# A Nonparametric Mean Estimator for Judgment Post-Stratified Data

Xinlei WANG, Johan LIM and Lynne STOKES*

July 2006

**Abstract**

MacEachern, Stasny and Wolfe (2004) introduced a data collection method, called judgment post-stratification (JP-S), based on ideas similar to those in ranked set sampling, and proposed methods for mean estimation from JP-S samples. In this paper we propose an improvement to their methods, which exploits the fact that the distributions of the judgment post-strata are often stochastically ordered, so as to form a mean estimator using isotonized sample means of the post-strata. This new estimator is strongly consistent with similar asymptotic properties to those in MacEachern, Stasny and Wolfe (2004). It is shown to be more efficient for small sample sizes, which appears to be attractive in applications requiring cost efficiency. Further, we extend our method to JP-S samples with imprecise ranking or multiple rankers. The performance of the proposed estimators is examined on three data examples through simulation.

*Keywords*: Imperfect Ranking; Imprecise Ranking; Isotonic Regression; Multiple Rankers; Ranked Set Sampling; Simple Stochastic Ordering.

*Xinlei Wang is Assistant Professor and Lynne Stokes is Professor, Department of Statistical Science, Southern Methodist University, 3225 Daniel Avenue, P O Box 750332, Dallas, Texas 75275-0332, swang@mail.smu.edu and slstokes@mail.smu.edu. Johan Lim is Assistant Professor, Department of Statistics, Texas A & M University, johanlim@stat.tamu.edu.

# 1 Introduction

Ranked set sampling (RSS) is an established cost efficient sampling method. It has been shown to be useful in situations where the characteristic of interest is expensive to measure, but sampling units can be easily gathered and ranked by some means not requiring quantification. Such situations arise naturally and frequently in fields of agriculture and environment sciences (Ross and Stokes, 1999). For an overview about RSS, see Chen et al. (2006) and the references therein.

Recently, judgment post-stratification (JP-S) has been proposed by MacEachern et al. (2004) as an alternative to ranked set sampling. RSS and JP-S are similar, both in practical implementation and in theoretical development. Thus, they can be applied in similar situations. Both sampling methods improve estimates of the population mean by artificially creating a stratified sample based on the (judgment) ranks of fully measured units. However, RSS relies on a special design, in which ranking is required to be done before measuring, and the number of units in each rank class is prespecified. In contrast, JP-S is based on a simple random sample (SRS), in which the experiment units are further post-stratified on ranks, and so the number of units in each rank class is a random variable. Since an underlying SRS is tractable in most statistical analyses, JP-S is attractive, for example, when data need to be collected for multiple purposes. Also, in some applications, researchers may be reluctant to beginning with a nonstandard design, but willing to beginning with an SRS with the option of using auxiliary rank information later. In addition, JP-S is flexible. It can allow rankers to express uncertainty about ranks (MacEachern et al., 2004). It can also formally incorporate ranking information from multiple auxiliary variables or rankers (Wang et al., 2006).

In this paper, we consider the problem of mean estimation based on JP-S samples. We propose new estimators that are motivated by the observation that the distributions from different judgment rank classes are stochastically ordered, even if ranking is imperfect. The order constraints in distributions yield a simple ordering in means, too. However, standard mean estimates of the judgment post-strata may not reflect the restrictions because of the inherent variability of the observations. Violations may often occur when the total sample size is small, as is the typical case in applications where accurate measurements are difficult or costly.

The literature indicates that by imposing the underlying order constraints, substantial reduction in mean square errors can be obtained (Feltz and Dykstra, 1985; Lee, 1981). Here, for JP-S data, we consider a method that explicitly takes into account the ordering via isotonic regression while estimating the means of the post-strata, and then uses the isotonized estimates to form a new estimator of the population mean.

The rest of the paper is organized as follows. In Section 2, we propose the new estimator of mean and show it is strongly consistent. Section 3 compares it with the existing one in both the asymptotic properties and small-sample behaviors. In Section 4, we further extend our method to JP-S data that allow for multiple rankers or imprecise ranking. In Section 5, three data examples are provided; the first example uses body mass data to examine the performance under the impact of imperfect ranking; the second illustrates the use of our estimator with two rankers through adjusted brain weight data; and the third illustrates the use with imprecise ranking on shelved Master's theses data.

# 2 Estimation of Mean Using Isotonic Regression

Suppose the variable of interest $Y$ is absolutely continuous with population mean $\mu$ and finite variance $\sigma^2$. A basic version of a judgment post-stratification sample is constructed as follows. First, select a simple random sample of $n$ units, on each of which the value of $Y$ is measured. For each $i$ $(1 \le i \le n)$, an additional $H-1$ units are randomly selected and the judgment order of $Y_i$ among its $H$ comparison units, denoted by $O_i$, is determined subjectively without measuring the $H-1$ units (hence ranking errors could occur). Thus, the JP-S sample consists of the data of the form $(Y_i, O_i)_{i=1}^n$, and the $n$ measured units fall into $H$ post-strata formed by the orders. Let $Y_{[h]}$ denote $Y|O = h$, any observation falling in the $h$-th post-stratum, $h = 1, \cdots H$. Let $n_h$, $\bar{Y}_{[h]}$, $\mu_{[h]}$, and $\sigma^2_{[h]}$ denote the number, sample mean, mean and variance of $Y_{[h]}$'s within the $h$-th stratum. Note that $(n_1, \cdots, n_H) \sim$ multinomial$(n, 1/H, \cdots, 1/H)$.

MacEachern et al. (2004) propose an unbiased JP-S estimator of $\mu$,

$$\hat{\mu} = \frac{1}{H}\sum_{h=1}^{H}\bar{Y}_{[h]}, \tag{1}$$

which is the average of all sample means of the $H$ post-strata. The asymptotic relative efficiency of $\hat{\mu}$ versus the estimator from balanced ranked set sampling is 1.

Assume that $Y_{[1]}, \cdots Y_{[H]}$ are stochastically ordered. That is, for any $y$,

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

where $F_{[h]}(y)$ denotes the cumulative density function within the $h$-th stratum. This assumption is true when ranking is perfect, since in this case $F_{[h]}(y)$ becomes $F_{(h)}(y)$, the distribution of the $h-$th order statistics. In the presence of ranking errors, it is true in many situations, for example, when linear ranking models (Dell and Clutter, 1972), or more generally, monotone likelihood ratio ranking models (Fligner and MacEachern, 2006) are suitable. It is also true for the expected spacings model for the probabilities of imperfect ranking (Bohn and Wolfe, 1994). Furthermore, Fligner and MacEachern (2006) argue that an appropriate imperfect-ranking model based on perceived values of units should satisfy (2).

We now consider the built-in ordering in the means of the strata. Note that for $1 \leq i \leq H$,

$$\begin{aligned}\mu_{[i]} &= \int_{-\infty}^{+\infty} y dF_{[i]}(y) \\ &= \int_{0}^{1} F_{[i]}^{-1}(z)dz. \end{aligned} \tag{3}$$

In the second equality above, we change the variable $z = F_{[i]}(y)$ such that $y = F_{[i]}^{-1}(z)$, where $F_{[i]}^{-1}(z) \equiv \inf\{y : F_{[i]}(y) = z\}$. Based on (2), it is easy to verify that for any $1 \leq i < j \leq H$, $0 \leq z \leq 1$,

$$F_{[i]}^{-1}(z) \leq F_{[j]}^{-1}(z)$$

which, combined with (3), yields the following result.

**Proposition 1.** *Assume that* $Y_{[1]}, \cdots Y_{[H]}$ *be stochastically ordered. Then*

$$\mu_{[1]} \leq \cdots \leq \mu_{[H]}. \tag{4}$$

However, the sample means $\bar{Y}_{[h]}$ may violate the simple order constraint (4), due to sampling variations. It might be helpful to replace each $\bar{Y}_{[h]}$ in (1) by its isotonized version $\bar{Y}^*_{[h]}$ to obtain the new estimator $\hat{\mu}^*$,

$$\hat{\mu}^* = \frac{1}{H} \sum_{h=1}^{H} \bar{Y}^*_{[h]}, \tag{5}$$

where

$$\bar{Y}^*_{[h]} = \max_{r \leq h} \min_{s \geq h} \sum_{g=r}^{s} \frac{n_g \bar{Y}_{[g]}}{n_{rs}}, \quad n_{rs} = \sum_{g=r}^{s} n_g.$$

Indeed, $\{\bar{Y}^*_{[h]}\}$ is the well-known isotonic regression estimator of $\{\bar{Y}_{[h]}\}$ with weights $(n_h)_{h=1}^{H}$, which minimizes the weighted least square $\sum_{h=1}^{H} (\bar{Y}_{[h]} - \mu_{[h]})^2 n_h$ over the restricted space $\{\boldsymbol{\mu} \in R^H, \mu_{[1]} \leq \cdots \leq \mu_{[H]}\}$. According to Chapter 1, Robertson et al. (1988), the following results hold for $\{\bar{Y}^*_{[h]}\}$:

1. $\sum_{h=1}^{H} (\bar{Y}_{[h]} - \mu_{[h]})^2 n_h \geq \sum_{h=1}^{H} (\bar{Y}_{[h]} - \bar{Y}^*_{[h]})^2 n_h + \sum_{h=1}^{H} (\bar{Y}^*_{[h]} - \mu_{[h]})^2 n_h$;
2. $\sum_{h=1}^{H} (\bar{Y}_{[h]} - \bar{Y}^*_{[h]}) \bar{Y}^*_{[h]} n_h = 0$; $\sum_{h=1}^{H} (\bar{Y}_{[h]} - \bar{Y}^*_{[h]}) \mu_{[h]} n_h \leq 0$;
3. $\sum_{h=1}^{H} \bar{Y}_{[h]} n_h = \sum_{h=1}^{H} \bar{Y}^*_{[h]} n_h$.

Note that if $n_h = n/H$, then $\hat{\mu}^* = \hat{\mu}$, which follows directly from the third result above. This indicates that for a balanced ranked set sample, adjusting for the underlying ordering has no effect at all. However, in JP-S samples, due to random allocation, it is rare to have an equal sample size of each stratum, especially when $n$ is small.

Using Theorem 2.2 of Barlow et al. (1972), we have that $\bar{Y}^*_{[h]}$ is a strongly consistent estimator of $\mu_{[h]}$, since $\bar{Y}_{[h]}$ is strongly consistent of $\mu_{[h]}$. Noting that

$$|\hat{\mu}^* - \mu| = \left| \frac{1}{H} \sum_{h=1}^{H} \left( \bar{Y}^*_{[h]} - \mu_{[h]} \right) \right| \leq \frac{1}{H} \sum_{h=1}^{H} \left| \bar{Y}^*_{[h]} - \mu_{[h]} \right|$$

yields the following result.

**Theorem 1.** *$\hat{\mu}^*$ is a strongly consistent estimator of $\mu$.*

Finally, we mention that $\hat{\mu}^*$ can be easily computed via a linear-time algorithm, called pool adjacent violators (PAV) (Ayer et al., 1955). Codes written in FORTRAN are available online in CMU StatLib.

# 3 Comparison

We first compare the large-sample properties of the JP-S estimators with and without using isotonic regression. To do this, the stratum means $\mu_{[1]}$,...,$\mu_{[H]}$ are assumed to be strictly increasing. This is reasonable as long as there are no adjacent rank classes in which the ranker totally mixes up and performs in a purely random pattern. Obviously, it is true for perfect ranking.

**Theorem 2.** *Assume $\mu_{[1]} < \cdots < \mu_{[H]}$. Then the JP-S estimator $\hat{\mu}^*$ with isotonic regression satisfies*

$$\sqrt{n}(\hat{\mu}^* - \mu) \to N(0, \frac{\sum_{h=1}^{H} \sigma^2_{[h]}}{H}).$$

The technical proof of the theorem is given in the appendix. It is easy to verify that the JP-S estimator $\hat{\mu}$ without isotonic regression has the same asymptotic distribution. Then it follows immediately that the asymptotic relative efficiency of $\hat{\mu}^*$ versus $\hat{\mu}$ is 1. So for large sample sizes, it is expected that $\hat{\mu}^*$ and $\hat{\mu}$ perform comparably.

Next, we compare the small-sample behaviors of $\hat{\mu}^*$ and $\hat{\mu}$ through simulation. Table 1 reports the relative efficiency of the two JP-S estimators (assuming perfect ranking) to the SRS estimator $\bar{Y}$ for six different distributions, the standard normal, uniform$(0,1)$, U-shaped, gamma with shape parameter 5 and scale parameter 1, standard exponential and standard lognormal distributions. We set the number of ranked sets $H = 2, 3, 4, 5, 10$, and the sample size $n = H \times \bar{n}$ where the average sample size in each set is chosen as $\bar{n} = 2, 3, 4, 5, 20$. Note that $\bar{n} = 20$ is included so as to verify the asymptotic property given in Theorem 2. Here, efficiency ($eff$) is defined as the ratio of the variance of $\bar{Y}$ to MSE of each JP-S estimator, where MSE is simulated from 5000 replicates except for the lognormal case where 5000 is replaced by 20,000.

Table 1: Comparing efficiency of the JP-S Estimators with and without isotonic regression

| | | **N(0,1)** | | **Uni(0,1)** | | **Ushape** | | **Gam** | | **Exp(1)** | | **LN(0,1)** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| H | $\bar{n}$ | $\hat{\mu}^*$ | $\hat{\mu}$ | $\hat{\mu}^*$ | $\hat{\mu}$ | $\hat{\mu}^*$ | $\hat{\mu}$ | $\hat{\mu}^*$ | $\hat{\mu}$ | $\hat{\mu}^*$ | $\hat{\mu}$ | $\hat{\mu}^*$ | $\hat{\mu}$ |
| 2 | 2 | *1.08* | 1.06 | *1.10* | 1.07 | *1.08* | 1.04 | *1.07* | 1.05 | *0.99* | 0.97 | *0.97* | 0.97 |
| | 3 | *1.16* | 1.14 | *1.18* | 1.14 | *1.18* | 1.13 | *1.13* | 1.10 | *1.10* | 1.08 | *0.98* | 0.97 |
| | 4 | *1.21* | 1.19 | *1.25* | 1.22 | *1.22* | 1.18 | *1.19* | 1.17 | *1.10* | 1.09 | *1.07* | 1.07 |
| | 5 | *1.27* | 1.26 | *1.32* | 1.31 | *1.28* | 1.25 | *1.27* | 1.26 | *1.15* | 1.14 | *1.03* | 1.03 |
| | 20 | *1.42* | 1.42 | *1.47* | 1.47 | *1.44* | 1.44 | *1.37* | 1.37 | *1.35* | 1.35 | *1.15* | 1.15 |
| 3 | 2 | *1.29* | 1.17 | *1.31* | 1.20 | *1.27* | 1.20 | *1.40* | 1.25 | *1.30* | 1.10 | *1.19* | 1.03 |
| | 3 | *1.41* | 1.30 | *1.50* | 1.39 | *1.40* | 1.29 | *1.43* | 1.33 | *1.29* | 1.20 | *1.11* | 1.04 |
| | 4 | *1.55* | 1.49 | *1.60* | 1.51 | *1.54* | 1.43 | *1.49* | 1.42 | *1.30* | 1.24 | *1.13* | 1.09 |
| | 5 | *1.56* | 1.53 | *1.66* | 1.61 | *1.58* | 1.50 | *1.54* | 1.50 | *1.41* | 1.37 | *1.12* | 1.10 |
| | 20 | *1.75* | 1.75 | *1.99* | 1.99 | *1.80* | 1.80 | *1.77* | 1.77 | *1.59* | 1.59 | *1.31* | 1.31 |
| 4 | 2 | *1.59* | 1.32 | *1.62* | 1.36 | *1.51* | 1.32 | *1.59* | 1.28 | *1.54* | 1.21 | *1.28* | 1.01 |
| | 3 | *1.74* | 1.55 | *1.85* | 1.65 | *1.73* | 1.50 | *1.72* | 1.51 | *1.58* | 1.38 | *1.19* | 1.08 |
| | 4 | *1.83* | 1.69 | *2.00* | 1.83 | *1.82* | 1.63 | *1.91* | 1.74 | *1.50* | 1.40 | *1.23* | 1.17 |
| | 5 | *1.94* | 1.85 | *2.16* | 2.04 | *1.85* | 1.71 | *1.76* | 1.70 | *1.56* | 1.49 | *1.20* | 1.18 |
| | 20 | *2.32* | 2.32 | *2.37* | 2.36 | *2.16* | 2.16 | *2.15* | 2.15 | *1.87* | 1.86 | *1.42* | 1.42 |
| 5 | 2 | *1.78* | 1.43 | *1.92* | 1.48 | *1.79* | 1.43 | *1.84* | 1.39 | *1.63* | 1.21 | *1.29* | 1.00 |
| | 3 | *2.02* | 1.72 | *2.22* | 1.83 | *2.11* | 1.74 | *2.06* | 1.71 | *1.66* | 1.40 | *1.32* | 1.18 |
| | 4 | *2.19* | 1.99 | *2.33* | 2.06 | *2.17* | 1.88 | *2.10* | 1.92 | *1.75* | 1.63 | *1.24* | 1.18 |
| | 5 | *2.31* | 2.18 | *2.53* | 2.34 | *2.27* | 2.08 | *2.11* | 2.01 | *1.86* | 1.76 | *1.24* | 1.21 |
| | 20 | *2.69* | 2.69 | *2.91* | 2.90 | *2.50* | 2.50 | *2.50* | 2.50 | *2.11* | 2.11 | *1.52* | 1.52 |
| 10 | 2 | *2.99* | 1.81 | *3.67* | 1.96 | *3.05* | 1.76 | *3.08* | 1.69 | *2.64* | 1.62 | *1.76* | 1.25 |
| | 3 | *3.50* | 2.47 | *4.25* | 2.62 | *3.57* | 2.34 | *3.35* | 2.30 | *2.66* | 2.02 | *1.69* | 1.43 |
| | 4 | *3.66* | 2.93 | *4.40* | 3.18 | *3.53* | 2.73 | *3.47* | 2.77 | *2.72* | 2.33 | *1.55* | 1.43 |
| | 5 | *3.96* | 3.43 | *4.86* | 4.03 | *3.80* | 3.12 | *3.57* | 3.14 | *2.68* | 2.45 | *1.65* | 1.58 |
| | 20 | *4.66* | 4.65 | *5.36* | 5.35 | *4.19* | 4.17 | *4.12* | 4.12 | *3.29* | 3.29 | *1.95* | 1.94 |

Figure 1: Improvement in percent from using isotonic regression for various distributions under perfect ranking. Each panel contains five lines, for different values of $H$, indicated by the number connecting each line. And $\bar{n}$ is the average sample size in each ranked set.

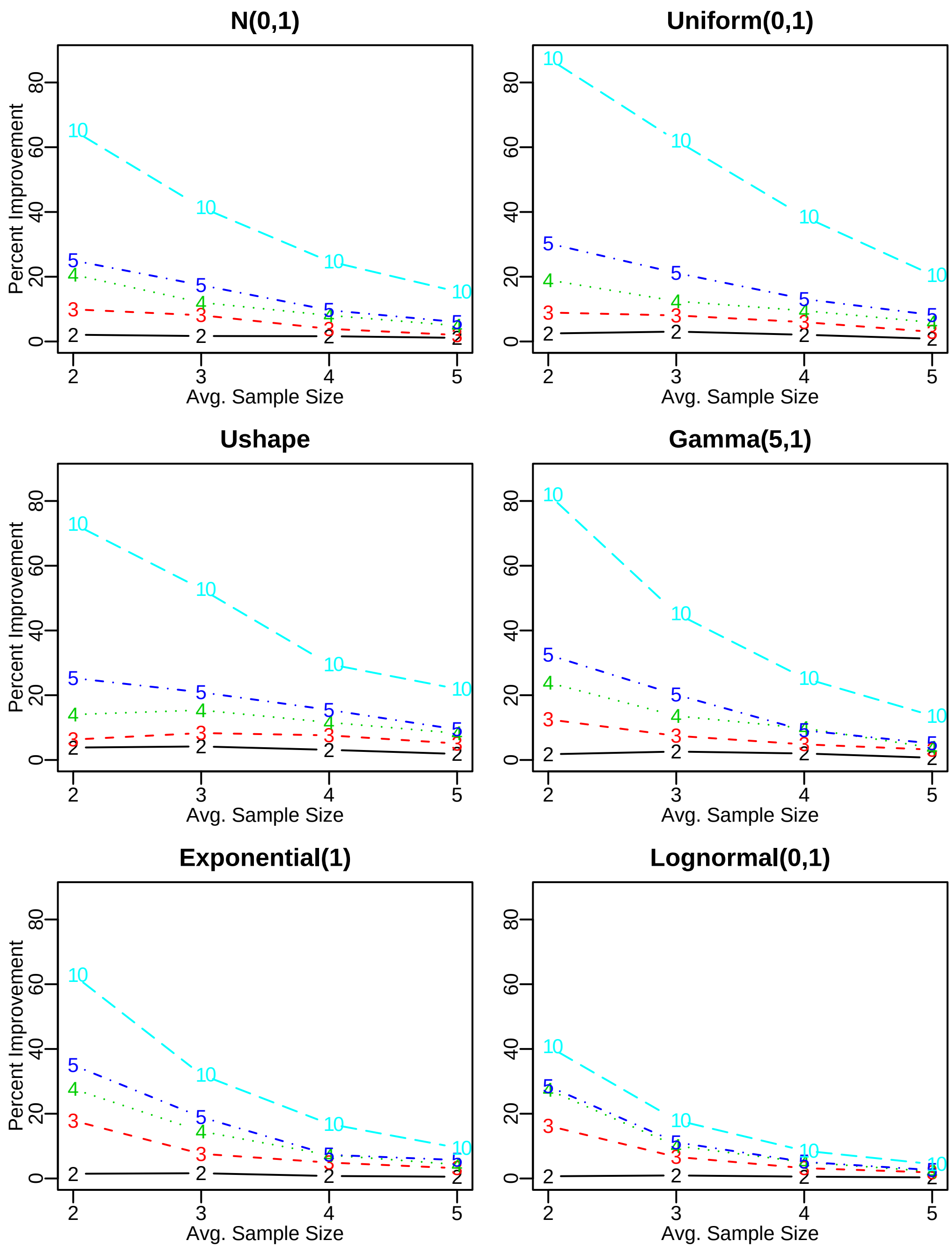



Table 1 shows that $\hat{\mu}^*$ is more efficient than $\hat{\mu}$ and $\bar{Y}$ in nearly all the settings. When $\bar{n} = 20$, the two JP-S estimators have virtually the same performance, which confirms that their asymptotic relative efficiency is one. Figure 1 further shows size of the improvement from using isotonic regression, defined as $[eff(\hat{\mu}^*) - eff(\hat{\mu})]/eff(\hat{\mu}) \times 100\%$. Clearly, the improvement increases as $H$ increases. This is reasonable, since the more sets we have, the more violations of the underlying order restrictions may occur in the sample means so that the isotonic regression takes greater effect. Another observation is that under the uniform distribution the size of improvement appears the biggest and under the lognormal distribution it appears the smallest. But for all the distributions, the improvement is considerable when $H$ is not small (say $H > 3$), and $n$ is small (say $\bar{n} < 5$).

We should mention that the better performance of $\hat{\mu}^*$ is achieved without any distributional assumptions. The improvement at small sample sizes is important in applications that require cost efficiency, where JP-S sampling is found most useful.

# 4 Extension to Multiple Rankers and Imprecise Ranking

Judgment post-stratification is known as a data collection method closely related to ranked set sampling, both of which are based on similar ideas and applied in similar situations. Certainly, JP-S has its own merits. As mentioned in the introduction, JP-S can incorporate multiple rankers and imprecise ranking to further improve mean estimation. Multiple rankers are useful in applications in which they can be recruited with some minimal cost. They can also be substituted by ranking through auxiliary variables that are easily attainable. Imprecise ranking allows a ranker to assign probabilities to ranks, and thus is useful in situations when ties occur or he/she has difficulty in producing a complete ordering for units in a set. We shall show below that the idea of isotonic regression can be applied in the presence of multiple rankers or imprecise ranking without any additional difficulty.

In the case where assessments of ranks are available from $m$ rankers, the data generated by a JP-S sample can be expressed by $\mathcal{D} = (Y_i, O_{i1}, ..., O_{im})_{i=1}^n$, where $O_{ij}$ is the judgment order of $Y_i$ assigned by ranker $j$ among its own set of unmeasured units, for $j = 1, ...m$. To combine information from the multiple rankers, MacEachern et al. (2004) proposed a nonparametric estimator of $\mu$,

$$\hat{\mu}^{(m)} = \frac{1}{H}\sum_{h=1}^{H}\hat{\mu}_{[h]}^{(m)} = \frac{1}{H}\sum_{h=1}^{H}\frac{\sum_{i=1}^{n} Y_i p_{ih}}{\sum_{i=1}^{n} p_{ih}}, \tag{6}$$

where $p_{ih} = \sum_{j=1}^{m} I(O_{ij} = h)/m$ is the proportion of rankers who classify $Y_i$ as having rank $h$. The corresponding estimator with isotonic regression can be constructed as

$$\hat{\mu}^{*(m)} = \frac{1}{H}\sum_{h=1}^{H}\hat{\mu}_{[h]}^{*(m)}, \tag{7}$$

where $\{\hat{\mu}_{[h]}^{*(m)}\}$ is the isotonized version of $\{\hat{\mu}_{[h]}^{(m)}\}$ with weights $(\tilde{n}_h)_{h=1}^{H}$,

$$\hat{\mu}_{[h]}^{*(m)} = \max_{r\le h}\min_{s\ge h}\sum_{g=r}^{s}\frac{\tilde{n}_g\hat{\mu}_{[g]}^{(m)}}{\tilde{n}_{rs}}, \quad \tilde{n}_g = m\sum_{i=1}^{n} p_{ih}; \quad \tilde{n}_{rs} = \sum_{g=r}^{s}\tilde{n}_g.$$

Computing $\hat{\mu}^{*(m)}$ when $m \geq 2$ is actually as simple as $\hat{\mu}^*$ for $m = 1$. Note that the original estimator $\hat{\mu}^{(m)}$ is equivalent to the following estimation process: first transpose the data $\mathcal{D}$ to $\tilde{\mathcal{D}} = [(Y_1, O_{11}), ..., (Y_1, O_{1m}), (Y_2, O_{21}), ..., (Y_2, O_{2m}), ..., (Y_n, O_{n1}), ..., (Y_n, O_{nm})]^T$, where each $Y_i$ value is replicated $m$ times; then use $\hat{\mu}$ in (1) with $\tilde{\mathcal{D}}$ to estimate $\mu$ as if there were only one ranker but $m \times n$ observations. This fact allows us to compute $\hat{\mu}^{*(m)}$ exactly the same as in the case of one ranker, but with the transposed data $\tilde{D}$.

In the case where imprecise ranking is allowed, each ranker assigns a distribution on the ranks so that the JP-S data can be expressed by $\mathcal{D} = (Y_i, \mathbf{p}_{i1}, ..., \mathbf{p}_{im})_{i=1}^{n}$, where $\mathbf{p}_{ij} = (p_{ij1}, ..., p_{ijH})$ satisfies $p_{ij1} + \cdots + p_{ijH} = 1$ and $p_{ijh}$ is the probability assigned by ranker $j$ that the $i$th fully measured unit has rank $h$. In formulas (6) and (7), by redefining

$$p_{ih} = \frac{1}{m}\sum_{j=1}^{m} p_{ijh},$$

both $\hat{\mu}^{(m)}$ and $\hat{\mu}^{*(m)}$ can be applied to estimate $\mu$.

As will be shown in next section, $\hat{\mu}^{*(m)}$ can improve $\hat{\mu}^{(m)}$ in both situations.

# 5 Empirical Studies

We have demonstrated in Section 2 that under perfect ranking from one ranker, our proposed estimator outperformed the one without using isotonic regression. The purpose of this section is to compare their performance under various practical situations and to investigate the impact of imperfect ranking, multiple rankers and imprecise ranking on three examples, respectively.

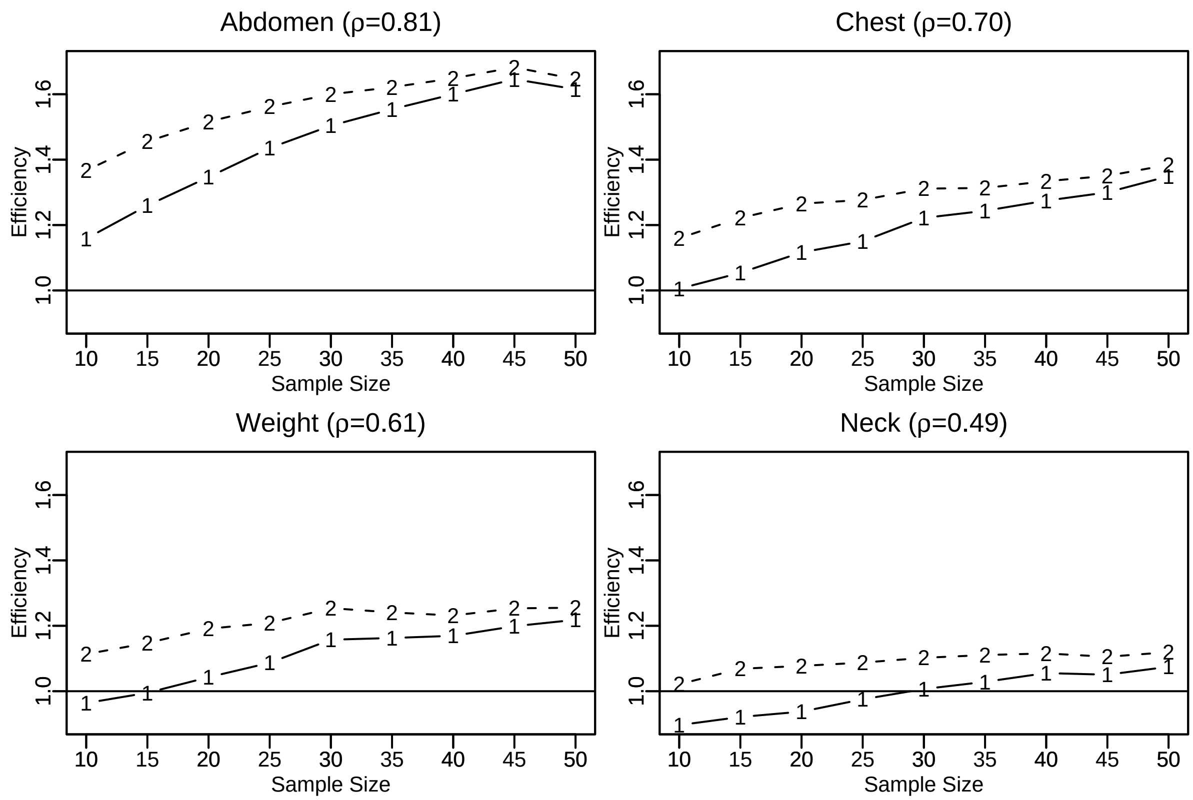
## 5.1 Imperfect Ranking: Body Fat

We first consider the data set containing the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men, which, along with a detailed description, is available at `http://lib.stat.cmu.edu/datasets/bodyfat`. We set our goal as estimation of the mean percentage of body fat for the 252 mean from simulated JP-S samples. To test the impact of ranking errors, ranking was not done directly on the percentage of body fat, but on other correlated measurements that are easier to obtain. We chose abdomen circumference, chest circumference, weight, and neck circumference as ranking variables, each having correlation with the percentage of body fat 0.81, 0.70, 0.61 and 0.49, respectively.

In this simulation, we fix $H$ at 5, and let the sample size $n$ vary from 10 to 50 with step size 5. To generate a JP-S sample of size $n$, we randomly selected a group of five subjects from the entire data set $n$ times. Among each of the $n$ groups, ranking was done based on one of the four ranking variables, and then one out of the five subjects was randomly selected to enter the JP-S sample.

Figure 2 shows values of the simulated relative efficiency of the two JP-S estimators (with and without isotonic regression) to $\bar{Y}$ for each sample size and each ranking variable. In each setting, MSE is estimated from 20,000 replicates. The figure shows that when the correlation is not weak, both JP-S estimators can do better than the SRS estimator. The stronger the correlation is, the more improvement the JP-S estimators can achieve over the SRS estimator. Also, $\hat{\mu}^*$ is uniformly better than $\hat{\mu}$. The performance difference between $\hat{\mu}$ and $\hat{\mu}^*$ is bigger for small $n$ and becomes smaller as $n$ gets large, which is consistent with

Figure 2: An empirical study of percentage of body fat. In each panel, the line connected by "1" is for $\hat{\mu}$ and the line connected by "2" is for $\hat{\mu}^*$; a horizontal line at efficiency equal to 1 is drawn as a reference line.




what we observe from Table 1.

## 5.2 Multiple Rankers: Adjusted Brain Weights of Mammals

We provide an illustration of our proposed method in the setting of multiple rankers. Here, we use the data set given in Section 4.2 of Stokes et al. (2006). It consists of 20 groups and each has 3 species of mammals with their adjusted brain weights $Y$ calculated from the formula $Y = \log\{\text{brain weight}/(\text{body weight})^{2/3}\}$. Within each group, ranks of $Y$ are available from two different rankers. The rankers did not know the value of $Y$ for each mammal before they assigned ranks. So they made judgments based on the conjecture that a "clever" species tends to have a large adjusted brain weight.

We assume that each of the 20 sets represents three independent draws from a large population of species, and set our goal as estimation of the mean of $Y$. To compare the estimators with two rankers, $\hat{\mu}^{(2)}$ and $\hat{\mu}^{*(2)}$, we conducted a simulation, in which $H = 3$ and $n$ varies from 3 to 18 with step size 3. In each iteration, a sample of $n$ species was selected, with one species from each set. Table 2 summarizes the results based on 20,000 iterations for each sample size $n$, which clearly shows that in presence of two rankers, using isotonic regression can improve mean estimation, especially for small sample sizes.

Table 2: Comparing efficiency of the JP-S estimators with and without isotonic regression through an empirical study of adjusted brain weights of mammals, in which two rankers are available.

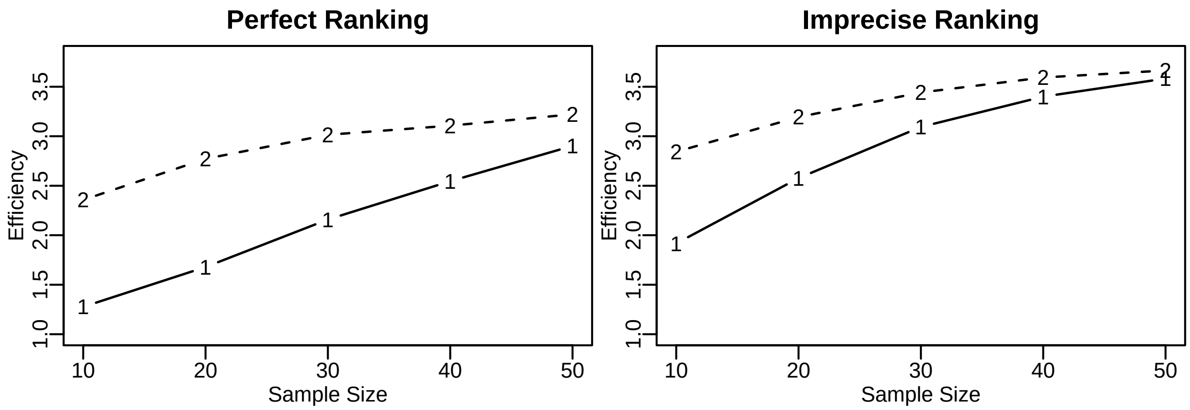
| n | 3 | 6 | 9 | 12 | 15 | 18 |
|---|---|---|---|---|---|---|
| $\hat{\mu}^{(2)}$ | 1.07 | 1.19 | 1.25 | 1.38 | 1.47 | 1.52 |
| $\hat{\mu}^{*(2)}$ | 1.28 | 1.36 | 1.37 | 1.46 | 1.52 | 1.56 |

## 5.3 Imprecise Ranking: Length of Master's Theses

To study the effect of imprecise ranking, we consider the data set in Stokes and Sager (1988), which contains the volumes of 300 master's theses from a library at the University of Texas of Austin. The data were collected by two rankers, each of which each time visually judged the largest and smallest volumes of three contiguous, randomly selected shelved master's theses in the library. Both rankers were forced to give exact ranks, and so made ranking errors due to the fact that the books were sometimes visually indistinguishable. Thus, it might be better to allow for imprecise ranking.

For illustrative purposes, we set our goal as estimation of the mean volume of the 300 theses. Note that the volume distribution in this data set is right skewed. To simulate a JP-S sample of size $n$ with imprecise ranking, a "perceived" ranker repeats the following procedure $n$ times. First, she randomly selects a group of $H$ books from the entire data set. Within the group, one book is then randomly selected to enter the JP-S sample, and all other $H - 1$ books are compared with it based on visual judgment. The ranker is assumed to claim that a tie occurs when the difference of two books in volume is not larger than 10 pages. She counts among the $H - 1$ books, how many longer than the selected book (say $l$), how many shorter (say $s$), how many ties (say $t$), where $l + s + t = H - 1$. Based on

Figure 3: An empirical study of the volumes of shelved master's theses. In each panel, the line connected by "1" is for $\hat{\mu}$ and the line connected by "2" is for $\hat{\mu}^*$.




$(l, s, t)$, the ranker distribution $(p_{i1}, \cdots, p_{iH})$ for the selected book (say the $i$ observation) is determined as $(0, \cdots, 0, \frac{1}{t+1}, \cdots, \frac{1}{t+1}, 0, \cdots 0)$, where there are $t+1$ nonzero probabilities, and $l$ and $s$ zero probabilities before and after the nonzero ones, respectively.

In our simulation, $H = 10$ and $n$ varies from 10 to 50 with step size 10. Again, MSE for each estimator is estimated from 20,000 replicates. We also consider the two estimators $\hat{\mu}$ and $\hat{\mu}^*$ under perfect ranking. Figure 3 plots the simulated efficiency of the two JP-S estimators to $\bar{Y}$ under each setting.

As we might expect, the figure shows that allowing the ranker to assign imprecise ranks improves mean estimation. Again, $\hat{\mu}^*$ is uniformly better than $\hat{\mu}$ under both perfect ranking and imprecise ranking. The size of improvement from using isotonic regression seems bigger under perfect ranking than that under imprecise ranking.

# 6 Discussion

For JP-S samples, we have shown that by imposing the ordering of the stratum means via isotonic regression, the proposed method can achieve significant improvement over the existing one in mean estimation. In parallel, such ordering arises naturally in RSS. Thus, our method could be used for data collected from an unbalanced RSS for potentially better mean estimates while for a balanced RSS, it yields exactly the same estimate.

In addition, our method can be extended to estimation of the distribution function for a population of interest. Research in this direction is given by Ozturk (2006) in the context of RSS.

## Appendix: Proof of Theorem 2

In spirit of the work by Barmi and Mukerjee (2005), we derive the asymptotic distribution of $\hat{\mu}^*$.

For $h = 1, 2, ..., H$,

$$\begin{aligned}\sqrt{n}(\bar{Y}^*_{[h]} - \mu_{[h]}) &= \sqrt{n} \max_{r \leq h} \min_{h \leq s} \sum_{j=r}^{s} \frac{n_j}{n_{rs}} \left(\bar{Y}_{[j]} - \mu_{[h]}\right) \\ &= \max_{r \leq h} \min_{h \leq s} \sum_{j=r}^{s} \frac{n_j}{n_{rs}} \left[\sqrt{n}(\bar{Y}_{[j]} - \mu_{[j]}) + \sqrt{n}(\mu_{[j]} - \mu_{[h]})\right]. \qquad \text{(A.1)}\end{aligned}$$

As $n \to +\infty$, it is well known that

$$\sqrt{n_j}(\bar{Y}_{[j]} - \mu_{[j]}) \to^d \quad N(0, \sigma^2_{[j]}), \qquad \text{(A.2)}$$

and $n_j/n \to 1/H$ with probability 1 for $j = 1, 2..., H$. Then based on the continuous mapping theorem, we have

$$\sqrt{n}(\bar{Y}_{[j]} - \mu_{[j]}) \to^d \quad N(0, H\sigma^2_{[j]}),$$

which indicates that $\{\sqrt{n}(\bar{Y}_{[j]} - \mu_{[j]})\}_{j=1}^{H}$ are bounded in probability. Further, noting that in (A.1), $\sqrt{n}(\mu_{[j]} - \mu_{[h]}) = +\infty(-\infty)$ as $n \to +\infty$ for $j > h(< h)$, we find that the $r$ and $s$ in the max and min will be restricted to $r = s = h$ with arbitrarily high probability for sufficiently large $n$, yielding

$$\sqrt{n}(\bar{Y}^*_{[h]} - \mu_{[h]}) \to^p \quad \sqrt{n}(\bar{Y}_{[j]} - \mu_{[j]}). \qquad \text{(A.3)}$$

From (A.2) and the independence of $\bar{Y}_{[j]}$'s,

$$\left(\sqrt{n}(\bar{Y}_{[1]} - \mu_{[1]}), \cdots, \sqrt{n}(\bar{Y}_{[H]} - \mu_{[H]})\right)^T \Rightarrow^w \quad (Z_1, \cdots, Z_H)^T,$$

which is a $H-$variate Gaussian process with independent components $Z_i =^d \quad N(0, H\sigma^2_{[j]})$. Combined with (A.3), it follows that

$$\sqrt{n}(\hat{\mu}^* - \mu) = \sum_{h=1}^{H} \sqrt{n}(\bar{Y}^*_{[h]} - \mu_{[h]})/H$$
$$\to^d \quad N(0, \frac{\sum_{h=1}^{H} \sigma^2_{[h]}}{H}).$$

# References


Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, 26:641–647.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*. Wiley, New York.

Barmi, H. E. and Mukerjee, H. (2005). Inferences under stochastic ordering constraint: the k-sample case. *Journal of the American Statistical Association*, 100:252–261.

Bohn, L. L. and Wolfe, D. A. (1994). The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the mann-whitney-wilcoxon statistic. *Journal of the American Statistical Association*, 89:168–176.

Chen, Z., 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.

Feltz, C. J. and Dykstra, R. L. (1985). Maximum likelihood estimator of the survival functions of N stochastically ordered distributions. *Journal of the American Statistical Association*, 80:1012–1019.

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

Lee, C.-I. C. (1981). The quadratic loss of isotonic regression under normality. *The Annals of Statistics*, 9:686–688.

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

Ozturk, O. (2006). Statistical inference under a stochastic ordering constraint in ranked set sampling. Technical report, Department of Statistics, The Ohio State Univeristy.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order-Restricted Inferences*. New York: Wiley.

Ross, N. P. and Stokes, L. (1999). Editorial: Special issue on statistical design and analysis with ranked set sampling. *Environmental and Ecological Statistics*, 6:5–9.

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. (2006). Judgment post-stratification with multiple rankers. Technical report, Southern Methodist University.

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*, to appear.