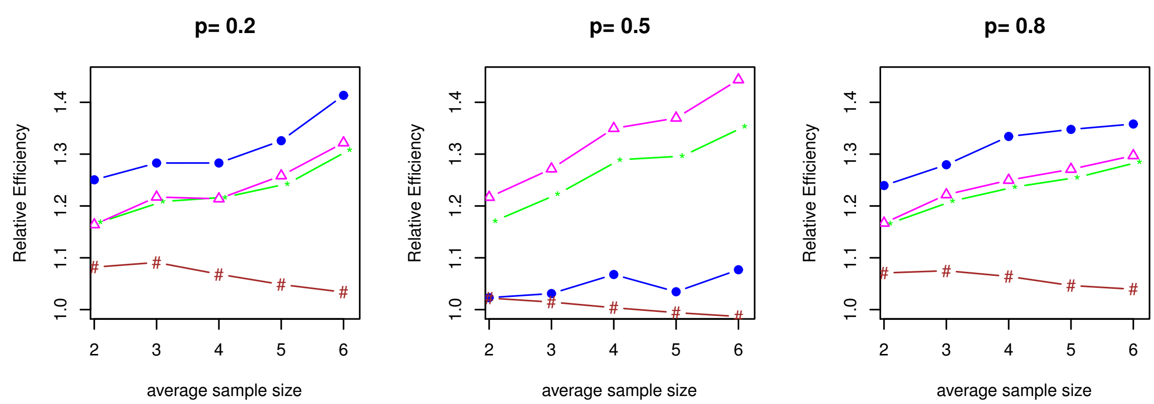

Figure 4: Simulated relative efficiencies of the three modified isotonized estimators $\hat{F}_1$ (represented by "●"), $\hat{F}_2$ (represented by "*"), $\hat{F}_3$(represented by "△"), and $\hat{F}^r$ (represented by "#", Frey and Ozturk 2011) over the standard estimator $\hat{F}^e$, as a function of average sample size (the number of rank classes H is fixed at 5), based on JPS data from imperfect ranking ($\rho = 0.75$) with at least one empty stratum.