A Nonparametric Mean Estimator for Judgment Post-Stratified Data

Xinlei Wang, Johan Lim and Lynne Stokes*

June 2007

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 et al. (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-effective sampling method. It is 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. Theoretical development showed that regardless of ranking errors, the RSS estimator of a population mean is at least as efficient as the estimator from a simple random sample of the same sample size. For an overview of RSS, see Chen, Bai and Sinha (2006) and the references therein.

Recently, judgment post-stratification (JP-S) has been proposed by MacEachern, Stasny and Wolfe (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, where measuring experimental units is much more expensive than recruiting and ranking them. Such situations arise naturally and frequently in fields of agriculture and environmental sciences (Ross and Stokes, 1999). For example, in an ecological study with the aim of estimating mean pool area for a stream, the pool area can be formally measured by a team of field assistants, which is costly; but it can be estimated visually and then ranked easily (Mode, Conquest and Marker, 1999). 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 experimental units are 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 begin with a nonstandard design, but willing to begin with an SRS with the option of using auxiliary rank information later. JP-S is more flexible than RSS in other ways as well. 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 are most likely to occur when the total sample size is small, which is the typical case in applications where accurate measurements are difficult or costly.

The literature indicates that by imposing underlying order constraints in some problems, substantial reduction in mean square errors can be obtained (Feltz and Dykstra, 1985; Lee, 1981). Here, for JP-S data, we propose 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 its 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 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 μ and finite variance σ^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 ($1 \leq i \leq n$), 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 hth 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 hth 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 μ,

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

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 (cdf) within the hth stratum. This assumption is true when ranking is perfect, since in this case $F_{[h]}(y)$ becomes $F_{(h)}(y)$, the distribution function of the hth order statistic (in an increasing order). 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 under the expected spacings model for the probabilities of imperfect ranking (Bohn and Wolfe, 1994). Even if ranking is perfectly random, or worthless, (2) is satisfied with $F_{[1]}(y) = \cdots = F_{[H]}(y) = F_Y(y)$, where $F_Y(y)$ is the cdf of Y. 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. It is well known that if two distributions are stochastically ordered (say $F(x) \geq G(x)$), then $E_F(\Phi(X)) \leq E_G(\Phi(X))$ for any nondecreasing function Φ. Taking $\Phi(X) = X$ yields the following result,

$$\mu_{[1]} \leq \cdots \leq \mu_{[H]}. \tag{3}$$

However, the sample means $\bar{Y}_{[h]}$ may violate the simple order constraint (3), due to sampling variation. We propose to replace each $\bar{Y}_{[h]}$ in (1) by its isotonized version $\bar{Y}^*_{[h]}$ to obtain a new estimator $\hat{\mu}^*$,

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

where

$$\bar{Y}^*_{[h]} = \max_{r \leq h} \min_{s \geq h} \sum_{g=r}^{s} \frac{n_g \bar{Y}_{[g]}}{n_{rs}}, \quad n_{rs} = \sum_{g=r}^{s} n_g. \tag{4}$$

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 $\{\boldsymbol{\mu} \in R^H, \mu_{[1]} \leq \cdots \leq \mu_{[H]}\}$. If $n_h = n/H$, it follows from the properties of the isotonic regression that $\hat{\mu}^* = \hat{\mu}$ (see Chapter 1, Robertson, Wright and Dykstra, 1988). 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 in each post-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 for $\mu_{[h]}$. 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 μ.*

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, we begin with the case where the stratum means $\mu_{[1]},...,\mu_{[H]}$ are strictly increasing. Obviously, this is true for perfect ranking. It is also true as long as the ranker is doing any better than purely random in distinguishing any two different ranks.

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

$$\sqrt{n}(\hat{\mu}^* - \mu) \rightarrow N(0, \frac{\sum_{h=1}^{H} \sigma^2_{[h]}}{H}). \tag{5}$$

The proof of the theorem can be found in the appendix. It is easy to verify that the JP-S estimator $\hat{\mu}$ without isotonic regression has the same asymptotic distribution as (5). 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 $\hat{\mu}^*$ provides no advantage over $\hat{\mu}$.

However, the assumption of strict ordering might be violated in some situations. For example, the ranker may always be able to identify the minimum correctly, but cannot distinguish any other ranks. In this case, $\mu_{[1]} < \mu_{[2]} = \cdots = \mu_{[H]}$. More generally, if $\mu_{[1]} \leq \cdots \leq \mu_{[H]}$, it can be verified that

$$\sqrt{n}(\hat{\mu}^* - \mu) \to^d \quad \frac{1}{\sqrt{H}} \sum_{h=1}^{H} \max_{r \leq h} \min_{s \geq h} \frac{\sum_{r \leq g \leq s;\, r,s \in S_h} Z_g}{s - r + 1}, \tag{6}$$

where $Z_g \sim N(0, \sigma^2_{[g]})$, $g = 1, ..., H$; and Z_g's are independent (see the appendix for details). Due to the dependence introduced by isotonic regression among the estimates of stratum means, it is difficult to derive a closed form formula for the variance of $\sqrt{n}(\hat{\mu}^* - \mu)$. Thus, theoretical comparison of the large sample performance may not be possible. Nevertheless, the distribution in (6) can be simulated to provide the computed asymptotic relative efficiency, in case that ties in stratum means are of interest.

Next, we compare the small-sample behavior of $\hat{\mu}^*$ and $\hat{\mu}$ through simulation. First, note that when n is small, empty cells may occur so that $\hat{\mu}$ is not directly applicable, since the means of the empty cells are not estimable. In this case, we calculate $\hat{\mu}$ as if empty cells do not exist, by taking the average of the sample means of those nonempty cells only. Since Y falls in each stratum with an equal probability $1/H$, doing so makes $\hat{\mu}$ still a consistent mean estimator even when there are empty cells. Also note that empty cells do not cause any problem for $\hat{\mu}^*$. This is because in (4), taking the maximum or minimum would automatically skip the empty cells so that the sample mean of each empty cell is substituted by the (pooled) sample mean of one (or more) of its adjacent cells.

Table 1 reports the simulated relative efficiency of $\hat{\mu}^*$ to $\hat{\mu}$ (assuming perfect ranking) 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, as in MacEachern et al. (2002). Here, relative efficiency (RE) is defined as the ratio of the MSE of $\hat{\mu}$ versus MSE of $\hat{\mu}^*$. 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 $\bar{n} = 1, 2, 3, 4, 5, 20$. Note that $\bar{n} = 20$ is included so as to verify the asymptotic property given in Theorem 2. Under each setting, RE is estimated from 20,000 replicates. That is, for each replicate, calculate $Z = (\hat{\mu} - \mu)^2$ and $Z^* = (\hat{\mu}^* - \mu)^2$; and then $\widehat{mse}(\hat{\mu}) = \bar{Z}$, $\widehat{mse}(\hat{\mu}^*) = \bar{Z}^*$ and $\widehat{RE} = \widehat{mse}(\hat{\mu})/\widehat{mse}(\hat{\mu}^*)$. The standard error of each RE estimate is also reported in Table 1, which is estimated from the delta method using the formula

$$Var(\frac{\bar{Z}}{\bar{Z}^*}) \approx \frac{1}{N(\bar{Z}^*)^2}(S_Z^2 + \widehat{RE}^2 \times S_{Z^*}^2 - 2\widehat{RE} \times S_{ZZ^*}),$$

where N is the number of replicates, S_Z^2, $S_{Z^*}^2$ and S_{ZZ^*} are sample variances and covariance of (Z, Z^*) over all the replicates.

Table 1 shows that $\hat{\mu}^*$ is more efficient than $\hat{\mu}$ in nearly all cases considered. When $\bar{n} = 20$, the two JP-S estimators have virtually the same performance, as expected. Also, for cases with $H = 2$ and $\bar{n} = 1$, they always give the same estimate $(y_1 + y_2)/2$, no matter whether the order is violated or not.

Figure 1 summarizes the results of the simulation by showing a graph of the relative efficiency as a function of average sample size for different numbers of classes. Clearly, the improvement increases as the number of classes H increases. This occurs because the more sets we have, the more violations of the underlying order restrictions may occur in the sample means. Also, the figure suggests that the improvement is not necessarily a decreasing function of the average sample size $\bar{n}$, but it finally diminishes to zero in all cases. For all the distributions, the improvement is considerable when the number of groups is not small and the average sample size is small (say $H > 3$ and $\bar{n} < 5$).

4 Extension to Multiple Rankers and Imprecise Ranking

Judgment post-stratification is a data collection method closely related to ranked set sampling and shares its advantages. But JP-S has its own merits, one being that it can incorporate multiple rankers and imprecise ranking into the ranking process. Multiple rankers are useful in applications in which rankers can be recruited with minimal cost. Another application would be the situation in which ranking is accomplished using an auxiliary variable. When several variables are available, multiple rankings exist. Imprecise ranking is the term used by MacEachern et al. (2004) to describe a ranking procedure in which the ranker is allowed to assign probabilities to ranks to indicate his or her uncertainty. It is useful in situations when ties occur (for example, for an auxiliary variable) or he/she has difficulty in producing a complete ordering for units in a set. We will show that the isotonic regression idea 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 multiple rankers, MacEachern et al. (2004) proposed a nonparametric estimator of μ,

$$\hat{\mu}^{(m)} = \frac{1}{H}\sum_{h=1}^{H}\hat{\mu}_{[h]}^{(m)} = \frac{1}{H}\sum_{h=1}^{H}\frac{\sum_{i=1}^n Y_i p_{ih}}{\sum_{i=1}^n p_{ih}}, \tag{7}$$

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

$$\hat{\mu}^{*(m)} = \frac{1}{H}\sum_{h=1}^{H}\hat{\mu}_{[h]}^{*(m)}, \tag{8}$$

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

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

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 μ 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 ith fully measured unit has rank h. In formulas (7) and (8), 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 μ.

We show in the next section that $\hat{\mu}^{*(m)}$ can improve $\hat{\mu}^{(m)}$ in both situations.

5 Empirical Studies

We have demonstrated in Section 3 that under perfect ranking from one ranker, our proposed estimator outperforms the usual JP-S estimator for small sample sizes. The purpose of this section is to compare the performance of the estimators in three practical situations, including when imperfect ranking, multiple rankers and imprecise ranking are present. To do this, we present three examples. In each, we simulate selection of JP-S samples from a real finite population and compare the performance of the new and standard JP-S estimators of population mean; relative efficiency and its standard error are estimated from 20,000 replicates.

5.1 Imperfect Ranking: Body Fat

We first consider a data set containing the percentage of body fat determined by underwater weighing and various body size measurements for 252 men. These data are available at `http://lib.stat.cmu.edu/datasets/bodyfat`. Our target parameter is the mean percentage of body fat for the 252 men. To test the impact of ranking errors, ranking was not done directly on the percentage of body fat, but on auxiliary variables in the data set that would be easier to obtain in a real application. We chose abdomen circumference, chest circumference, weight, and neck circumference as ranking variables, each having correlation ρ with the percentage of body fat 0.81, 0.70, 0.61 and 0.49, respectively. Note that ρ directly conveys the strength of the ranking information.

In this simulation, we fix H at 5, and let the average sample size $\bar{n}$ vary from 1 to 10. 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. This procedure was repeated for each of the ranking variables.

Figure 2 shows values of the simulated relative efficiency ($\pm 2s.e.$) of the two JP-S estimators (with and without isotonic regression) for each average sample size and each ranking variable. The figure shows that $\hat{\mu}^*$ maintains an advantage over $\hat{\mu}$, even as ranking becomes less accurate (ρ decreases). Also, it appears that the advantage is not affected much by the strength of correlation except for very small $\bar{n}$, where the advantage becomes smaller as the correlation gets weaker.

We end this section by commenting that our proposed JP-S estimator does not require the values of the auxiliary variables $\mathbf{X}$, as in RSS. This makes JP-S easy to implement and applicable in nearly all situations where RSS is found useful. For example, if one is interested in estimating the mean volume of trees in a forest, he can rank the trees easily by eyes based on some perceived variable that combines information from the tree height and trunk circumference, which (at least the height) cannot be measured easily. Another example, as will be given in Section 5.2, is the adjusted brain weights example. The rankers assigned ranks based on their prior knowledge, whose values are not easy to quantify.

However, in some applications such as the body fat example, the values of $\mathbf{X}$ could be collected at some additional cost for all units (including those selected for ranking). In these cases, one might consider using a modified double-sampling estimator $\hat{\mu}_{mDS}$ in spirit of Yu and Lam (1997): $\hat{\mu}_{mDS} = \bar{Y}_{JP-S} + \hat{\mathbf{B}}(\bar{\mathbf{X}}' - \bar{\mathbf{X}}_{JP-S})$ where $\hat{\mathbf{B}}$ is the vector of regression coefficients, $\bar{\mathbf{X}}'$ is the sample mean of $\mathbf{X}$ based on all the selected Hn units; and $\bar{Y}_{JP-S}$ and $\bar{\mathbf{X}}_{JP-S}$ are our proposed JP-S estimators of μ_y and μ_x based on the n units with measured y values, respectively. As in Yu and Lam (1997), under the normality assumption, $\hat{\mu}_{mDS}$ would improve the original estimator $\hat{\mu}_{DS} = \bar{Y} + \hat{\mathbf{B}}(\bar{\mathbf{X}}' - \bar{\mathbf{X}})$ when the correlations $\boldsymbol{\rho}_{\mathbf{x}y}$ are positive ($\bar{Y}$ and $\bar{\mathbf{X}}$ are the SRS estimators of μ_y and μ_x based on the same n units).

5.2 Multiple Rankers: Adjusted Brain Weights of Mammals

This example provides an illustration of our proposed method in the setting of multiple rankers. Following Section 5 of MacEachern et al. (2004), we use a data set that consists of allometric measurements for 62 species of mammals, and consider our target parameter to be the mean of Y, the log of adjusted brain weight, defined as $Y = \log\{\text{brain weight}/(\text{body weight})^{2/3}\}$.

The data were randomly grouped into 20 sets of 3 species (2 species were randomly selected and 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 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. The population, along with the ranks assigned by the two rankers are shown in Table 2.

We assume that each of the 20 sets represents three independent draws from a large population of species. To compare the estimators with two rankers, $\hat{\mu}^{(2)}$ and $\hat{\mu}^{*(2)}$, we conducted a simulation, in which $H = 3$ and $\bar{n}$ varies from 1 to 6. In each iteration, a sample of n species was selected, with one species from each set. Figure 3 shows the relative efficiency of the two JP-S estimators along with $\pm 2s.e.$ for each average sample size. The figure shows that, in comparison with the cases of $H = 3$ in Table 1, the advantage provided by the new estimator with two rankers is substantial, especially for small sample sizes. This example establishes that the new estimator can provide an advantage even when the error-prone ranking process is a "natural" one that does not follow a known model.

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 target parameter to be the mean number of pages of the 300 theses. The page length distribution in this data set is right skewed, which can be approximated well by a gamma distribution. 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 page length difference in the two books 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 (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, we set $H = 10$ and investigated average sample sizes $\bar{n}$ ranging from 1 to 5. We also calculated the two estimators $\hat{\mu}$ and $\hat{\mu}^*$ under perfect ranking. Figure 4 plots the simulated efficiency of the two JP-S estimators in each setting.

We see from the figure that 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.

We should mention that our new estimator does not require any distributional assumptions. The improvement at small sample sizes is important in applications that require cost efficiency. Although the main focus of this paper is on point estimation, we note that interval estimates and standard errors of our estimator can be obtained through several methods. For samples of reasonably large n, we can simply produce the approximate confidence intervals based on Theorem 2, using the same asymptotic distribution as for the standard JP-S estimator, but by centering them at $\hat{\mu}^*$ instead of $\hat{\mu}$. A more promising approach would be to use bootstrapping. This entails (i) taking a sample of size q ($q \leq n$) with replacement from the data; (ii) calculating $\hat{\mu}^*$ based on the sample; (iii) repeat the above steps Q times. Then we can calculate the sample SD for $\hat{\mu}^{*(1)}, \cdots \hat{\mu}^{*(Q)}$, which can be rescaled to approximate the standard error of $\hat{\mu}^*$. Also, we can get a confidence interval from the corresponding quantiles of the Q estimates. Other methods of variance estimation include random grouping, jackknife, etc. And we refer readers to Lohr (1999) for details.

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.

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.

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

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.

Lohr, S. L. (1999). *Sampling: Design and Analysis.* New York: Duxbury.

MacEachern, S. N., Ömer Öztürk, Wolfe, D. A., and Stark, G. V. (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.

Mode, N. A., Conquest, L. L., and Marker, D. A. (1999). Ranked set sampling for ecological research: Accounting for the total costs of sampling. *Environmetrics*, 10:179–194.

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.

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.

Yu, P. L. H. and Lam, K. (1997). Regression estimator in ranked set sampling. *Biometrics*, 53:1070–1080.

Appendix: The Asymptotic Distribution

In spirit of the work by El Barmi and Mukerjee (2005), we derive the asymptotic distribution of $\hat{\mu}^*$ under the constraint $\mu_{[1]} \leq \cdots \leq \mu_{[H]}$.

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

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

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

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

and $n_g/n \to 1/H$ with probability 1 for $g = 1, 2..., H$. Then based on Slutsky's Theorem, we have

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

which also indicates that $\{\sqrt{n}(\bar{Y}_{[g]} - \mu_{[g]})\}$ are bounded in probability. Further, due to the independence of $\bar{Y}_{[g]}$'s,

$$\left(\sqrt{n}(\bar{Y}_{[1]} - \mu_{[1]}), \cdots, \sqrt{n}(\bar{Y}_{[H]} - \mu_{[H]})\right)^T \Rightarrow^w \quad \sqrt{H}\,(Z_1, \cdots, Z_H)^T, \qquad \text{(A.2)}$$

which is a $H-$variate normal distribution with independent components $Z_g \sim N(0, \sigma^2_{[g]})$.

Let

$$S_h = \{g : \mu_{[g]} = \mu_{[h]}\}, \quad h = 1, ..., H,$$

be the set of consecutive integers from $\{1, ..., H\}$ such that $\mu_{[g]} = \mu_{[h]}$ for $g \in S_h$. As

$n \to +\infty$,

$$\begin{aligned} \sqrt{n}(\mu_{[g]} - \mu_{[h]}) &\to +\infty \text{ if } \quad g > \max S_h \\ \sqrt{n}(\mu_{[g]} - \mu_{[h]}) &\to -\infty \text{ if } \quad g < \min S_h \end{aligned} \tag{A.3}$$

Because of (A.3) and the fact that $\{\sqrt{n}(\bar{Y}_{[g]} - \mu_{[g]})\}$ are bounded in probability, the r and s in the $\max_{r \le h} \min_{s \ge h}$ of (A.1) will be restricted to S_h with arbitrarily high probability for sufficiently large n. Combined with (A.2) and Slutsky's Theorem, this yields

$$\sqrt{n}(\bar{Y}^*_{[h]} - \mu_{[h]}) \to^d \quad \sqrt{H} Z^*_h$$

where

$$Z^*_h \equiv \max_{r \le h} \min_{s \ge h} \frac{\sum_{r \le g \le s;\, r,s \in S_h} Z_g}{s - r + 1}.$$

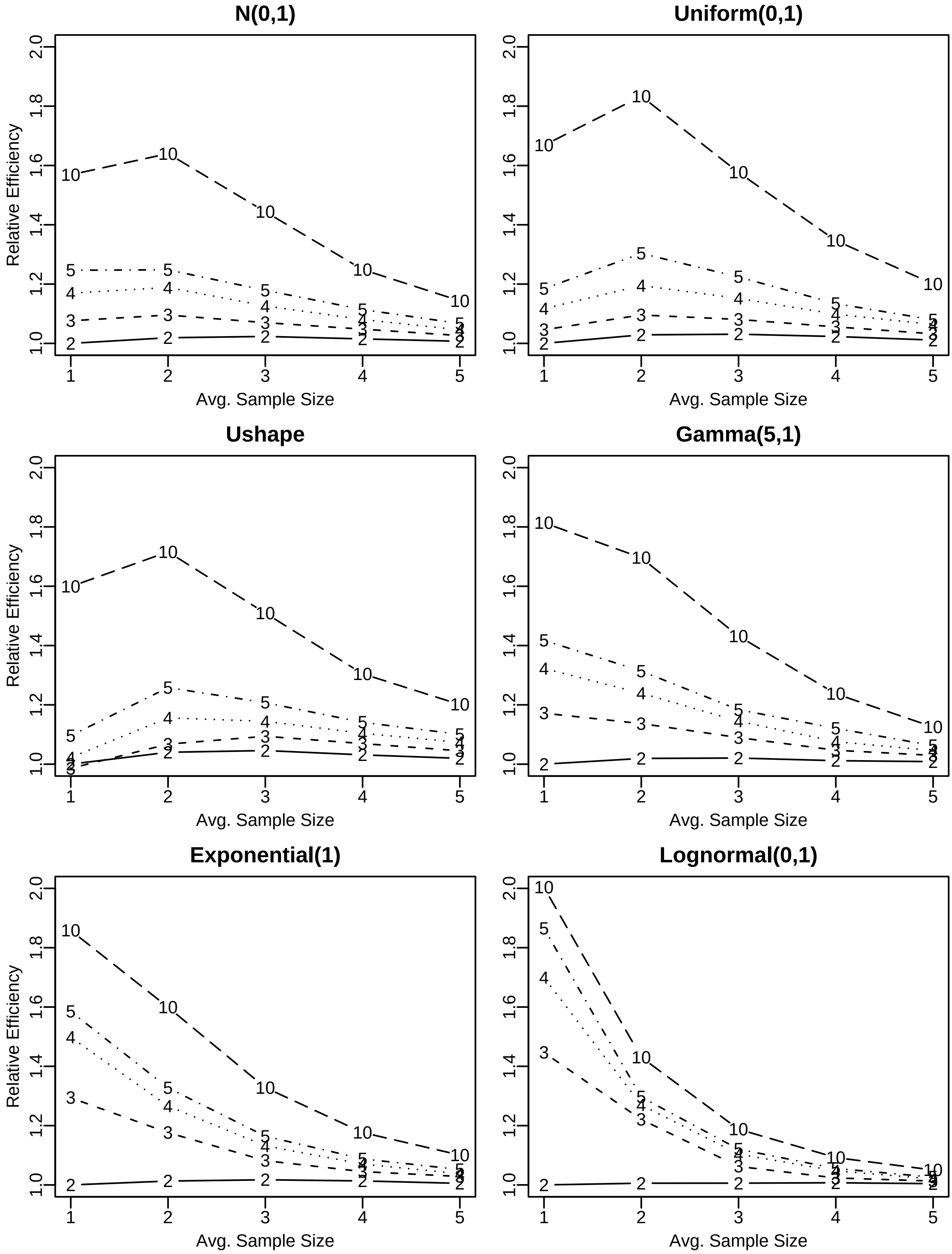
Finally, noting that $\sqrt{n}(\hat{\mu}^* - \mu) = \sum_{h=1}^{H} \sqrt{n}(\bar{Y}^*_{[h]} - \mu_{[h]})/H$, we have

$$\sqrt{n}(\hat{\mu}^* - \mu) \to^d \quad \frac{1}{\sqrt{H}} \sum_{h=1}^{H} Z^*_h.$$

Now consider the strictly increasing case $\mu_{[1]} < \cdots < \mu_{[H]}$. It is obvious that $S_h = \{h\}$, $h = 1, ..., H$. Thus, Z^*_h is reduced to Z_h, which leads to (5) in Theorem 2.

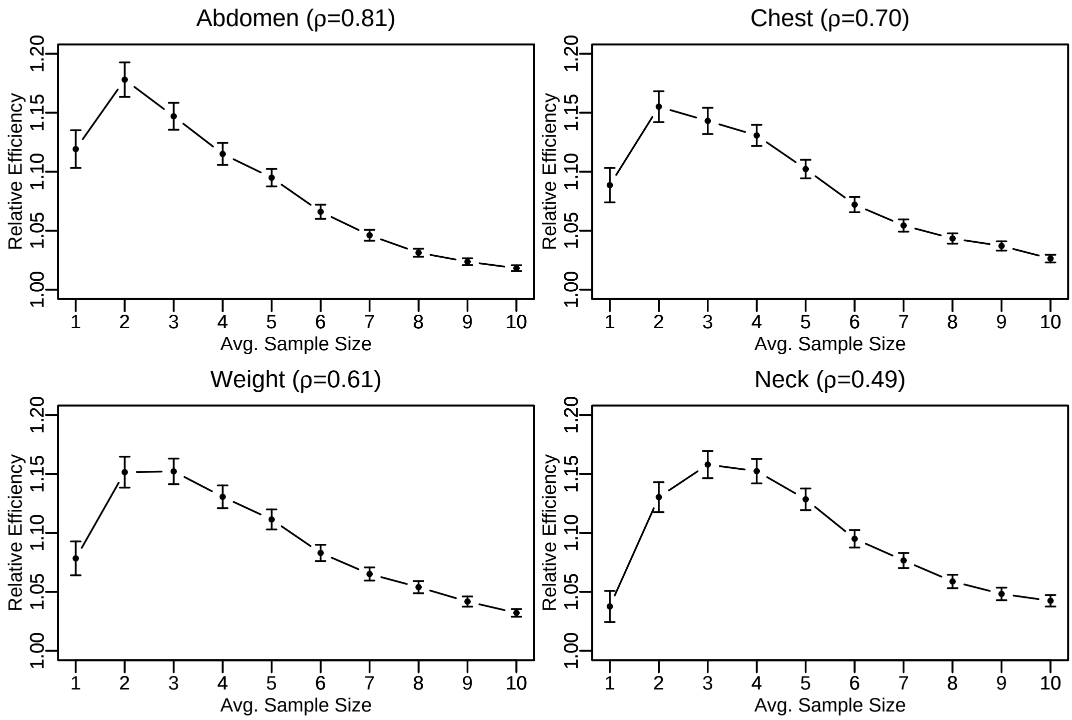
List of Figures

Figure 1:



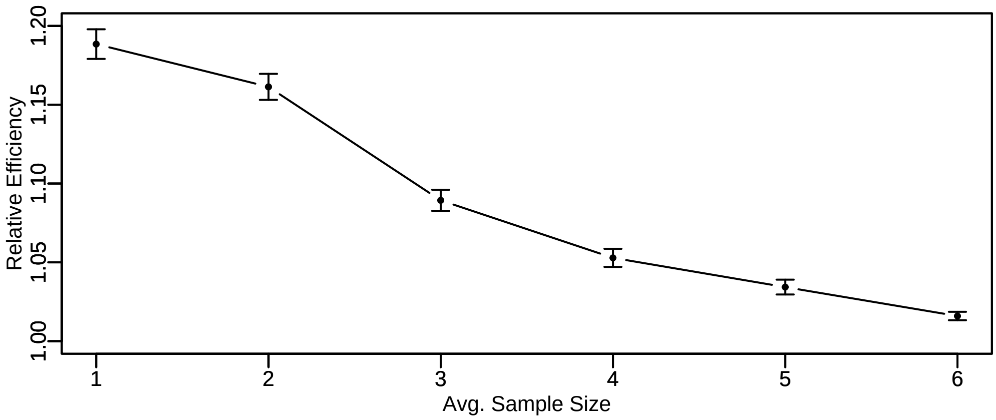
N(0,1)
Uniform(0,1)
Ushape
Gamma(5,1)
Exponential(1)
Lognormal(0,1)
Relative Efficiency
Avg. Sample Size

Figure 2:



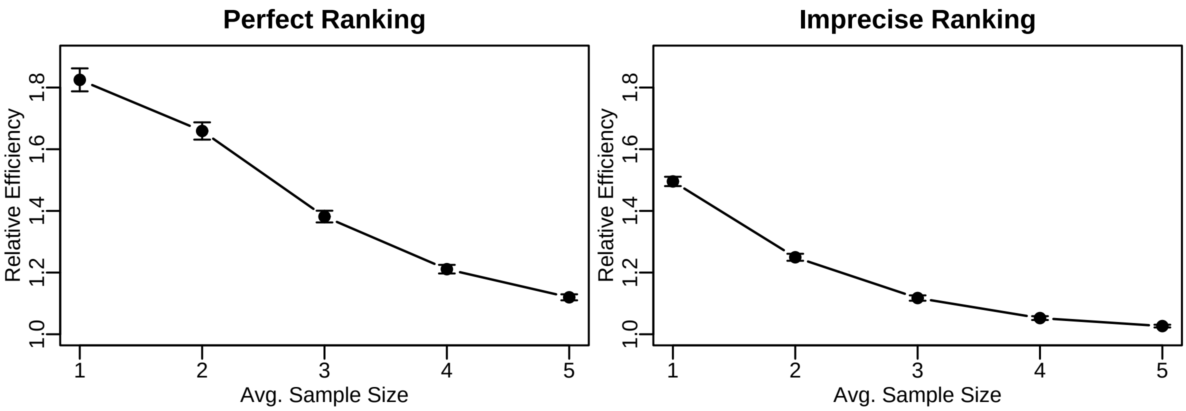
Abdomen (ρ=0.81)
Chest (ρ=0.70)
Weight (ρ=0.61)
Neck (ρ=0.49)
Relative Efficiency
Avg. Sample Size
1.00
1.05
1.10
1.15
1.20
1 2 3 4 5 6 7 8 9 10

Figure 3:



Relative Efficiency
1.00
1.05
1.10
1.15
1.20
1
2
3
4
5
6
Avg. Sample Size

Figure 4:



Perfect Ranking
Relative Efficiency
1.0
1.2
1.4
1.6
1.8
1
2
3
4
5
Avg. Sample Size
Imprecise Ranking
Relative Efficiency
1.0
1.2
1.4
1.6
1.8
1
2
3
4
5
Avg. Sample Size

Table 1: Simulated relative efficiency of the isotonic regression JP-S over the standard JP-S estimator reported with standard error

		AVG. Sample Size $\bar{n}$					
		1	2	3	4	5	20
	H	$\widehat{RE}(s.e.)$	$\widehat{RE}(s.e.)$	$\widehat{RE}(s.e.)$	$\widehat{RE}(s.e.)$	$\widehat{RE}(s.e.)$	$\widehat{RE}(s.e.)$
N(0,1)	2	1 (0)	1.02 (.00)	1.02 (.00)	1.02 (.00)	1.01 (.00)	1.00 (.00)
	3	1.08 (.01)	1.10 (.01)	1.07 (.00)	1.05 (.00)	1.03 (.00)	1.00 (.00)
	4	1.17 (.01)	1.19 (.01)	1.13 (.01)	1.08 (.00)	1.05 (.00)	1.00 (.00)
	5	1.25 (.01)	1.25 (.01)	1.18 (.01)	1.11 (.01)	1.07 (.00)	1.00 (.00)
	10	1.57 (.02)	1.64 (.02)	1.44 (.01)	1.25 (.01)	1.14 (.01)	1.00 (.00)
Uni(0,1)	2	1 (0)	1.03 (.00)	1.03 (.00)	1.02 (.00)	1.01 (.00)	1.00 (.00)
	3	1.05 (.00)	1.10 (.01)	1.08 (.00)	1.06 (.00)	1.03 (.00)	1.00 (.00)
	4	1.12 (.01)	1.20 (.01)	1.15 (.01)	1.10 (.00)	1.06 (.00)	1.00 (.00)
	5	1.18 (.01)	1.30 (.01)	1.22 (.01)	1.13 (.01)	1.08 (.00)	1.00 (.00)
	10	1.67 (.02)	1.83 (.02)	1.58 (.01)	1.35 (.01)	1.20 (.01)	1.00 (.00)
Ushape	2	1 (0)	1.04 (.00)	1.05 (.00)	1.03 (.00)	1.02 (.00)	1.00 (.00)
	3	0.99 (.00)	1.07 (.00)	1.09 (.00)	1.07 (.00)	1.05 (.00)	1.00 (.00)
	4	1.02 (.00)	1.16 (.01)	1.14 (.01)	1.10 (.00)	1.07 (.00)	1.00 (.00)
	5	1.10 (.01)	1.26 (.01)	1.21 (.01)	1.14 (.01)	1.10 (.00)	1.00 (.00)
	10	1.60 (.02)	1.72 (.02)	1.51 (.01)	1.30 (.01)	1.20 (.01)	1.00 (.00)
Gam(5,1)	2	1 (0)	1.02 (.00)	1.02 (.02)	1.01 (.00)	1.01 (.00)	1.00 (.00)
	3	1.17 (.01)	1.14 (.01)	1.09 (.00)	1.05 (.00)	1.03 (.00)	1.00 (.00)
	4	1.32 (.01)	1.24 (.01)	1.14 (.01)	1.08 (.00)	1.05 (.00)	1.00 (.00)
	5	1.42 (.01)	1.31 (.01)	1.18 (.01)	1.12 (.01)	1.06 (.00)	1.00 (.00)
	10	1.81 (.02)	1.70 (.02)	1.43 (.01)	1.24 (.01)	1.13 (.01)	1.00 (.00)
Exp(1)	2	1 (0)	1.01 (.00)	1.02 (.00)	1.01 (.00)	1.01 (.00)	1.00 (.00)
	3	1.29 (.01)	1.18 (.01)	1.08 (.00)	1.04 (.00)	1.03 (.00)	1.00 (.00)
	4	1.50 (.01)	1.27 (.01)	1.13 (.01)	1.07 (.00)	1.04 (.00)	1.00 (.00)
	5	1.59 (.02)	1.33 (.01)	1.16 (.01)	1.09 (.00)	1.05 (.00)	1.00 (.00)
	10	1.86 (.02)	1.60 (.01)	1.33 (.01)	1.18 (.01)	1.10 (.00)	1.00 (.00)
LN(0,1)	2	1 (0)	1.01 (.00)	1.01 (.00)	1.01 (.00)	1.00 (.00)	1.00 (.00)
	3	1.45 (.03)	1.22 (.04)	1.06 (.01)	1.02 (.00)	1.01 (.00)	1.00 (.00)
	4	1.70 (.04)	1.27 (.03)	1.10 (.01)	1.05 (.01)	1.02 (.00)	1.00 (.00)
	5	1.86 (.03)	1.30 (.01)	1.12 (.01)	1.06 (.00)	1.03 (.00)	1.00 (.00)
	10	2.00 (.03)	1.43 (.02)	1.19 (.01)	1.09 (.00)	1.05 (.00)	1.00 (.00)

Table 2: Data for log adjusted brain weights of mammals; in each set of 3 mammals, the ranks assigned by two rankers are available.

mammals	set	R1	R2	y	mammals	set	R1	R2	y
Genet	1	2	1	2.63	Cat	11	2	2	2.45
Rat	1	3	2	1.49	Human	11	1	1	4.43
Cow	1	1	3	1.95	Rabbit	11	3	3	1.88
African giant pouched rat	2	3	2	1.89	Artic fox	12	1	1	2.98
Kangaroo	2	2	3	1.66	Nine-banded armadillo	12	3	3	1.54
Red fox	2	1	1	2.96	Brazilian tapir	12	2	2	1.75
Lesser short-tailed shrew	3	3	3	1.57	Tree hyrax	13	3	3	2.05
Jaguar	3	1	1	1.99	Pig	13	1	2	1.69
Rock hyrax-a	3	2	2	2.70	Guinea pig	13	2	1	1.68
Baboon	4	1	1	3.62	Water opossum	14	3	3	0.53
Phalanger	4	2	2	2.11	Rhesus monkey	14	1	1	3.91
Sheep	4	3	3	2.49	African elephant	14	2	2	2.78
Gorilla	5	1	1	2.45	N.A. opossum	15	3	3	1.49
Giant armadillo	5	3	3	1.66	Roe deer	15	1	2	2.79
Yellow-bellied marmot	5	2	2	1.90	Okapi	15	2	1	2.51
Echidna	6	3	2	2.49	Donkey	16	1	2	2.55
Owl monkey	6	1	1	3.23	Mountian beaver	16	3	3	1.89
Giraffe	6	2	3	2.34	Horse	16	2	1	2.31
Grey wolf	7	1	1	2.39	Tree shrew	17	3	3	2.43
Big brown bat	7	3	3	1.31	Galago	17	1	1	2.68
Goat	7	2	2	2.53	Golden hamster	17	2	2	1.41
Little brown bat	8	1	3	1.68	European hedgehog	18	3	3	1.41
Tenrec	8	2	1	1.03	Ground squirrel	18	1	1	2.91
Desert hedgehog	8	3	2	1.27	Rock hyrax-b	18	2	2	2.19
Asian elephant	9	3	1	3.21	Raccoon	19	1	2	2.70
E. American mole	9	2	2	1.91	Mouse	19	2	1	1.60
Verbet	9	1	3	3.11	Musk shrew	19	3	3	0.92
Chinchilla	10	2	2	2.43	Star-nosed mole	20	3	3	1.88
Grey seal	10	1	3	2.82	Slow loris	20	2	2	2.30
Mole rat	10	3	1	2.50	Patas monkey	20	1	1	3.21