Ann. Inst. Statist. Math.
Vol. 52, No. 2, 239–254 (2000)

OPTIMAL ALLOCATION FOR SYMMETRIC DISTRIBUTIONS IN RANKED SET SAMPLING*

AMARJOT KAUR, G. P. PATIL AND C. TAILLIE

Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park, PA 16802, U.S.A.

(Received August 5, 1996; revised October 22, 1998)

Abstract. Ranked set sampling (RSS) is a cost efficient method of sampling that provides a more precise estimator of population mean than *simple random sampling*. The benefits due to ranked set sampling further increase when appropriate allocation of sampling units is made. For highly skew distributions, allocation based on the Neyman criterion achieves a substantial precision gain over equal allocation. But the same is not true for symmetric distributions; in fact, the gains due to using the Neyman allocation are typically very marginal for symmetric distributions. This paper, determines optimal RSS allocations for two classes of symmetric distributions. Depending upon the class, the optimal allocation assigns all measurements either to the extreme ranks or to the middle rank(s). This allocation outperforms both equal and Neyman allocations in terms of the precision of the estimator which remains unbiased. The two classes of distributions are distinguished by different growth patterns in the variance of their order statistics regarded as a function of the rank order. For one class, the variance peaks for middle rank orders and tapers off in the tails; for the other class, the variance peaks for the two extreme rank orders and tapers off toward the middle. Kurtosis appears to effectively discriminate between the two classes of symmetic distributions. The Neyman allocation is required to quantify all rank orders at least once (to ensure general unbiasedness) but then quantifies most frequently the more variable rank orders. Under symmetry, unbiasedness can be obtained without quantifying all rank orders and the optimal allocation quantifies the least variable rank order(s), resulting in a high precision estimator.

Key words and phrases: Equal allocation, kurtosis, Neyman allocation, order statistics, relative precision, skewness, symmetry.

1. Introduction

It is well established that under equal allocation, ranked set sampling (RSS) is a more precise method of sampling than simple random sampling (SRS) in estimating the population mean. McIntyre (1952) first recognized the potential of RSS in estimation of the herbage mass. Halls and Dell (1966) formalized this concept and coined the term RSS. Takahasi and Wakimoto (1968) established a rigorous foundation for the theory of RSS, and thereafter various facets of RSS have been discussed in the literature. Dell and Clutter (1972) examined the effect of ranking error, Stokes (1980*a*,*b*) discussed

* Prepared with partial support from the Statistical Analysis and Computing Branch, Environmental Statistics and Information Division, Office of Policy, Planning, and Evaluation, United States Environmental Protection Agency, Washington, DC under a Cooperative Agreement Number CR-821531. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

RSS in estimating the population variance and estimation of correlation coefficient. The estimation of distribution function using RSS was considered by Stokes and Sager (1988), along with its application in estimating the tree volume. Cobby *et al.* (1985) provide several applications of RSS in agriculture. See Patil *et al.* (1994) and Kaur *et al.* (1996) for an overview of RSS literature.

The performance of RSS can be further improved by using an appropriate allocation. McIntyre (1952) suggested allocation proportional to the standard deviations of rank order statistics (Neyman allocation) to maximize precision in estimating population mean. When the standard deviations of the order statistics are unknown, Kaur *et al.* (1997) describe some allocation rules based on the knowledge of skewness, kurtosis, or coefficient of variation of the underlying distribution. In general, RSS estimator of mean can be expressed as a weighted average of rank order statistics, and under equal allocation all weights are the same. Equal allocation is most appropriate under complete lack of underlying distribution, but when there is some knowledge of the form of the underlying distribution, appropriate weights can be chosen to obtain the best linear unbiased estimator of the mean. Kvam and Samaneigo (1993) established inadmissibility of the equally allocated RSS estimator of population mean when it is either the location or scale parameter of the underlying distribution.

For symmetric distributions, the gains due to Neyman allocation instead of equal allocation are usually modest. Yanagawa and Shirahata (1976) proposed a minimum variance linear unbiased median-mean estimator of population mean for a family of symmetric distributions. Shirahata (1982) further examined more general procedures that are unbiased for symmetric distributions. In this paper, we consider allocation models for symmetric distributions by exploiting the type of symmetry and provide the minimum variance linear unbiased estimator of the population mean. Two classes of symmetric distributions, characterized by the qualitative behavior of the variances of their order statistics, are considered.

Section 2 reviews the RSS procedure under equal and unequal allocation. In Section 3, the optimal allocation for symmetric distributions is discussed. Section 4 gives examples of the two classes of symmetric distributions and compares the performance of the optimal allocation with equal and Neyman allocation. Finally, the role of kurtosis in determining the precision of the proposed estimator and also in characterizing the type of symmetry is discussed in Section 5.

2. Ranked set sampling

The procedure to obtain a ranked set sample of size m with equal allocation involves randomly drawing m^2 units from the population and then randomly partitioning them into m equal sized subsets. The units are then ranked within each subset. Here ranking could be based on judgment, visual perception, covariates, or any other method that does not require actual measurement of the units. The unit receiving the smallest rank is quantified from the first set, the unit receiving the second smallest rank is quantified from the second set, and so forth until the unit with the largest rank is quantified from the m-th set. Thus m units are quantified out of m^2 selected originally. This procedure is repeated r times (cycles) in order to get $n = mr$ quantifications.

Let $X_{(i:m)j}$, $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, r$, denote the quantification of the i-th rank order in the j-th cycle. For fixed i, the $X_{(i:m)j}$, $j = 1, 2, \ldots, r$, are iid with $E(X_{(i:m)j}) = \mu_{(i:m)}$ and $\text{var}(X_{(i:m)j}) = \sigma^2_{(i:m)}$.

Let μ and σ^2 be the mean and variance of the population. The ranked set sample mean given as

$$\bar{X}_{(m)\text{eq}} = \frac{1}{mr}\sum_{i=1}^{m}\sum_{j=1}^{r} X_{(i:m)j},$$

is an unbiased estimator of μ, having variance

$$\text{var}(\bar{X}_{(m)\text{eq}}) = \frac{1}{m^2 r}\sum_{i=1}^{m}\sigma^2_{(i:m)}.$$

The subscript "eq" signifies equal allocation.

Under SRS with $n = mr$ quantifications the variance of the mean is σ^2/n, so the relative precision of RSS with respect to SRS is given as

$$\begin{aligned}\text{RP}_{\text{eq}} &= \frac{\sigma^2/n}{\text{var}(\bar{X}_{(m)\text{eq}})}\\ &= \frac{\sigma^2}{\frac{1}{m}\sum_{i=1}^{m}\sigma^2_{(i:m)}} = \frac{\sigma^2}{\overline{\sigma^2}},\end{aligned}$$

where $\overline{\sigma^2} = \frac{1}{m}\sum_{1=1}^{m}\sigma^2_{(i:m)}$ is the average within-rank variance.

Under the scenario of general allocation, n sets each containing m sampling units are ranked and the i-th rank order is quantified from r_i of the sets, $i = 1,\ldots,m$. This gives $n = r_1+\cdots+r_m$ quantifications denoted by $X_{(i:m)j}$, $i = 1,2,\ldots,m$; $j = 1,2,\ldots,r_i$. When all r_i are positive, an unbiased estimator of the population mean is

$$\bar{X}_{(m)\text{ueq}} = \frac{1}{m}\sum_{i=1}^{m}\frac{T_i}{r_i},$$

where

$$T_i = \sum_{j=1}^{r_i} X_{(i:m)j}.$$

The variance of this estimator is

$$\text{var}(\bar{X}_{(m)\text{ueq}}) = \frac{1}{m^2}\sum_{i=1}^{m}\frac{\sigma^2_{(i:m)}}{r_i}. \tag{2.1}$$

Still assuming that $r_i > 0$, the optimal Neyman allocation is

$$r_i = \frac{n\sigma_{(i:m)}}{\sum_{i=1}^{m}\sigma_{(i:m)}}. \tag{2.2}$$

Using (2.2) in (2.1) yields

$$\text{var}(\bar{X}_{(m)\text{ney}}) = \frac{1}{nm^2}\left(\sum_{i=1}^{m}\sigma_{(i:m)}\right)^2 = \frac{\bar{\sigma}^2}{n}, \tag{2.3}$$

where $\bar{\sigma} = \frac{1}{m}\sum_{i=1}^{m}\sigma_{(i:m)}$ is the average within-rank standard deviation.

Note that equation (2.2) does not generally yield integral values for the r_i so that some adjustments are needed to obtain the optimal allocation. These adjustments have the effect of increasing the variance above the value given by (2.3). Thus (2.3) is only an approximate (lower bound) expression for the variance of the Neyman allocation; however, it is asymptotically correct as $n \to \infty$. This point becomes important for the comparisons of the next section where we obtain the exact (integral) optimal allocation for symmetric distributions. Compared with SRS, the asymptotic relative precision of the Neyman allocation is

$$\mathrm{RP}_{\mathrm{ney}} = \frac{\sigma^2/n}{\operatorname{var}(\bar{X}_{(m)\mathrm{ney}})} = \left(\frac{\sigma}{\bar{\sigma}}\right)^2.$$

3. Allocation models for symmetric distributions

Neyman's formula provides an optimal allocation; nonetheless, for symmetric distributions, there are allocations yielding even better performance than the Neyman allocation. This can be explained by noting that optimality of the Neyman allocation is established under the constraint that all r_i be positive. The constraint is imposed because, in general, there does not exist a linear combination of the $X_{(i:m)j}$ that is unbiased for μ unless each rank order is quantified at least once, i.e., unless each r_i is positive.

In the case of symmetric distributions, the symmetry can be exploited to allow some (in fact, most) of the r_i to vanish and thereby to broaden the class of linear unbiased estimators of μ. The resulting optimal allocation strategy is precisely the opposite of the Neyman strategy which quantifies most heavily those rank orders having the largest variances. For symmetric distributions, the strategy is to ignore the rank orders with large variances and to quantify only the rank orders having the smallest variances. Since the estimator is constructed from variates that already have a small variance, the performance improvement over the Neyman allocation can be quite large. Along these lines, David ((1981), pp. 138–140) discuss the use of the trimmed mean and the two point mean for estimating the location parameter of the normal distribution. Another estimator of location is the midpoint or midrange $(X_{(1:m)} + X_{(m:m)})/2$, which is optimal for uniform population. David and Groeneveld (1982) showed the asymptotic variance of the middle rank order statistics to be either locally minimum or maximum for a symmetric distribution.

Let $M = (m+1)/2$ so that M is the unique middle rank order when m is odd. If m is even, then M is not an integer and the two middle rank orders are $M - \frac{1}{2}$ and $M + \frac{1}{2}$.

For a symmetric distribution, we have that

$$\mu = \frac{1}{2}(\mu_{(i:m)} + \mu_{(m-i+1:m)}), \qquad 1 \le i < M,$$

and

$$\mu = \mu_{(i:m)} \quad \text{if} \quad m \text{ is odd and } i = M.$$

Accordingly,

$$\hat{\mu}_i(r_i, r_{m-i+1}) = \frac{1}{2}\left(\frac{T_i}{r_i} + \frac{T_{m-i+1}}{r_{m-i+1}}\right), \qquad 1 \le i < M, \tag{3.1}$$

is an unbiased estimator of μ provided r_i and r_{m-i+1} are both positive. In what follows, it will be convenient to write

$$r_i^* = r_{m-i+1}, \qquad 1 \le i \le M.$$

If m is odd and $i = M$, we also have the unbiased estimator

$$\hat{\mu}_M(r_M) = \hat{\mu}_M(r_M, r_M) = \frac{T_M}{r_M},$$

as long as r_M is positive. The estimator defined by equation (3.1) has its variance given by

$$\mathrm{var}(\hat{\mu}_i(r_i, r_i^*)) = \frac{\sigma^2_{(i:m)}}{4}\left(\frac{1}{r_i} + \frac{1}{r_i^*}\right), \qquad 1 \le i < M,$$

since $\sigma^2_{(i:m)} = \sigma^2_{(m-i+1:m)}$. Similarly,

$$\mathrm{var}(\hat{\mu}_M(r_M)) = \frac{\sigma^2_{(M:m)}}{r_M}. \tag{3.2}$$

When $n_i = r_i + r_i^*$ is fixed, the variance of $\hat{\mu}_i(r_i, r_i^*)$ is minimized by making the allocation as nearly balanced as possible, e.g.,

$$r_i = \mathrm{floor}(n_i/2) \qquad \text{and} \qquad r_i^* = \mathrm{ceil}(n_i/2).$$

We write $\hat{\mu}_i(n_i)$, $1 \le i < M$, for the corresponding estimator. One finds that

$$\mathrm{var}(\hat{\mu}_i(n_i)) = \begin{cases} \dfrac{\sigma^2_{(i:m)}}{n_i} & \text{if } \; n_i \; \text{ is even} \\[2ex] \dfrac{n_i^2}{n_i^2 - 1}\dfrac{\sigma^2_{(i:m)}}{n_i} & \text{if } \; n_i \; \text{ is odd,} \end{cases} \tag{3.3}$$

where $1 \le i < M$.

Let the number n of possible measurements be fixed. Then, an allocation $(r_1, r_2, \ldots, r_m)$ with $n = r_1 + r_2 + \cdots + r_m$ is said to be *G-allowable* if each r_i is positive. The letter "G" refers to general distribution, since G-allowability is a necessary and sufficient condition for the existence of a linear unbiased estimator of μ for arbitrary distributions (with finite first moments). The allocation $(r_1, r_2, \ldots, r_m)$ is *S-allowable* if, for each i with $1 \le i < M$, either r_i and r_i^* are both positive or both vanish. Note that every G-allowable allocation is S-allowable. For symmetric distributions and S-allowable allocations, the linear unbiased estimators $\hat{\mu}_i$ defined above are available and we can incorporate all the observations by considering estimators of the form

$$\sum_{1 \le i \le M}{}' W_i \hat{\mu}_i(r_i, r_i^*) \quad \text{with} \quad \sum_{1 \le i \le M}{}' W_i = 1, \tag{3.4}$$

where the prime on the summation signifies that the summation extends over those i for which $r_i > 0$.

Note that S-allowability is not necessary for the existence of a linear estimator of μ that is unbiased for all symmetric distributions. For example, such an estimator exists whenever there is at least one i for which r_i and r_i^* are both positive. However, S-allowability does appear to be necessary if all observations are to appear in the linear combination. A closely related question asks, for symmetric distributions and S-allowable allocations, if every linear unbiased estimator of μ has variance greater than or equal to the variance of an estimator of the form (3.4), i.e., does the class (3.4) include all admissible linear unbiased estimators.

Here, we restrict attention to the estimators of form (3.4) and determine the S-allowable allocation that minimizes the variance within this class of estimators. There are several cases, depending upon the parity of m and n:

Case 1. If n or m or both are even, then let j be the rank order i that minimizes $\sigma^2_{(i:m)}$ for $1 \le i \le M$.

(a) If $j < M$, then the optimal allocation has $r_j = \text{floor}(n/2)$, $r_j^* = \text{ceil}(n/2)$, and all other r_i vanish. The optimal estimator is $\hat{\mu}_j(n)$ whose variance is given by equation (3.3).

(b) If $j = M$, then the optimal allocation has $r_M = n$ and all other r_i vanish. The optimal estimator is $\hat{\mu}_M(n)$ whose variance is given by equation (3.2).

Case 2. If n and m are both odd, then let j be the rank order i that minimizes $\sigma^2_{(i:m)}$ for $1 \le i < M$. Note that $i = M$ is excluded from this minimization.

(a) If $\sigma^2_{(j:m)} \le \frac{n-1}{n}\sigma^2_{(M:m)}$ then the optimal allocation and estimator are as in Case 1(a).

(b) If $\sigma^2_{(M:m)} \le \sigma^2_{(j:m)}$, then the optimal allocation and estimator are as in Case 1(b).

(c) If $\frac{n-1}{n}\sigma^2_{(M:m)} < \sigma^2_{(j:m)} < \sigma^2_{(M:m)}$, then the optimal allocation has $r_j = r_j^* = (n-1)/2$, $r_M = 1$, and all other r_i vanish. The optimal estimator is

$$W\hat{\mu}_j(n-1) + (1-W)\hat{\mu}_M(1)$$

with

$$W = \frac{(n-1)/\sigma^2_{(j:m)}}{(n-1)/\sigma^2_{(j:m)} + 1/\sigma^2_{(M:m)}}$$

whose variance is

$$\frac{1}{(n-1)/\sigma^2_{(j:m)} + 1/\sigma^2_{(M:m)}}.$$

The proof of this result appears in the Appendix. The Case 2(c) is a bit anomalous, but the prescription for the other cases is simple and intuitive: determine the rank order j whose order statistic has the smallest variance,

$$\sigma^2_{\min} = \min\{\sigma^2_{(i:m)} : 1 \le i \le M\},$$

and allocate all the measurements to this rank order and to the symmetric rank order $m - j + 1$, in as balanced a manner as possible. The optimal variance is then

$$\sigma^2_{\text{sym}} = \begin{cases} \dfrac{\sigma^2_{\min}}{n} & \text{if } \; j = M \; \text{ or } \; n \text{ is even} \\ \dfrac{n^2}{n^2-1}\dfrac{\sigma^2_{\min}}{n} & \text{if } \; j < M \; \text{ and } \; n \text{ is odd.} \end{cases}$$

Note that

$$\frac{\sigma^2_{\min}}{n} \le \sigma^2_{\text{sym}} \le \frac{n^2}{n^2-1}\frac{\sigma^2_{\min}}{n}. \tag{3.5}$$

It is easy to see that these inequalities hold for Case 2(c) also. This gives the asymptotic expression

$$\sigma^2_{\text{sym}} = \frac{\sigma^2_{\min}}{n} \quad \text{as} \quad n \to \infty, \tag{3.6}$$

analogous to equation (2.3) for the Neyman allocation. The asymptotic expression (3.6) is exact for finite n if either (i) n is even or (ii) m is odd and the middle rank order M minimizes $\sigma^2_{(i:m)}$ for $1 \le i \le M$. Otherwise, the asymptotic expression provides a slight underestimate for σ^2_{sym}, but still an excellent approximation since the upper bound in (3.5) approaches $\sigma^2_{\min}/n$ very rapidly as n gets large. For example, the upper bound is $1.01\sigma^2_{\min}/n$ for n as small as 10.

If the above prescription were applied to the situation of Case 2(c), then all n observations would be allocated to rank orders j and $m-j+1$. But this would yield an unbalanced allocation since n is odd. If $\sigma^2_{(M:m)}$ is only slightly larger than $\sigma^2_{(j:m)}$ then it is better to balance the allocation and assign the extra remaining measurement to rank order M. Observe that Case 2(c) disappears in the limit of large n.

Recall that every G-allowable allocation is S-allowable. Since the Neyman allocation is the optimal G-allowable allocation, it follows that the precision of the optimal S-allowable allocation must be at least as good as that of the Neyman allocation for symmetric distributions. Numerical comparison is easiest using the asymptotic variances, for which the precision of the S-optimal allocation relative to the Neyman allocation is

$$\text{RP} = \frac{\bar{\sigma}^2}{\sigma^2_{\min}} \ge 1.$$

At a practical level, determination of the S-optimal allocation requires knowledge of the order statistic variances. In reality, these variances are not known. However, we have found for many distributions that the plot of $\sigma^2_{(i:m)}$ versus i is either mound-shaped or U-shaped. In the first case, the minimal variance occurs in the tails ($j = 1$) and, in the second case, it occurs in the middle ($j = \text{floor}(M)$). Thus, only a qualitative judgment regarding the two shapes is needed to arrive at the optimal S-allocation. The next section studies a number of examples to develop an intuitive feel for helping to make this judgment.

More complicated multimodal shapes are possible. But even here, we often find that the appropriate choice of either $j = 1$ or $j = \text{floor}(M)$ yields a good approximation to $\sigma^2_{\min}$.

4. Examples

A symmetric distribution will be said to belong to the family $\mathcal{F}_{S_1} = \mathcal{F}_{S_1}(m)$ if the plot of $\sigma^2_{(i:m)}$ versus i is mound shaped, i.e.,

$$\sigma^2_{(i:m)} \text{ is increasing in } i \quad \text{for} \quad 1 \le i \le M.$$

For simplicity, we consider only the asymptotic case of large n. The smallest order statistic variance occurs for $j = 1$ and, according to the previous section, the optimal allocation assigns floor($n/2$) measurements to rank order 1 and ceil($n/2$) measurements to rank order m. We call this the "extreme" rank orders allocation and indicate it by a subscript "ext" For large n, the asymptotic variance of the estimator is

$$\sigma^2_{\text{ext}} = \frac{\sigma^2_{(1:m)}}{n}. \tag{4.1}$$

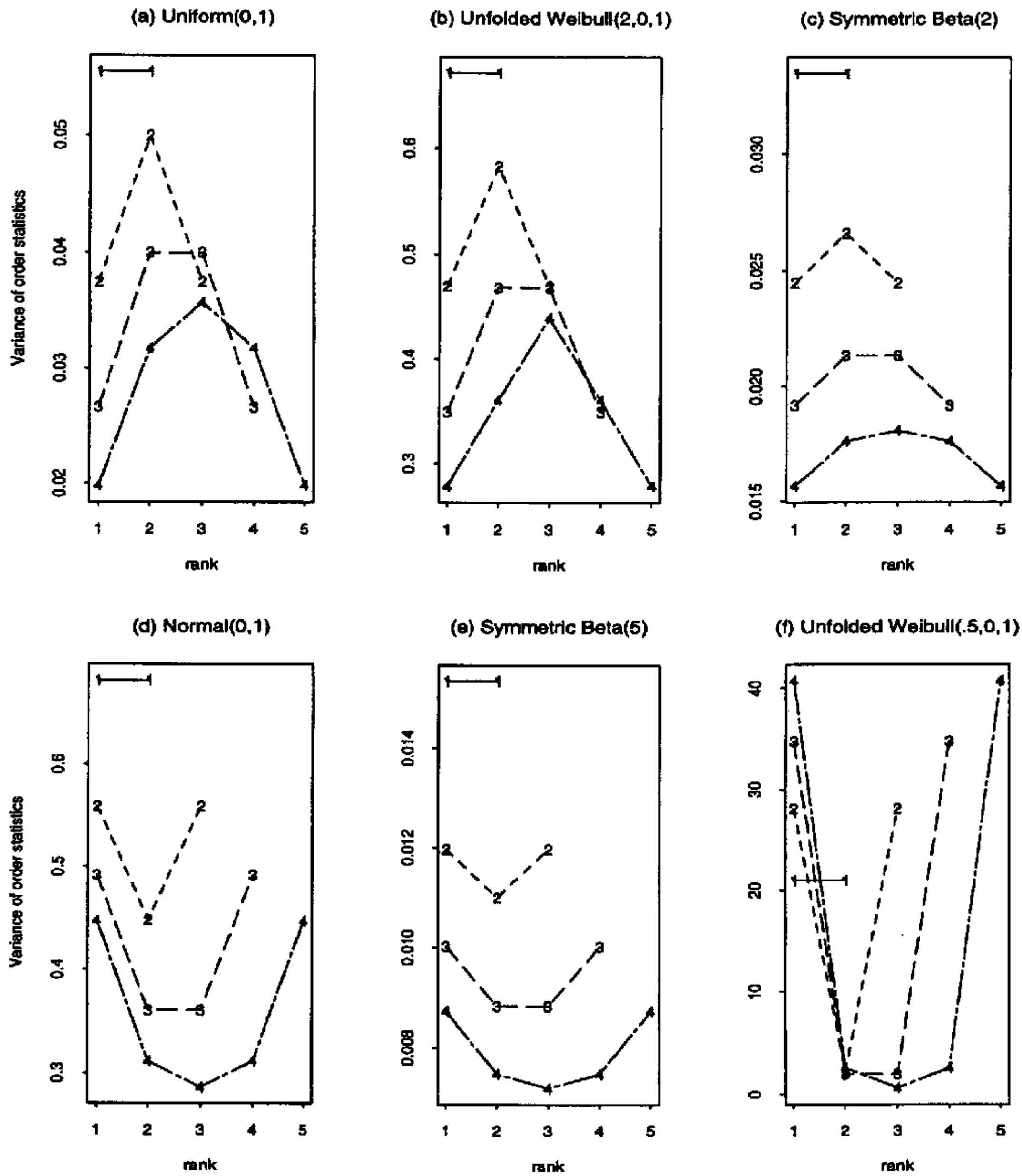

Fig. 1. Variances of order statistics of Normal, Uniform, Symmetric Beta and Unfolded Weibull distributions. The line types 1,2,3,4 correspond to the set sizes $m = 2, 3, 4, 5$ respectively.

The family $\mathcal{F}_{S_1}$ includes the Uniform and Sine distributions as well as certain members of the Unfolded Weibull and of the Symmetric Beta families as shown in Fig. 1(a–c). A random variable X follows Unfolded Weibull Distribution $(\alpha, 0, 1)$, if symmetric about $x = 0$ and $|X|$ follows the Weibull Distribution $(\alpha, 1)$; having pdf $\alpha|x|^{(\alpha-1)}e^{-|x|^{\alpha}}$, for $-\infty < x < \infty$. A special peculiarity of this distribution is that when $|X|$ follows J-shaped Weibull, then its unfolded version X may have a distribution with very high kurtosis. In Symmetric Beta distribution with parameter α, bounded range and the shape parameter control the type of symmetry, i.e., bell of U-shaped.

The asymptotic relative precision of the extreme rank orders allocation compared

Table 1. Relative precisions $\mathrm{RP}_{\mathrm{eq}}$, $\mathrm{RP}_{\mathrm{ney}}$ and $\mathrm{RP}_{\mathrm{ext}}$ of some selected distributions for the set sizes 2 through 5.

Distribution	set size	$\mathrm{RP}_{\mathrm{eq}}$	$\mathrm{RP}_{\mathrm{ney}}$	$\mathrm{RP}_{\mathrm{ext}}$
Uniform (0,1)	m=2	1.50	1.50	1.50
Kurtosis= 1.8	3	2.00	2.00	2.22
	4	2.50	2.53	3.12
	5	3.00	3.05	4.20
Sine $(0,\pi)$	m=2	1.49	1.49	1.49
Kurtosis=2.19	3	1.98	1.98	2.02
	4	2.46	2.46	2.55
	5	2.95	2.95	3.09
Unfolded Weibull (2,0,1)	m=2	1.49	1.49	1.49
Kurtosis=2	3	1.97	1.98	2.13
	4	2.44	2.46	2.86
	5	2.91	2.93	3.59
Symmetric Beta (2)	m=2	1.49	1.49	1.49
Kurtosis=2.14	3	1.98	1.98	2.04
	4	2.47	2.47	2.61
	5	2.96	2.96	3.19

with SRS is obtained from equation (4.1) as

$$\mathrm{RP}_{\mathrm{ext}} = \left(\frac{\sigma}{\sigma_{(1:m)}}\right)^2.$$

The relative precisions $\mathrm{RP}_{\mathrm{eq}}$, $\mathrm{RP}_{\mathrm{ney}}$ and $\mathrm{RP}_{\mathrm{ext}}$ are computed for the Uniform (0,1) distribution, the Sine $(0,\pi)$ distribution, an Unfolded Weibull distribution, and a Symmetric Beta distribution in Table 1. It is seen that the performance of equal and Neyman allocation is very close but the extreme allocation model performs better than both. Further, we note that the relative precision in all the three cases increases with m. A symmetric distribution will be said to belong to the family $\mathcal{F}_{S_2} = \mathcal{F}_{S_2}(m)$ if the plot of $\sigma^2_{(i:m)}$ versus i is U-shaped, i.e.,

$$\sigma^2_{(i:m)} \text{ is decreasing} \quad \text{for} \quad 1 \le i \le M.$$

The optimal allocation depends upon whether m is even or odd. For even m, the smallest order statistic variance occurs for the two middle rank orders, $M-\frac{1}{2}$ and $M+\frac{1}{2}$. The asymptotic optimal allocation assigns floor$(n/2)$ measurements to the first of these rank orders and ceil$(n/2)$ measurements to the second. For odd m, the unique middle rank order is M and the asymptotic optimal allocation assigns all n measurements to this rank order. We call this the middle rank order(s) allocation. The asymptotic variance of the estimator is given by

$$\sigma^2_{\mathrm{mid}} = \frac{\sigma^2_{(L:m)}}{n},$$

where L is the floor of M. The relative precision compared with SRS is $\mathrm{RP}_{\mathrm{mid}} = \{\sigma/\sigma_{(L:m)}\}^2$.

The $\mathcal{F}_{S_2}$ family includes the Normal, Logistic, Laplace, and Triangular distributions, as well as certain members of the Unfolded Weibull family and of the Symmetric Beta family as shown in Fig. 1(d–f).

Table 2. Relative precisions RP_{eq}, RP_{ney} and RP_{mid} of some selected distributions for the set sizes 2 through 5.

Distribution	set size	RP_{eq}	RP_{ney}	RP_{mid}
Normal (0,1)	m=2	1.47	1.47	1.47
Kurtosis=3	3	1.91	1.92	2.23
	4	2.35	2.36	2.77
	5	2.77	2.80	3.49
Logistic (0,1)	m=2	1.44	1.44	1.44
Kurtosis=4.2	3	1.84	1.86	2.55
	4	2.22	2.27	3.16
	5	2.58	2.67	4.17
Triangular $(-1, 1)$	m=2	1.49	1.49	1.49
Kurtosis=2.4	3	1.96	1.96	2.03
	4	2.43	2.43	2.53
	5	2.90	2.90	3.13
Laplace (0,1)	m=2	1.39	1.39	1.39
Kurtosis=6	3	1.73	1.78	3.13
	4	2.04	2.16	3.84
	5	2.33	2.54	5.70
Unfolded Weibull(.5,0,1)	m=2	1.15	1.15	1.15
Kurtosis=70	3	1.24	1.49	11.42
	4	1.31	1.82	12.72
	5	1.38	2.15	41.43
Symmetric Beta(5)	m=2	1.48	1.48	1.48
Kurtosis=2.53	3	1.95	1.95	2.07
	4	2.41	2.41	2.57
	5	2.87	2.87	3.15

The relative precisions RP_{eq}, RP_{ney} and RP_{mid} are computed for the Normal (0,1), Logistic (0,1), Triangular $(-1, 1)$, Laplace (0,1), Unfolded Weibull (.5,0,1) and Symmetric Beta (5) distributions for $m = 2, 3, 4, 5$ in Table 2. It is seen that the performance of equal and Neyman allocation is almost the same but the middle rank order allocation outperforms both. In case of unfolded weibull (.5,0,1) distribution, high relative precision of 41.43 is due to very sharp form of density function at the origin which results in a large kurtosis value of 70. Consequently, the variability in the middle is very small thus yielding high precision for middle allocation.

When the assumption of symmetry is violated, then the bias is introduced in the estimator of mean. If the allocation is made incorrectly at the middle rank instead of correct exterme rank order statistics, or vice-versa, then the magnitude of loss is directly proportional to the difference in variability at extreme and middle order statistics. Kurtosis can be a relevant measure of this difference, as its increasing values indicate smaller variability in the middle as compared to extremes. Further, ranking errors undermine somewhat the benefits of RSS.

5. Role of kurtosis

The general RSS procedure is distribution-free, requiring only finiteness of the low order moments of the parent distribution. Nonetheless, even going back to McIntyre

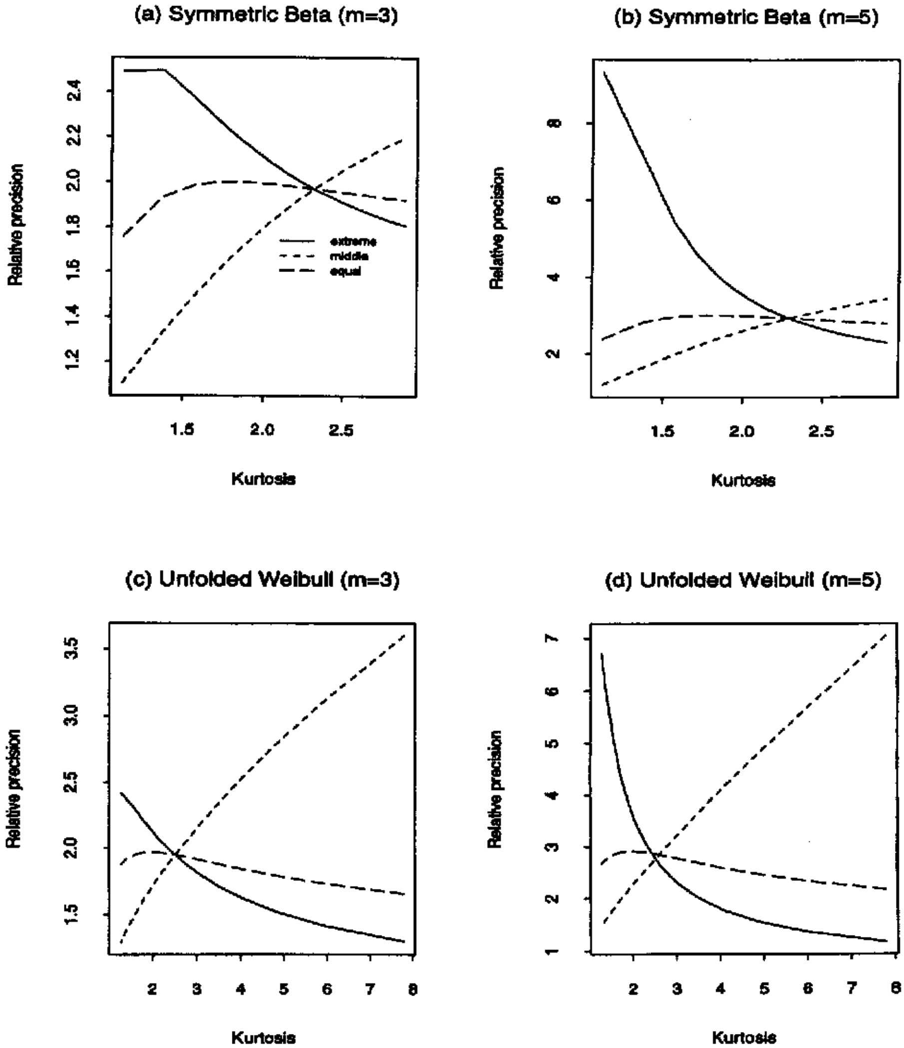

Fig. 2. Relative precision RP_{ext}, RP_{mid}, RP_{eq} versus kurtosis for Symmetric Beta and Unfolded Weibull distributions for $m = 3, 5$.

(1952), one tries to reference the behavior of RSS procedures by summary distributional characteristics like skewness and kurtosis. Since the procedures studied in this paper are limited to symmetric distributions where skewness is noninformative, we have studied kurtosis as a referencing characteristic.

Figure 2 plots the relative precision of middle, extreme, and equal allocations against kurtosis for two distributional families (symmetric beta and unfolded Weibull) and two set sizes ($m = 3$ and $m = 5$). For all four combinations, relative precision of the extreme allocation is a steadily declining function of kurtosis. This can be explained by the hign variability of the extreme order statistics when kurtosis is large. Conversely, relative precision of the middle allocation increases with increasing kurtosis. Large kurtosis is associated with a bunching up of observations in the center of the distribution and this limits the variability of the middle order statistic and results in improved performance of the middle allocation. The intersection of the RP curves for middle and extreme allocations occurs when kurtosis is about 2.3 for symmetric beta family and about 2.5 for unfolded Weibull family. The nearness of these values suggests that kurtosis may be an effective discriminator between the two classes of symmetric distributions considered in this paper. We have not examined this question for other families of distributions, however.

Compared with middle and extreme allocations, the RP curve for equal allocation is comparatively flat (Fig. 2). Interestingly, the figure also indicates that the intersection points for the three curves are nearly concident. Apparently, when the optimality changeover from extreme to middle allocation occurs, all of the order statistics have about the same variance and, consequently, all allocations perform about the same.

Appendix

Here, we derive the optimal S-allocation described in Section 3. Some preliminary results are needed. Result 1 is well-known, but we make repeated use of the equation (A.1) for the optimal variances.

Result 1. Let $Z_1, \ldots, Z_p$ be independent random variables with a common mean μ and with variances $\sigma_1^2, \ldots, \sigma_p^2$. The linear combination

$$W_1 Z_1 + W_2 Z_2 + \cdots + W_p Z_p$$

with $W_1 + W_2 + \cdots + W_p = 1$ that has the smallest variance is obtained by taking W_i inversely proportional to σ_i^2. The resulting minimum variance is

$$\frac{1}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2} + \cdots + \dfrac{1}{\sigma_p^2}}. \tag{A.1}$$

Result 2. If X and Y are positive random variables, then

$$E\left[\frac{1}{\dfrac{1}{X} + \dfrac{1}{Y}}\right] \le \frac{1}{\dfrac{1}{E[X]} + \dfrac{1}{E[Y]}}.$$

Result 3. If x_i, y_i, $i = 1, \ldots, p$, are positive numbers, then

$$\sum_{i=1}^{p}\left[\frac{1}{\dfrac{1}{x_i} + \dfrac{1}{y_i}}\right] \le \frac{1}{\dfrac{1}{\sum_{i=1}^{p} x_i} + \dfrac{1}{\sum_{i=1}^{p} y_i}}.$$

Since Result 3 is a special case of Result 2, we prove Result 2. Define

$$f(x, y) = \frac{1}{\dfrac{1}{x} + \dfrac{1}{y}}, \qquad x, y > 0.$$

The Hessian matrix of second partials of f is

$$H = \frac{2}{(x+y)^3}\begin{bmatrix} -y^2 & xy \\ xy & -x^2 \end{bmatrix}.$$

The quadratic form associated with H is negative semidefinite, implying that $f(x,y)$ is concave downward. Result 2 is then a consequence of Jensen's inequality.

Consider an S-allowable allocation $(r_1, r_2, \ldots, r_m)$ and a linear combination

$$L = \sum_{1\le i\le M}{}' \; W_i \hat{\mu}_i(r_i, r_i^*)$$

with $\sum'_{1\le i\le M} W_i = 1$. Recall that the prime indicates that the summation is over all i for which $r_i > 0$. The case of a unique middle rank order requires special treatment, so we assume for now that such a rank order does not occur in L, i.e., the summation is over $1 \le i < M$. Let j be the index i that minimizes $\sigma^2_{(i:m)}$ for $1 \le i < M$. We are going to show that

$$\text{(A.2)} \qquad \text{var}(L) \ge \text{var}\left(\hat{\mu}_j\left(\sum_i{}'(r_i + r_i^*)\right)\right).$$

In fact, we show that

$$\text{(A.3)} \qquad \text{var}(L) \ge \text{var}\left(\hat{\mu}_j\left(\sum_i{}' r_i, \sum_i{}' r_i^*\right)\right).$$

Equation (A.2) then follows since the right hand side of equation (A.3) is greater than or equal to the right hand side of equation (A.2). By Result 1, var(L) is greater than or equal to

$$\frac{1}{\sum_i' \frac{1}{\sigma_i^2}}$$

where

$$\begin{aligned}\sigma_i^2 &= \text{var}(\hat{\mu}_i(r_i, r_i^*)) \\ &= \frac{\sigma^2_{(i:m)}}{4}\left(\frac{1}{r_i} + \frac{1}{r_i^*}\right) \\ &\ge \frac{\sigma^2_{(j:m)}}{4}\left(\frac{1}{r_i} + \frac{1}{r_i^*}\right).\end{aligned}$$

Thus,

$$\text{var}(L) \ge \frac{1}{\sum_i' \left[\frac{\sigma^2_{(j:m)}}{4}\left(\frac{1}{r_i} + \frac{1}{r_i^*}\right)\right]^{-1}} = \frac{\sigma^2_{(j:m)}}{4}\left[\frac{1}{\sum_i' \left(\frac{1}{r_i} + \frac{1}{r_i^*}\right)^{-1}}\right].$$

On the other hand, the right hand side of equation (A.3) equals

$$\frac{\sigma^2_{(j:m)}}{4}\left[\frac{1}{\sum_i' r_i} + \frac{1}{\sum_i' r_i^*}\right].$$

Thus, (A.3) will be true provided

$$\frac{1}{\sum_i' \left[\frac{1}{r_i} + \frac{1}{r_i^*}\right]^{-1}} \geq \frac{1}{\sum_i' r_i} + \frac{1}{\sum_i' r_i^*}.$$

But this is a consequence of Result 3. Thus, we are finished when m is even.

Now, we consider the case where m is odd so the unique middle rank order M may occur in L. As before, let j be the rank order i that minimizes $\sigma^2_{(i:m)}$ for $1 \leq i < M$. Write L as

$$L = \sum_{1\leq i<M}' W_i \hat{\mu}_i(r_i, r_i^*) + W_M \hat{\mu}_M(r_M).$$

Applying the preceding result to the first term on the right, we can conclude that

$$\text{var}(L) \geq \text{var}(L_1),$$

where L_1 is given by

$$L_1 = (1 - W_M)\hat{\mu}_j \left(\sum_{1\leq i<M}' r_i, \sum_{1\leq i<M}' r_i^* \right) + W_M \hat{\mu}_M(r_M).$$

Thus, we are reduced to minimizing the variance for estimators of the form L_1. For notational simplicity, we drop the subscript M on W_M. Balancing the allocation assigned to $\hat{\mu}_j$ reduces the variance and we can restrict to estimators of the form

$$L_2 = (1 - W)\hat{\mu}_j(t) + W\hat{\mu}_M(n - t),$$

for $t = 0, 1, \ldots, n$. But note that $t = 1$ must be excluded since $\hat{\mu}_j(t)$ is not defined for $t = 1$. We can assume that W is the optimal weight as given by Result 1, so that

$$\text{var}(L_2) = \frac{1}{\frac{1}{\text{var}(\hat{\mu}_j(t))} + \frac{1}{\text{var}(\hat{\mu}_n(n-t))}}.$$

Taking reciprocals, we need to find the value of t that *maximizes*

$$\begin{aligned} G(t) &= \frac{1}{\text{var}(\hat{\mu}_j(t))} + \frac{1}{\text{var}(\hat{\mu}_M(n-t))} \\ &= K(t) \cdot \frac{t}{\sigma^2_{(j:m)}} + \frac{n-t}{\sigma^2_{(M:m)}}, \end{aligned}$$

where $K(t) = 1$ if t is even and $K(t) = (t^2 - 1)/t^2$ for odd t. To simplify the notation, write

$$\sigma^2 = \sigma^2_{(j:m)} \quad \text{and} \quad \tau^2 = \sigma^2_{(M:m)}.$$

We need to maximize

$$H(t) = \sigma^2 G(t) = \begin{cases} t - \dfrac{1}{t} + (n-t) \cdot \dfrac{\sigma^2}{\tau^2} & \text{if } t \text{ is odd} \\ t + (n-t)\dfrac{\sigma^2}{\tau^2} & \text{if } t \text{ is even,} \end{cases}$$

for $t = 0, 1, 2, \ldots, n$, but $t \neq 1$.

If $\tau^2 \leq \sigma^2$, then one checks directly that $H(t) \leq H(0) = n\sigma^2/\tau^2$ for all t. Thus the optimal allocation assigns all n measurements to the unique middle rank M.

Next, suppose that $\sigma^2 < \tau^2$. If t is odd, then

$$H(t+1) - H(t) = 1 + \frac{1}{t} - \frac{\sigma^2}{\tau^2} > 0,$$

so that $H(t+1) > H(t)$. Thus, the maximum does not occur for odd t, except possibly at the upper end $t = n$. When restricted to even t, it is obvious that $H(t)$ is monotone increasing. Thus the maximum occurs either at the largest even value of t or at n. If n is even, then the maximum is at n and the optimal allocation assigns all n measurements to $\hat{\mu}_j$.

Now suppose that n is odd. The maximum occurs at either $t = n - 1$ or $t = n$. But, since n is odd,

$$\begin{aligned} H(n) &= n - \frac{1}{n} \\ H(n-1) &= n - 1 + \frac{\sigma^2}{\tau^2} \\ H(n) - H(n-1) &= 1 - \frac{1}{n} - \frac{\sigma^2}{\tau^2}. \end{aligned}$$

Thus, the maximum occurs at n exactly when $1 - \frac{1}{n} - \frac{\sigma^2}{\tau^2} \geq 0$ or $\sigma^2 \leq \frac{n-1}{n}\tau^2$. For this case, the optimal allocation assigns all n measurements to $\hat{\mu}_j$.

In the final case,

$$\frac{n-1}{n}\tau^2 < \sigma^2 < \tau^2,$$

and the maximum occurs at $t = n-1$. The optimal allocation assigns $n-1$ measurements to $\hat{\mu}_j$ and one measurement to $\hat{\mu}_M$.

References

Cobby, J. M., Ridout, M. S., Bassett, P. J. and Large, R. V. (1985). An investigation into the use of ranked set sampling on grass and grass-clover swards, *Grass and Forage Science*, **40**, 257–263.

David, H. A. (1981). *Order Statistics*, Wiley, New York.

David, H. A. and Groeneveld, R. A. (1982). Measures of local variation in a distribution: Expected length of spacings and variances of order statistics, *Biometrika*, **69**, 227–232.

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

Halls, L. K. and Dell, T. R. (1966). Trials of ranked set sampling for forage yields, *Forest Science*, **12**, 22–26.

Kaur, A., Patil, G. P. and Taillie, C. (1997). Unequal allocation models for ranked set sampling with skew distributions, *Biometrics*, **53**, 123–130.

Kaur, A., Patil, G. P., Sinha, A. K. and Taillie, C. (1996). Ranked set sampling: An annotated bibliography, *Environmental and Ecological Statistics*, **2**, 25–54.

Kvam, P. and Samaniego, F. J. (1993). On the inadmissibility of empirical averages as estimators in ranked set sampling, *J. Statist. plann. Inference*, **36**, 39–55.

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

Patil, G. P., Boswell, M. T. and Ratnaparkhi, M. V. (1984). *Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work*, Vol. 2, International Co-operative Publishing House, Fairland, MD.

Patil, G. P., Sinha, A. K. and Taillie, C. (1994). Ranked set sampling, *Handbook of Statistics*, Vol. 12, (eds. G. P. Patil and C. R. Rao), 167–198, North-Holland, Amsterdam.

Shirahata, S. (1982). An extension of the ranked set sampling theory, *J. Statist. Plann. Inference*, **6**, 65–72.

Stokes, S. L. (1980*a*). Estimation of variance using judgment ordered ranked set samples, *Biometrics*, **36**, 35–42.

Stokes, S. L. (1980*b*). Inferences on the correlation coefficient in bivariate normal populations from ranked set samples, *J. Amer. Statist. Assoc.*, **75**, 989–995.

Stokes, S. L. and Sager, T. (1988). Characterization of a ranked set sample with application to estimating distribution functions, *J. Amer. Statist. Assoc.*, **83**, 374–381.

Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering, *Ann. Inst. Statist. Math.*, **20**, 1–31.

Yanagawa, T. and Shirahata, S. (1976). Ranked set sampling theory with selective probability matrix, *Austral. J. Statist.*, **18**, 45–52.