2

Balanced Ranked Set Sampling I: Nonparametric

In a balanced ranked set sampling, the number of measurements made on each ranked statistic is the same for all the ranks. A balanced ranked set sampling produces a data set as follows.

$$\begin{array}{cccc} X_{[1]1} & X_{[1]2} & \cdots & X_{[1]m} \\ X_{[2]1} & X_{[2]2} & \cdots & X_{[2]m} \\ \cdots & \cdots & \cdots & \cdots, \\ X_{[k]1} & X_{[k]2} & \cdots & X_{[k]m}. \end{array} \tag{2.1}$$

It should be noted that the $X_{[r]i}$'s in (2.1) are all mutually independent and, in addition, the $X_{[r]i}$'s in the same row are identically distributed. We denote by $f_{[r]}$ and $F_{[r]}$, respectively, the density function and the distribution function of the common distribution of the $X_{[r]i}$'s. The density function and the distribution function of the underlying distribution are denoted, respectively, by f and F. In this chapter, no assumption on the underlying distribution is made. We shall first discuss the ranking mechanisms of RSS and then discuss the properties of various statistical procedures using ranked set samples. In Section 1, the concept of consistent ranking mechanism is proposed and several consistent ranking mechanisms are discussed. In Section 2.2, the estimation of means using RSS sample is considered. The unbiasedness, asymptotic distribution and relative efficiency of the RSS estimates with respect to the corresponding SRS estimates are treated in detail. In Section 2.3, the results on the estimation of means are extended to the estimation of smooth functions of means. In Section 2.4, a special treatment is devoted to the estimation of variance. A minimum variance unbiased non-negative estimate of variance based on an RSS sample is proposed and studied. In Section 2.5, tests and confidence interval procedures for the population mean are discussed. In Section 2.6, the estimation of quantiles is tackled. The RSS sample quantiles are defined and their properties, similar to those of the SRS sample quantiles, such as strong consistency, Bahadur representation and asymptotic normality are established. In Section 2.7, the estimation of density function is treated. A kernel estimate based on an RSS sample is defined similarly as in the case of SRS. Its properties and relative efficiency with respect to SRS are investigated. In Section 2.8, the properties of M-estimates using RSS samples are discussed. Some technical details on the estimation of variance are given in Section 2.9. Readers who are not interested in technicalities might skip the final section.

2.1 Ranking mechanisms

In chapter 1, we introduced the general procedure of RSS. The procedure is a two-stage scheme. At the first stage, simple random samples are drawn and a certain ranking mechanism is employed to rank the units in each simple random sample. At the second stage, actual measurements of the variable of interest are made on the units selected based on the ranking information obtained at the first stage. The judgment ranking relating to the latent values of the variable of interest, as originally considered by McIntyre [96], provides one ranking mechanism. We mentioned that mechanisms other than this one can be used as well. In this section, we discuss the ranking mechanisms which are to be used in various practical situations.

Let us start with McIntyre's original ranking mechanism, i.e., ranking with respect to the latent values of the variable of interest. If the ranking is perfect, that is, the ranks of the units tally with the numerical orders of their latent values of the variable of interest, the measured values of the variable of interest are indeed order statistics. In this case, $f_{[r]} = f_{(r)}$, the density function of the rth order statistic of a simple random sample of size k from distribution F. We have

$$f_{(r)}(x) = \frac{k!}{(r-1)!(k-r)!} F^{r-1}(x)[1-F(x)]^{k-r} f(x).$$

It is then easy to verify that

$$f(x) = \frac{1}{k} \sum_{r=1}^{k} f_{(r)}(x),$$

for all x. This equality plays a very important role in RSS. It is this equality that gives rise to the merits of RSS. We are going to refer to equalities of this kind as fundamental equalities.

A ranking mechanism is said to be consistent if the following fundamental equality holds:

$$F(x) = \frac{1}{k} \sum_{r=1}^{k} F_{[r]}(x), \text{ for all } x. \tag{2.2}$$

Obviously, perfect ranking with respect to the latent values of X is consistent. We discuss other consistent ranking mechanisms in what follows.

(i) Imperfect ranking with respect to the variable of interest. When there are ranking errors, the density function of the ranked statistic with rank r is no longer $f_{(r)}$. However, we can express the corresponding cumulative distribution function $F_{[r]}$ in the form:

$$F_{[r]}(x) = \sum_{s=1}^{k} p_{sr} F_{(s)}(x),$$

where p_{sr} denotes the probability with which the sth (numerical) order statistic is judged as having rank r. If these error probabilities are the same within each cycle of a balanced RSS, we have $\sum_{s=1}^{k} p_{sr} = \sum_{r=1}^{k} p_{sr} = 1$. Hence,

$$\begin{aligned}\frac{1}{k}\sum_{r=1}^{k} F_{[r]}(x) &= \frac{1}{k}\sum_{r=1}^{k}\sum_{s=1}^{k} p_{sr} F_{(s)}(x) \\ &= \frac{1}{k}\sum_{s=1}^{k}(\sum_{r=1}^{k} p_{sr}) F_{(s)}(x) = F(x).\end{aligned}$$

(ii) Ranking with respect to a concomitant variable. There are cases in practical problems where the variable of interest, X, is hard to measure and difficult to rank as well but a concomitant variable, Y, can be easily measured. Then the concomitant variable can be used for the ranking of the sampling units. The RSS scheme is adapted in this situation as follows. At the first stage of RSS, the concomitant variable is measured on each unit in the simple random samples, and the units are ranked according to the numerical order of their values of the concomitant variable. Then the measured X values at the second stage are induced order statistics by the order of the Y values. Let $Y_{(r)}$ denote the rth order statistic of the Y's and $X_{[r]}$ denote its corresponding X. Let $f_{X|Y_{(r)}}(x|y)$ denote the conditional density function of X given $Y_{(r)} = y$ and $g_{(r)}(y)$ the marginal density function of $Y_{(r)}$. Then we have

$$f_{[r]}(x) = \int f_{X|Y_{(r)}}(x|y) g_{(r)}(y) dy.$$

It is easy to see that

$$\begin{aligned}f(x) &= \int \sum_{r=1}^{k} \frac{1}{k} f_{X|Y_{(r)}}(x|y) g_{(r)}(y) dy \\ &= \frac{1}{k}\sum_{r=1}^{k} f_{[r]}(x).\end{aligned}$$

(iii) Multivariate samples obtained by ranking one of the variables. Without loss of generality, let us consider the bivariate case. Suppose that inferences are to be made on the joint distribution of X and Y. The RSS scheme can

be similarly adapted in this case. The scheme goes the same as the standard RSS. The sampling units are ranked according to one of the variables, say Y. However, for each item to be quantified, both variables are measured. Let $f(x,y)$ denote the joint density function of X and Y and $f_{[r]}(x,y)$ the joint density function of $X_{[r]}$ and $Y_{[r]}$. Then

$$f_{[r]}(x,y) = f_{X|Y_{[r]}}(x|y)g_{[r]}(y)$$

and

$$f(x,y) = \frac{1}{k}\sum_{r=1}^{k} f_{[r]}(x,y).$$

(iv) Ranking mechanisms based on multiple concomitant variables. If there is more than one concomitant variable, any function of the concomitant variables can be used as a ranking criterion and the resultant ranking mechanism is consistent. Some of the ranking mechanisms based on functions of concomitant variables are discussed in detail in Chapter 6. We also develop a multi-layer RSS scheme using multiple concomitant variables, which is consistent, in Chapter 6.

2.2 Estimation of means using ranked set sample

Let $h(x)$ be any function of x. Denote by μ_h the expectation of $h(X)$, i.e., $\mu_h = Eh(X)$. We consider in this section the estimation of μ_h by using a ranked set sample. Examples of $h(x)$ include: (a) $h(x) = x^l, l = 1, 2, \cdots$, corresponding to the estimation of population moments, (b) $h(x) = I\{x \leq c\}$ where $I\{\cdot\}$ is the usual indicator function, corresponding to the estimation of distribution function, (c) $h(x) = \frac{1}{\lambda}K(\frac{t-x}{\lambda})$, where K is a given function and λ is a given constant, corresponding to the estimation of density function. We assume that the variance of $h(X)$ exists. Define the moment estimator of μ_h based on data (2.1) as follows.

$$\hat{\mu}_{h\cdot\text{RSS}} = \frac{1}{mk}\sum_{r=1}^{k}\sum_{i=1}^{m} h(X_{[r]i}).$$

We consider first the statistical properties of $\hat{\mu}_{h\cdot\text{RSS}}$ and then the relative efficiency of RSS with respect to SRS in the estimation of means.

First, we have the following result.

Theorem 2.1. *Suppose that the ranking mechanism in RSS is consistent. Then,*

(i) The estimator $\hat{\mu}_{h\cdot RSS}$ is unbiased, i.e., $E\hat{\mu}_{h\cdot RSS} = \mu_h$.

(ii) $Var(\hat{\mu}_{h\cdot RSS}) \leq \frac{\sigma_h^2}{mk}$, where σ_h^2 denotes the variance of $h(X)$, and the inequality is strict unless the ranking mechanism is purely random.

(iii)As $m \to \infty$,

$$\sqrt{mk}(\hat{\mu}_{h\cdot RSS} - \mu_h) \to N(0, \sigma^2_{h\cdot RSS}),$$

in distribution, where,

$$\sigma^2_{h\cdot RSS} = \frac{1}{k}\sum_{r=1}^{k}\sigma^2_{h[r]}.$$

Here $\sigma^2_{h[r]}$ *denotes the variance of* $h(X_{[r]i})$.

Proof: (i) It follows from the fundamental equality that

$$\begin{aligned}
E\hat{\mu}_{h\cdot \mathrm{RSS}} &= \frac{1}{mk}\sum_{r=1}^{k}\sum_{i=1}^{m} Eh(X_{[r]i}) = \frac{1}{k}\sum_{r=1}^{k} Eh(X_{[r]1}) \\
&= \frac{1}{k}\sum_{r=1}^{k}\int h(x)dF_{[r]}(x) = \int h(x)d\frac{1}{k}\sum_{r=1}^{k} F_{[r]}(x) \\
&= \int h(x)dF(x) = \mu_h.
\end{aligned}$$

(ii)

$$\begin{aligned}
\mathrm{Var}(\hat{\mu}_{h\cdot \mathrm{RSS}}) &= \frac{1}{(mk)^2}\sum_{r=1}^{k}\sum_{i=1}^{m}\mathrm{Var}(h(X_{[r]i})) = \frac{1}{mk^2}\sum_{r=1}^{k}\mathrm{Var}(h(X_{[r]})) \\
&= \frac{1}{mk}\left(\frac{1}{k}\sum_{r=1}^{k}(E[h(X_{[r]})]^2 - [Eh(X_{[r]})]^2)\right) \\
&= \frac{1}{mk}\left(m_{h2} - \frac{1}{k}\sum_{r=1}^{k}[Eh(X_{[r]})]^2\right),
\end{aligned}$$

where m_{h2} denotes the second moment of $h(X)$. It follows from the Caushy-Schwarz inequality that

$$\frac{1}{k}\sum_{r=1}^{k}[Eh(X_{[r]})]^2 \geq \left(\frac{1}{k}\sum_{r=1}^{k} Eh(X_{[r]})\right)^2 = \mu_h^2,$$

where the equality holds only when $Eh(X_{[1]}) = \cdots = Eh(X_{[r]})$ in which case the ranking mechanism is purely random.

(iii) By the fundamental equality, $\mu_h = \frac{1}{k}\sum_{r=1}^{k}\mu_{h[r]}$, where $\mu_{h[r]}$ is the expectation of $h(X_{[r]i})$. Then, we can write

$$\begin{aligned}
\sqrt{mk}(\hat{\mu}_{h\cdot \mathrm{RSS}} - \mu_h) &= \frac{1}{\sqrt{k}}\sum_{r=1}^{k}\sqrt{m}[\frac{1}{m}\sum_{i=1}^{m} h(X_{[r]i}) - \mu_{h[r]}] \\
&= \frac{1}{\sqrt{k}}\sum_{r=1}^{k} Z_{mr}, \text{ say.}
\end{aligned}$$

By the multivariate central limit theorem, $(Z_{m1}, \cdots, Z_{mk})$ converges to a multivariate normal distribution with mean vector zero and covariance matrix given by $Diag(\sigma^2_{h[1]}, \cdots, \sigma^2_{h[k]})$. Part (iii) then follows.

This proves the theorem.

We know that $\sigma_h^2/(mk)$ is the variance of the moment estimator of μ_h based on a simple random sample of size mk. Theorem 2.1 implies that the moment estimator of μ_h based on an RSS sample always has a smaller variance than its counterpart based on an SRS sample of the same size. In the context of RSS, we have tacitly assumed that the cost or effort for drawing sampling units from the population and then ranking them is negligible. When we compare the efficiency of a statistical procedure based on an RSS sample with that based on an SRS sample, we assume that the two samples have the same size. Let $\hat{\mu}_{h\cdot\text{SRS}}$ denote the sample mean of a simple random sample of size mk. We define the relative efficiency of RSS with respect to SRS in the estimation of μ_h as follows:

$$\text{RE}(\hat{\mu}_{h\cdot\text{RSS}}, \hat{\mu}_{h\cdot\text{SRS}}) = \frac{\text{Var}(\hat{\mu}_{h\cdot\text{SRS}})}{\text{Var}(\hat{\mu}_{h\cdot\text{RSS}})}. \tag{2.3}$$

Theorem 2.1 implies that $\text{RE}(\hat{\mu}_{h\cdot\text{RSS}}, \hat{\mu}_{h\cdot\text{SRS}}) \geq 1$. In order to investigate the relative efficiency in more detail, we derive the following:

$$\begin{aligned}
\sigma^2_{h\cdot\text{RSS}} &= \frac{1}{k}\sum_{r=1}^{k}\sigma^2_{h[r]} \\
&= \frac{1}{k}\sum_{r=1}^{k}(E[h(X_{[r]})]^2 - [Eh(X_{[r]})]^2) \\
&= \frac{1}{k}\sum_{r=1}^{k}E[h(X_{[r]})]^2 - \mu_h^2 + \mu_h^2 - \frac{1}{k}\sum_{r=1}^{k}[Eh(X_{[r]})]^2 \\
&= \sigma_h^2 - \frac{1}{k}\sum_{r=1}^{k}(\mu_{h[r]} - \mu_h)^2. \qquad (2.4)
\end{aligned}$$

Thus, we can express the relative efficiency as

$$\text{RE}(\hat{\mu}_{h\cdot\text{RSS}}, \hat{\mu}_{h\cdot\text{SRS}}) = \frac{\sigma_h^2}{\sigma^2_{h\cdot\text{RSS}}} = \left[1 - \frac{\frac{1}{k}\sum_{r=1}^{k}(\mu_{h[r]} - \mu_h)^2}{\sigma_h^2}\right]^{-1}.$$

It is clear from the above expression that, as long as there is at least one r such that $\mu_{h[r]} \neq \mu_h$, the relative efficiency is greater than 1. For a given underlying distribution and a given function h, the relative efficiency can be computed, at least, in principle.

In the remainder of this section, we discuss the relative efficiency in more detail for the special case that $h(x) = x$. Based on the computations on a number of underlying distributions, McIntyre [96] made the following conjecture: the relative efficiency of RSS with respect to SRS, in the estimation

Table 2.1. $\mu, \sigma^2, \gamma, \kappa$ and the relative efficiency of RSS with $k = 2, 3, 4$ for some distributions

Distribution	μ	σ^2	γ	κ	2	3	4
Uniform	0.500	0.083	0.000	1.80	1.500	2.000	2.500
Exponential	1.000	1.000	2.000	9.00	1.333	1.636	1.920
Gamma(0.5)	0.500	0.500	2.828	15.0	1.254	1.483	1.696
Gamma(1.5)	1.500	1.500	1.633	7.00	1.370	1.710	2.030
Gamma(2.0)	2.000	2.000	1.414	6.00	1.391	1.753	2.096
Gamma(3.0)	3.000	3.000	1.155	5.00	1.414	1.801	2.169
Gamma(4.0)	4.000	4.000	1.000	4.50	1.427	1.827	2.210
Gamma(5.0)	5.000	5.000	0.894	4.20	1.434	1.843	2.236
Normal	0.000	1.000	0.000	3.00	1.467	1.914	2.347
Beta (4,4)	0.500	0.028	0.000	2,45	1.484	1.958	2.425
Beta(7,4)	0.636	0.019	-0.302	2.70	1.475	1.936	2.389
Beta(13,4)	0.765	0.010	-0.557	3.14	1.460	1.903	2.333
Weibull(0.5)	2.000	20.00	6.619	87.7	1.127	1.236	1.334
Weibull(1.5)	0.903	0.376	1.072	4.39	1.422	1.822	2.205
Weibull(2.0)	0.886	0.215	0.631	3.24	1.458	1.897	2.325
Weibull(3.0)	0.893	0.105	0.168	2.73	1.476	1.936	2.387
Weibull(4.0)	0.906	0.065	-0.087	2.75	1.474	1.932	2.380
Weibull(5.0)	0.918	0.044	-0.254	2.88	1.469	1.921	2.361
Weibull(6.0)	0.918	0.032	-0.373	3.04	1.464	1.909	2.341
Weibull(7.0)	0.935	0.025	-0.463	3.19	1.459	1.898	2.324
Weibull(8.0)	0.942	0.019	-0.534	3.33	1.456	1.890	2.309
$\chi^2(1)$	0.789	0.363	0.995	3.87	1.430	1.841	2.239
Triangular	0.500	0.042	0.000	2.40	1.485	1.961	2.430
Extreme value	0.577	1.645	1.300	5.40	1.413	1.793	2.153

of population mean, is between 1 and $(k+1)/2$ where k is the set size; For symmetric underlying distributions, the relative efficiency is not much less than $(k+1)/2$, however, as the underlying distribution becomes asymmetric, the relative efficiency drops down. Takahasi and Wakimoto [167] showed that, when ranking is perfect, $\frac{1}{k}\sum_{r=1}^{k}\sigma_{h[r]}^2$, as a function of k, decreases as k increases, which implies that the relative efficiency increases as k increases. A practical implication of this result is that, in the case of judgment ranking relating to the latent values of the variable of interest, when ranking accuracy can still be assured or, in other cases, when the cost of drawing sampling units and ranking by the given mechanism can still be kept at a negligible level, the set size k should be taken as large as possible. Dell and Clutter [50] computed the relative efficiency for a number of underlying distributions. They noticed that the relative efficiency is affected by the underlying distribution, especially by the skewness and kurtosis. Table 2.1 below is partially reproduced from Table 1 of Dell and Clutter (1972). The notations μ, σ^2, γ and κ in the table stand, respectively, for the mean, variance, skewness and kurtosis.

2.3 Estimation of smooth-function-of-means using ranked set sample

In this section we deal with the properties of RSS for a particular model, the smooth-function-of-means model, which refers to the situation where we are interested in the inference on a smooth function of population moments. Typical examples of smooth-function-of-means are (i) the variance, (ii) the coefficient of variation, and (iii) the correlation coefficient, etc. Let $m_1, \ldots, m_p$ denote p moments of F and g a p-variate smooth function with first derivatives. We consider the method-of-moment estimation of $g(m_1, \ldots, m_p)$.

The following notation will be used in this section. Let $Z_l, l = 1, \ldots, p$, be functions of $X(\sim F)$ such that $E[Z_l] = m_l$. Let $n = km$. A simple random sample of size n is represented by $\{(Z_{1j}, ..., Z_{pj}) : j = 1, \ldots, n\}$. A general RSS sample of size n is represented by $\{(Z_{1(r)i}, \ldots, Z_{p(r)i}) : r = 1, \ldots, k; i = 1, \ldots, m\}$. The simple random and ranked set sample moments are denoted, respectively, by

$$\bar{Z}_l = \frac{1}{n} \sum_{j=1}^{n} Z_{lj}, l = 1, \ldots, p$$

and

$$\tilde{Z}_l = \frac{1}{km} \sum_{r=1}^{k} \sum_{i=1}^{m} Z_{l(r)i}, l = 1, \ldots, p.$$

Let $\bar{\boldsymbol{Z}}_{\text{SRS}} = (\bar{Z}_1, \ldots, \bar{Z}_p)^T$ and $\bar{\boldsymbol{Z}}_{\text{RSS}} = (\tilde{Z}_1, \ldots, \tilde{Z}_p)^T$. Denote by Σ_{SRS} and Σ_{RSS} the variance-covariance matrices of $\sqrt{n}\bar{\boldsymbol{Z}}_{\text{SRS}}$ and $\sqrt{n}\bar{\boldsymbol{Z}}_{\text{RSS}}$, respectively. Let $\boldsymbol{\partial g}$ denote the vector of the first partial derivatives of g evaluated at $(m_1, \ldots, m_p)$. Define

$$\begin{aligned} \eta &= g(m_1, \ldots, m_p), \\ \hat{\eta}_{\text{SRS}} &= g(\bar{Z}_1, \ldots, \bar{Z}_p), \\ \hat{\eta}_{\text{RSS}} &= g(\tilde{Z}_1, \ldots, \tilde{Z}_p). \end{aligned}$$

We first state the asymptotic normality of $\hat{\eta}_{\text{SRS}}$ and $\hat{\eta}_{\text{RSS}}$, and then consider the asymptotic relative efficiency (ARE) of $\hat{\eta}_{\text{RSS}}$ with respect to $\hat{\eta}_{\text{SRS}}$.

Theorem 2.2. *As $m \to \infty$ (hence $n \to \infty$), we have*

$$\sqrt{n}(\hat{\eta}_{SRS} - \eta) \to N(0, \boldsymbol{\partial g}^{\boldsymbol{T}} \Sigma_{SRS} \boldsymbol{\partial g})$$

in distribution and

$$\sqrt{n}(\hat{\eta}_{RSS} - \eta) \to N(0, \boldsymbol{\partial g}^{\boldsymbol{T}} \Sigma_{RSS} \boldsymbol{\partial g})$$

in distribution.

The above result follows from the multivariate central limit theorem. The proof is omitted. The ARE of $\hat{\eta}_{\text{RSS}}$ with respect to $\hat{\eta}_{\text{SRS}}$ is defined as

$$\text{ARE}(\hat{\eta}_{\text{RSS}}, \hat{\eta}_{\text{SRS}}) = \frac{\boldsymbol{\partial g^T} \Sigma_{\text{SRS}} \boldsymbol{\partial g}}{\boldsymbol{\partial g^T} \Sigma_{\text{RSS}} \boldsymbol{\partial g}}.$$

The next theorem implies that the ARE of $\hat{\eta}_{\text{RSS}}$ with respect to $\hat{\eta}_{\text{SRS}}$ is always greater than 1.

Theorem 2.3. *Suppose that the ranking mechanism in RSS is consistent. Then we have that*

$$\Sigma_{SRS} \geq \Sigma_{RSS},$$

where $\Sigma_{SRS} \geq \Sigma_{RSS}$ *means that* $\Sigma_{SRS} - \Sigma_{RSS}$ *is non-negative definite.*

Proof: It suffices to prove that, for any vector of constants $\boldsymbol{a}$,

$$\boldsymbol{a^T} \Sigma_{\text{SRS}} \boldsymbol{a} - \boldsymbol{a^T} \Sigma_{\text{RSS}} \boldsymbol{a} \geq 0. \tag{2.5}$$

Define

$$Y = \boldsymbol{a^T Z} = \sum_{j=1}^{p} a_j Z_j \quad \text{and} \quad \mu_Y = \boldsymbol{a^T} \mathbf{m} = \sum_{j=1}^{p} a_j m_j.$$

Then we have

$$\hat{\mu}_{Y \cdot \text{SRS}} = \boldsymbol{a^T} \bar{Z}_{\text{SRS}} \quad \text{and} \quad \hat{\mu}_{Y \cdot \text{RSS}} = \boldsymbol{a^T} \bar{Z}_{\text{RSS}}.$$

It follows from Theorem 2.1 that

$$\text{Var}(\hat{\mu}_{Y \cdot \text{RSS}}) \leq \text{Var}(\hat{\mu}_{Y \cdot \text{SRS}}),$$

i.e.,

$$\boldsymbol{a^T} \Sigma_{\text{SRS}} \boldsymbol{a} \geq \boldsymbol{a^T} \Sigma_{\text{RSS}} \boldsymbol{a}.$$

The theorem is proved.

In fact, it can be proved that, as long as there are at least two ranks, say r and s, such that $F_{[r]} \neq F_{[s]}$, then $\Sigma_{\text{SRS}} > \Sigma_{\text{RSS}}$.

It should be noted that, unlike in the estimation of means, the estimator of a smooth-function-of-means is no longer necessarily unbiased. It is only asymptotically unbiased. In this case, the relative efficiency of RSS with respect to SRS should be defined as the ratio of the mean square errors of the two estimators. The ARE, which is the limit of the relative efficiency as the samples size goes to infinity, does not take into account the bias for finite sample sizes. In general, the ARE can not be achieved when sample size is small. We consider this issue in more detail for the special case of the estimation of population variance σ^2 in the next section.

2.4 Estimation of variance using an RSS sample

2.4.1 Naive moment estimates

The natural estimates of σ^2 using an SRS sample and an RSS sample are given, respectively, by

$$s^2_{\text{SRS}} = \frac{1}{mk-1}\sum_{r=1}^{k}\sum_{i=1}^{m}(X_{ri} - \bar{X}_{\text{SRS}})^2,$$

where $\bar{X}_{\text{SRS}} = \frac{1}{mk}\sum_{r=1}^{k}\sum_{i=1}^{m} X_{ri}$, and

$$s^2_{\text{RSS}} = \frac{1}{mk-1}\sum_{r=1}^{k}\sum_{i=1}^{m}(X_{[r]i} - \bar{X}_{\text{RSS}})^2,$$

where $\bar{X}_{\text{RSS}} = \frac{1}{mk}\sum_{r=1}^{k}\sum_{i=1}^{m} X_{[r]i}$.

The properties of s^2_{RSS} were studied by Stokes [159]. Unlike the SRS version s^2_{SRS}, the RSS version s^2_{RSS} is biased. It can be derived, see Stokes [159], that

$$Es^2_{\text{RSS}} = \sigma^2 + \frac{1}{k(mk-1)}\sum_{r=1}^{k}(\mu_{[r]} - \mu)^2.$$

An appropriate measure of the relative efficiency of s^2_{RSS} with respect to s^2_{SRS} is then given by

$$\begin{aligned}\text{RE}(s^2_{\text{RSS}}, s^2_{\text{SRS}}) &= \frac{\text{Var}(s^2_{\text{SRS}})}{\text{MSE}(s^2_{\text{RSS}})} \\ &= \frac{\text{Var}(s^2_{\text{SRS}})}{\text{Var}(s^2_{\text{RSS}}) + \left[\frac{1}{k(mk-1)}\sum_{r=1}^{k}(\mu_{[r]} - \mu)^2\right]^2}.\end{aligned}$$

It can be easily seen that

$$\text{RE}(s^2_{\text{RSS}}, s^2_{\text{SRS}}) < \text{ARE}(s^2_{\text{RSS}}, s^2_{\text{SRS}}).$$

Since

$$\frac{1}{k}\sum_{r=1}^{k}(\mu_{[r]} - \mu)^2 < \sigma^2,$$

it is clear that $\frac{1}{k(mk-1)}\sum_{r=1}^{k}(\mu_{[r]} - \mu)^2$ will decrease as either k or m increases. That is, the RE will converge increasingly to the ARE as either k or m increases. Stokes [159] computed both the RE when $m = 1$ and the ARE for a few underlying distributions. Table 2.2 below is reproduced from Table 1 of [159]. In the table, the U-shaped distribution refers to the distribution with

Table 2.2. Relative efficiency, $\mathrm{Var}(s^2_{\mathrm{SRS}})/\mathrm{MSE}(s^2_{\mathrm{RSS}})$, for $m = 1$ and $m \to \infty$.

Distribution	k	$m = 1$	$m \to \infty$
(i) U-shaped	2	0.85	1.00
	3	1.22	1.10
	4	1.28	1.21
(ii) Uniform	2	0.72	1.00
	3	0.92	1.11
	4	1.09	1.25
	5	1.20	1.40
(iii) Normal	2	0.68	1.00
	3	0.81	1.08
	4	0.93	1.18
	5	1.03	1.27
(iv) Gamma	2	0.71	1.02
	3	0.81	1.08
	4	0.91	1.16
	5	1.00	1.23
	6	1.09	1.35
(v) Exponential	2	0.78	1.03
	3	0.84	1.08
	4	0.91	1.12
	5	0.97	1.17
	6	1.02	1.22
(vi) Lognormal	2	0.93	1.00
	3	0.95	1.01
	4	0.96	1.01
	5	0.97	1.02
	6	0.98	1.02
	7	0.99	1.03
	8	1.00	1.03
	9	1.01	1.04

density function $f(x) = (3/2)x^2 I\{-1 \le x \le 1\}$, and the Gamma distribution has density function $f(x) = x^4 \exp(-x)/\Gamma(5) I\{x \ge 0\}$.

It should be remarked that, in the estimation of variance, RSS is not necessarily more efficient than SRS when sample size is small, and the relative efficiency is much smaller than in the estimation of population mean even when RSS is beneficial. Therefore, if the estimation of variance is the primary purpose, it is not worthwhile to apply RSS. RSS is most useful when both the population mean and variance are to be estimated.

It is indeed a natural question whether or not better estimates of σ^2 based on an RSS sample can be found. We take up this question in the next subsection.

2.4.2 Minimum variance unbiased non-negative estimate

We demonstrate in this subsection that it is possible to construct a class of nonnegative unbiased estimates of σ^2 based on a balanced RSS sample, whatever the nature of the underlying distribution. Towards this end, we need the following basic identity which follows directly from (2.4):

$$\sigma^2 = \frac{1}{k}[\sum_{r=1}^{k} \sigma_{[r]}^2 + \sum_{r=1}^{k} \mu_{[r]}^2] - \mu^2. \tag{2.6}$$

Recall that $\bar{X}_{i:\text{RSS}} = (1/k)\sum_{r=1}^{k} X_{[r]i}$ provides an unbiased estimate of the mean μ based on the data of the ith cycle of an RSS . Let

$$W_i = \sum_{r=1}^{k}(X_{[r]i} - \bar{X}_{i:\text{RSS}})^2, \;\; i = 1, \cdots, m. \tag{2.7}$$

From the basic identity (2.6), it is clear that an unbiased estimate of σ^2 can be obtained by plugging in unbiased estimates of $\sigma_{[r]}^2 + \mu_{[r]}^2$ and μ^2. Since $\sum_{i=1}^{m} X_{[r]i}^2/m$ is an unbiased estimate of the former term and $\bar{X}_{i:\text{RSS}}\bar{X}_{j:\text{RSS}}$ for $i \neq j$ is an unbiased estimate of μ^2, it follows easily that an unbiased estimate of σ^2 is given by

$$\hat{\sigma}^2 = \frac{\sum_{r=1}^{k}\sum_{i=1}^{m}[X_{[r]i}]^2}{mk} - \frac{\sum_{i\neq j}\bar{X}_{i:\text{RSS}}\bar{X}_{j:\text{RSS}}}{m(m-1)}. \tag{2.8}$$

The above estimate can be readily simplified as

$$\hat{\sigma}^2 = \frac{W}{mk} + \frac{B}{(m-1)k} \tag{2.9}$$

where B and W represent, respectively, the between- and within-cycle sum of squares of the entire balanced data, defined as

$$B = k\sum_{i=1}^{m}(\bar{X}_{i:\text{RSS}} - \bar{X}_{\text{RSS}})^2, \;\; W = \sum_{i=1}^{m} W_i. \tag{2.10}$$

It is obvious that $\hat{\sigma}^2$ is nonnegative.

Let

$$\boldsymbol{X}_{[r]} = (X_{[r]1}, \cdots, X_{[r]m})', \; r = 1, \cdots, k,$$
$$\boldsymbol{X} = (\boldsymbol{X}'_{[1]}, \cdots, \boldsymbol{X}'_{[k]})'.$$

Denote by $\mathbf{1}_m$ an m-dimensional vector of elements 1's and I_m an identity matrix of order m. It can be easily verified that

$$\hat{\sigma}^2 = \sum_{r=1}^{k}\sum_{s=1}^{k}[\tilde{a}_{rs}\boldsymbol{X}_{[r]}^{'}(I_m - \frac{\mathbf{1}_m\mathbf{1}_m^{'}}{m})\boldsymbol{X}_{[s]} + \tilde{d}_{rs}\boldsymbol{X}_{[r]}^{'}\frac{\mathbf{1}_m\mathbf{1}_m^{'}}{m}\boldsymbol{X}_{[s]}],$$

where

$$\tilde{a}_{rs} = \begin{cases} \frac{1}{m(m-1)k^2}, \ r \neq s, \\ \frac{(m-1)k+1}{m(m-1)k^2}, \ r = s, \end{cases}$$

$$\tilde{d}_{rs} = \begin{cases} 1-m, & r \neq s, \\ 1-m+\frac{1}{mk}, & r = s. \end{cases}$$

Furthermore, let

$$\tilde{A} = (\tilde{a}_{rs})_{k\times k}, \quad \tilde{D} = (\tilde{d}_{rs})_{k\times k}.$$

We can express $\hat{\sigma}^2$ as follows:

$$\hat{\sigma}^2 = \boldsymbol{X}^{'}[\tilde{A}\otimes(I_m - \frac{\mathbf{1}_m\mathbf{1}_m^{'}}{m}) + \tilde{D}\otimes\frac{\mathbf{1}_m\mathbf{1}_m^{'}}{m}]\boldsymbol{X},$$

where $\otimes$ denotes the operation of Kronecker product.

The form of $\hat{\sigma}^2$ above motivates us to consider a class of quadratic estimators of σ^2. Let

$$Q = A\otimes(I_m - \frac{\mathbf{1}_m\mathbf{1}_m^{'}}{m}) + D\otimes\frac{\mathbf{1}_m\mathbf{1}_m^{'}}{m},$$

where $A = (a_{rs})_{k\times k}$ and $D = (d_{rs})_{k\times k}$ are arbitrary $k\times k$ symmetric matrices. We consider the class of quadratic estimators $\boldsymbol{X}^{'}Q\boldsymbol{X}$ with Q given by the above form. The following theorem provides characterizations of unbiased estimators of σ^2 and of the minimum variance unbiased estimator of σ^2 among estimators in the above class.

Theorem 2.4. *(a) $\hat{\sigma}^2 = \boldsymbol{X}^{'}Q\boldsymbol{X}$ in the above class is unbiased for σ^2 if and only if*

$$a_{rr} = \frac{1}{k(m-1)}[1 - \frac{k-1}{mk}] = \frac{1}{mk}[1 + \frac{1}{k(m-1)}]$$

for $r = 1, 2, \ldots, k$, and $D = (1/mk)[I_k - \mathbf{1}_k\mathbf{1}_k^{'}/k]$.

(b) Among all unbiased estimators of σ^2 in the above class, $\hat{\sigma}^2$ has the minimum variance if and only if $a_{rs} = 0$, for all $r \neq s$.

The proof of the theorem is given in the appendix at the end of this chapter.

Corollary 1 *The minimum variance unbiased nonnegative estimate of σ^2 in the class considered can be simplified as*

$$\hat{\sigma}^2_{UMVUE} = \frac{k(m-1)+1}{k^2m(m-1)}W^* + \frac{1}{mk}B^*,$$

where B^* and W^* are the between- and within-rank sum of squares of the RSS data defined by, respectively,

$$B^* = m\sum_{r=1}^{k}(\bar{X}_{[r]} - \bar{X}_{RSS})^2,$$

$$W^* = \sum_{r=1}^{k} W_r^*,$$

where $\bar{X}_{[r]} = (1/m)\sum_{i=1}^{m} X_{[r]i}$ and $W_r^* = \sum_{i=1}^{m}(X_{[r]i} - \bar{X}_{[r]})^2$, $r = 1, \cdots, k$. Moreover, we have

$$\begin{aligned} Var(\hat{\sigma}^2_{UMVUE}) = &\frac{1}{mk^2}\sum_{r=1}^{k}\kappa^*_{[r]} + 2\frac{(k(m-1)+1)^2}{k^4m^2(m-1)}\sum_{r=1}^{k}\sigma^4_{[r]} \\ &+\frac{2}{k^4m^2}[\sum_{r=1}^{k}\sigma^2_{[r]}]^2 + \frac{2k(k-2)}{k^4m^2}\sum_{r=1}^{k}\sigma^4_{[r]} \\ &+\frac{4}{k^2m}\sum_{r=1}^{k}\tau^2_{[r]}\sigma^2_{[r]} + \frac{4}{k^2m}\sum_{r=1}^{k}\tau_{[r]}\gamma^*_{[r]}, \end{aligned}$$

where $\mu_{[r]}, \sigma_{[r]}, \gamma^*_{[r]}$ and $\kappa^*_{[r]}$ are, respectively the mean, standard deviation, third and fourth cumulants of $X_{[r]i}$, and $\tau_{[r]} = \mu_{[r]} - \mu$.

Remarks:

(i) It is easy to verify that two other unbiased nonnegative estimates of σ^2 with special choices of off-diagonal elements of A are given by

$$\hat{\sigma}^2_1 = \frac{k(m-1)+1}{km(m-1)}B + \frac{1}{mk}B^* \text{ and } \hat{\sigma}^2_2 = \frac{1}{mk}W + \frac{1}{k^2(m-1)}W^*.$$

(ii) The variance of Stokes's (1980) estimate of σ^2 can be easily obtained by choosing $A = \frac{1}{mk-1}\boldsymbol{I}_m$ and $D = \frac{1}{mk-1}[I_k - \mathbf{1}_k\mathbf{1}_k^{'}/k]$.

(iii) The asymptotic relative efficiency (ARE) of $\hat{\sigma}^2_{\text{UMVUE}}$ compared to Stokes's estimate, defined as the limit of $\text{Var}(\hat{\sigma}^2_{Stokes})/\text{Var}(\hat{\sigma}^2_{\text{UMVUE}})$ as $m \to \infty$, turns out to be unity. However, for small m, the relative efficiency computed for several distributions shows the superiority of the UMVUE over Stokes's estimate. For details, see Perron and Sinha [132].

2.5 Tests and confidence intervals for population mean

In this section, we present asymptotic testing and confidence interval procedures for the population mean based on a balanced RSS under a nonparametric set up.

2.5.1 Asymptotic pivotal method

Assume that the population of concern has finite mean μ and variance σ^2. We discuss procedures for constructing confidence intervals and testing hypotheses for μ based on a pivot. A pivot for μ is a function of μ and the data, whose distribution or asymptotic distribution does not depend on any unknown parameters, and usually can be obtained by standardizing an unbiased estimate of μ. Thus if $\hat{\mu}_{\text{RSS}}$ is an unbiased estimate of μ based on a balanced ranked set sample and $\hat{\sigma}_{\hat{\mu}_{\text{RSS}}}$ is a consistent estimate of the standard deviation of $\hat{\mu}_{\text{RSS}}$, a pivot for μ can be formed as

$$Z_1 = \frac{\hat{\mu}_{\text{RSS}} - \mu}{\hat{\sigma}_{\hat{\mu}_{\text{RSS}}}}.$$

Then an equal tailed $100(1-\alpha)\%$ confidence interval of μ can be constructed as

$$[\hat{\mu}_{\text{RSS}} - z_{1-\alpha/2}\hat{\sigma}_{\hat{\mu}_{\text{RSS}}}, \hat{\mu}_{\text{RSS}} - z_{\alpha/2}\hat{\sigma}_{\hat{\mu}_{\text{RSS}}}]$$

where $z_{\alpha/2}$ denotes the $(\alpha/2)$th quantile of the asymptotic distribution of Z_1. A hypothesis for μ, either one-sided or two-sided, can be tested based on the pivot Z_1 in a straightforward manner. On the other hand, based on a simple random sample of size N, a pivot Z_2 is given by

$$Z_2 = \frac{\bar{X} - \mu}{s/\sqrt{N}}$$

where $\bar{X}$ is the sample mean and s^2 is the usual unbiased estimate of σ^2. In what follows, we derive the counterparts of Z_2 by choosing Z_1 appropriately under a balanced ranked set sampling scheme.

Obviously, in the case of a *balanced* RSS, the ranked set sample mean, $\bar{X}_{\text{RSS}} = 1/(mk)\sum_{r=1}^{k}\sum_{i=1}^{m} X_{[r]i}$, is an unbiased estimate of μ. Therefore, a pivot for μ can be formed as

$$Z = \frac{\bar{X}_{\text{RSS}} - \mu}{\hat{\sigma}_{\bar{X}_{\text{RSS}}}}$$

where $\hat{\sigma}_{\bar{X}_{\text{RSS}}}$ is a consistent estimate of $\sigma_{\bar{X}_{\text{RSS}}}$, the standard deviation of $\bar{X}_{\text{RSS}}$. It follows from the central limit theorem that Z follows asymptotically a standard normal distribution. However, it turns out that there are several consistent estimates of $\sigma_{\bar{X}_{\text{RSS}}}$, each of which gives rise to a pivotal statistic. The question then obviously arises as to which of the estimates of $\sigma_{\bar{X}_{\text{RSS}}}$ should be used in the pivot. We shall now discuss different consistent estimates of $\sigma_{\bar{X}_{\text{RSS}}}$ and make a choice in the next section.

2.5.2 Choice of consistent estimates of $\boldsymbol{\sigma_{\bar{X}_{\text{RSS}}}}$

In this section we discuss several consistent estimates of $\sigma_{\bar{X}_{\text{RSS}}}$ and make a choice for the one to be used in the pivot considered in the previous section.

First, we give different expressions for $\bar{X}_{\text{RSS}}$ that motivate various estimates of $\sigma_{\bar{X}_{\text{RSS}}}$. We can write $\bar{X}_{\text{RSS}}$ as

$$\bar{X}_{\text{RSS}} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{k}\sum_{r=1}^{k}X_{[r]i} = \frac{1}{m}\sum_{i=1}^{m}\bar{X}_{i\text{RSS}}$$

where $\bar{X}_{i\text{RSS}}$, $i = 1, \cdots, m$, are i.i.d. with mean μ and variance, say, τ^2. Then we have

$$\text{Var}(\bar{X}_{\text{RSS}}) = \frac{\tau^2}{m}. \tag{2.11}$$

Obviously, an unbiased estimate of τ^2 is given by

$$\hat{\tau}_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{X}_{i\text{RSS}} - \bar{X}_{\text{RSS}})^2.$$

On the other hand, we can express $\bar{X}_{\text{RSS}}$ as

$$\bar{X}_{\text{RSS}} = \frac{1}{k}\sum_{r=1}^{k}\frac{1}{m}\sum_{i=1}^{m}X_{[r]i} = \frac{1}{k}\sum_{r=1}^{k}\bar{X}_{[r]}$$

where $\bar{X}_{[r]}$, $r = 1, \cdots, k$, are independent with means and variances given by, respectively, $E(\bar{X}_{[r]}) = \mu_{[r]}$ and $\text{Var}(\bar{X}_{[r]}) = \sigma_{[r]}^2/m$. Thus we have another expression of $\text{Var}(\bar{X}_{\text{RSS}})$ as follows.

$$\text{Var}(\bar{X}_{\text{RSS}}) = \frac{1}{mk^2}\sum_{r=1}^{k}\sigma_{[r]}^2. \tag{2.12}$$

Comparing (2.11) and (2.12), we have

$$\tau^2 = \frac{1}{k^2}\sum_{r=1}^{k}\sigma_{[r]}^2. \tag{2.13}$$

Hence an unbiased estimate of τ^2 is obtained through the expression (2.13) by forming unbiased estimates of $\sigma_{[r]}^2$'s. The usual unbiased estimate of $\sigma_{[r]}^2$ is given by

$$\hat{\sigma}_{[r]}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_{[r]i} - \bar{X}_{[r]})^2,$$

since for fixed r, $X_{[r]i}, i = 1, \cdots, m$, are i.i.d. with variance $\sigma_{[r]}^2$. Hence, an unbiased estimate of τ^2 is given by

$$\hat{\tau}_2^2 = \frac{1}{k^2}\sum_{r=1}^{k}\hat{\sigma}_{[r]}^2.$$

Since $\hat{\sigma}^2_{[r]}$ is consistent for $\sigma^2_{[r]}$, obviously $\hat{\tau}^2_2$ is consistent for τ^2.

Finally, we can express τ^2 as

$$\tau^2 = \frac{\sigma^2}{k} - \frac{1}{k^2}\sum_{r=1}^{k}(\mu_{[r]} - \mu)^2. \tag{2.14}$$

Thus, if we plug in consistent estimates of $\sigma^2, \mu_{[r]}$ and μ into (2.14), we obtain a third consistent estimate of τ^2 as follows.

$$\hat{\tau}^2_3 = \frac{\hat{\sigma}^2}{k} - \frac{1}{k^2}\sum_{r=1}^{k}(\hat{\mu}_{[r]} - \hat{\mu})^2. \tag{2.15}$$

We can take

$$\begin{aligned}
\hat{\mu} &= \bar{X}_{\text{RSS}}, \\
\hat{\mu}_{[r]} &= \frac{1}{m}\sum_{i=1}^{m} X_{[r]i} = \bar{X}_{[r]}, \\
\hat{\sigma}^2 &= \frac{1}{mk}\sum_{r=1}^{k}\sum_{i=1}^{m}(X_{[r]i} - \bar{X}_{\text{RSS}})^2.
\end{aligned}$$

It can be verified that (see Stokes, 1980)

$$\begin{aligned}
E\hat{\sigma}^2 &= (1 - \frac{1}{mk})\sigma^2 + \frac{1}{mk^2}\sum_{r=1}^{k}(\mu_{[r]} - \mu)^2 \\
&= \sigma^2 + O(\frac{1}{m})
\end{aligned}$$

and that

$$\begin{aligned}
E[\frac{1}{k^2}\sum_{r=1}^{k}(\hat{\mu}_{[r]} - \hat{\mu})^2] &= \frac{1}{k^2}\sum_{r=1}^{k}(\mu_{[r]} - \mu)^2 + \frac{k-1}{mk}\tau^2 \\
&= \frac{1}{k^2}\sum_{r=1}^{k}(\mu_{[r]} - \mu)^2 + O(\frac{1}{m}).
\end{aligned}$$

Some other consistent estimates of τ^2 can also be obtained by plugging in bias-corrected estimators of σ^2 and $\sum_{r=1}^{k}(\mu_{[r]} - \mu)^2$ into (2.14).

In the remainder of this section, we compare the three estimates of τ^2 defined above. After some manipulation, we can obtain

$$\hat{\tau}^2_3 = \frac{m-1}{m}\hat{\tau}^2_2.$$

Therefore, $\hat{\tau}^2_3$ and $\hat{\tau}^2_2$ are asymptotically equivalent. In fact, any estimate of τ^2 obtained by bias-correcting $\hat{\sigma}^2$ and $\sum_{r=1}^{k}(\hat{\mu}_{[r]} - \hat{\mu})^2$ is asymptotically equivalent to $\hat{\tau}^2_2$. Such a bias-corrected estimator takes the form

$$a(m,k)\hat{\tau}_2^2 + b(m,k)\sum_{r=1}^{k}(\hat{\mu}_{[r]} - \hat{\mu})^2.$$

In order to make it consistent, we must have, as $m \to \infty$,

$$a(m,k) \to 1, \quad b(m,k) \to 0.$$

Hence, the estimate is essentially of the form

$$\hat{\tau}_2^2 + o(1).$$

The term $o(1)$ is of order $O(1/m^2)$ or higher with bias-correction.

Therefore, we only need to consider $\hat{\tau}_1^2$ and $\hat{\tau}_2^2$. Some straightforward algebra yields that

$$\hat{\tau}_1^2 = \hat{\tau}_2^2 + \frac{m}{(m-1)k^2}\sum_{r\neq s}(\bar{X}_{[rs]} - \bar{X}_{[r]}\bar{X}_{[s]}),$$

where

$$\bar{X}_{[rs]} = \frac{1}{m}\sum_{i=1}^{m} X_{[r]i}X_{[s]i}.$$

Note that

$$E\sum_{r\neq s}(\bar{X}_{[rs]} - \bar{X}_{[r]}\bar{X}_{[s]}) = 0.$$

We have the following result.

Lemma 2.5. *Among all unbiased estimates of τ^2 of the form*

$$\hat{\tau}_2^2 + \lambda\sum_{r\neq s}(\bar{X}_{[rs]} - \bar{X}_{[r]}\bar{X}_{[s]}),$$

the estimate $\hat{\tau}_2^2$ has the smallest variance.

Proof: In order to prove the lemma, we only need to verify that, for any l, r, s such that $r \neq s$,

$$\text{Cov}(\bar{X}_{[ll]} - \bar{X}_{[l]}^2, \bar{X}_{[rs]} - \bar{X}_{[r]}\bar{X}_{[s]}) = 0. \tag{2.16}$$

If both r and s are not equal to l, (2.16) holds trivially since $X_{[l]i}$'s are independent from $X_{[r]i}$'s and $X_{[s]i}$'s. In the following, we verify (2.16) for the case that either r or s equals l. Without loss of generality, let $r = l$. We have

$$\begin{aligned}
&\text{Cov}(\bar{X}_{[ll]} - \bar{X}_{[l]}^2, \bar{X}_{[ls]} - \bar{X}_{[l]}\bar{X}_{[s]}) \\
&= E[\text{Cov}(\bar{X}_{[ll]} - \bar{X}_{[l]}^2, \bar{X}_{[ls]} - \bar{X}_{[l]}\bar{X}_{[s]}|X_{[s]i}, i = 1, \ldots, m)] \\
&= E[\frac{1}{m}\sum_{i=1}^{m} X_{[s]i}\text{Cov}(\bar{X}_{[ll]} - \bar{X}_{[l]}^2, X_{[l]i})] - E[\bar{X}_{[s]}\text{Cov}(\bar{X}_{[ll]} - \bar{X}_{[l]}^2, \bar{X}_{[l]})] \\
&= 0.
\end{aligned}$$

As a special case of Lemma 2.5, we have

$$\mathrm{Var}(\hat{\tau}_2^2) < \mathrm{Var}(\hat{\tau}_1^2).$$

We summarize what we have found so far to conclude this section. We have investigated possible ways of estimating the variance of a standard RSS estimate of the population mean. We compared different estimates of this variance and found the one with the smallest variance. We recommend that the estimate with the smallest variance be used in the pivot for the construction of confidence intervals or testing hypotheses for the population mean.

Thus, the pivot recommended is

$$Z = \frac{\sqrt{mk}(\bar{X}_{\text{RSS}} - \mu)}{\hat{\tau}_2}.$$

2.5.3 Comparison of power between tests based on RSS and SRS

In this section, we make a power comparison between the test using pivot Z based on RSS and the test using pivot Z_2 based on SRS by a simulation study. In the simulation study, the values of k and m are taken as $k = 3, 4, 5$, and $m = 15, 20, 25$. The underlying distribution is taken as normal distributions $N(\mu, 1)$ with a number of μ values. The power of the SRS-based test is computed by using the standard normal distribution since the sample size $N = mk$ is large. The power of the RSS based test is simulated. In Table 2.3 below, we report the power of the two tests for testing $H_0 : \mu = 0$ against $H_1 : \mu > 0$ at the significance level $\alpha = 0.05$. The efficiency of RSS over SRS is obvious from Table 2.3.

2.6 Estimation of quantiles

This section is devoted to the estimation of population quantiles by using a balanced ranked set sample. We first give the definition of the ranked set sample quantiles analogous to the simple random sample quantiles and investigate their properties. Then we consider inference procedures such as the construction of confidence intervals and the testing of hypotheses for the quantiles. Finally we make a comparison between RSS quantile estimates and SRS quantile estimates in terms of their asymptotic variances.

2.6.1 Ranked set sample quantiles and their properties

Let the ranked-set empirical distribution function be defined as

$$\hat{F}_{\text{RSS}}(x) = \frac{1}{mk} \sum_{r=1}^{k} \sum_{i=1}^{m} I\{X_{[r]i} \leq x\}.$$

Table 2.3. Power of tests based on Z and Z_2

m	k	Pivot\ μ	0	.01	.02	.05	.1	.2	.3	.4	.5
15	3	Z	.05	.07	.08	.12	.25	.59	.97	1	1
		Z_2	.05	.06	.06	.09	.16	.38	.85	.96	1
	4	Z	.06	.09	.13	.19	.40	.83	1	1	1
		Z_2	.05	.06	.07	.10	.19	.46	.93	.99	1
	5	Z	.06	.07	.09	.19	.46	.90	1	1	1
		Z_2	.05	.06	.07	.11	.22	.53	.96	1	1
20	3	Z	.05	.06	.07	.12	.29	.64	.99	1	1
		Z_2	.05	.06	.07	.10	.19	.46	.93	.99	1
	4	Z	.06	.09	.12	.22	.43	.90	1	1	1
		Z_2	.05	.06	.07	.12	.23	.56	.97	1	1
	5	Z	.05	.08	.09	.22	.51	.94	1	1	1
		Z_2	.05	.06	.07	.13	.26	.64	.99	1	1
25	3	Z	.05	.06	.07	.16	.33	.78	1	1	1
		Z_2	.05	.06	.07	.11	.22	.53	.97	.99	1
	4	Z	.06	.07	.09	.12	.46	.92	1	1	1
		Z_2	.05	.06	.07	.13	.26	.72	1	1	1
	5	Z	.04	.06	.11	.22	.58	.98	1	1	1
		Z_2	.05	.06	.08	.14	.30	.72	1	1	1

Let $n = mk$. For $0 < p < 1$, the pth ranked-set sample quantile, denoted by $\hat{x}_n(p)$, is then defined as the pth quantile of $\hat{F}_{\text{RSS}}$, i.e.,

$$\hat{x}_n(p) = \inf\{x : \hat{F}_{\text{RSS}}(x) \geq p\}.$$

We also define the ranked-set order statistics as follows. Let the $X_{[r]i}$'s be ordered from the smallest to the largest and denote the ordered quantities by

$$Z_{(1:n)} \leq \cdots \leq Z_{(j:n)} \leq \cdots \leq Z_{(n:n)}.$$

The $Z_{(j:n)}$'s are then referred to as the ranked-set order statistics.

The following results are parallel to those on simple random sample quantiles.

Let the pth quantile of F be denoted by $x(p)$. First, we have that $\hat{x}_n(p)$ converges with probability one to $x(p)$. Indeed, more strongly, we have

Theorem 2.6. *Suppose that the ranking mechanism in RSS is consistent. Then, with probability 1,*

$$|\hat{x}_n(p) - x(p)| \leq \frac{2(\log n)^2}{f(x(p))n^{1/2}},$$

for all sufficiently large n.

Next, we have a Bahadur representation for the ranked-set sample quantile.

Theorem 2.7. *Suppose that the ranking mechanism in RSS is consistent and that the density function f is continuous at $x(p)$ and positive in a neighborhood of $x(p)$. Then,*

$$\hat{x}_n(p) = x(p) + \frac{p - \hat{F}_{RSS}(x(p))}{f(x(p))} + R_n,$$

where, with probability one,

$$R_n = O(n^{-3/4}(\log n)^{3/4}),$$

as $n \to \infty$.

From the Bahadur representation follows immediately the asymptotic normality of the ranked-set sample quantile.

Theorem 2.8. *Suppose that the same conditions as in Theorem 2.7 hold. Then*

$$\sqrt{n}(\hat{x}_n(p) - x(p)) \to N\left(0, \frac{\sigma^2_{k,p}}{f^2(x(p))}\right),$$

in distribution, where,

$$\sigma^2_{k,p} = \frac{1}{k}\sum_{r=1}^{k} F_{[r]}(x(p))[1 - F_{[r]}(x(p))].$$

In particular, if ranking is perfect, noting that $F_{[r]}(x(p)) = B(r, k+r-1, p)$ in this case,

$$\sigma^2_{k,p} = \frac{1}{k}\sum_{r=1}^{k} B(r, k+r-1, p)[1 - B(r, k+r-1, p)],$$

where $B(r, s, x)$ denotes the distribution function of the beta distribution with parameters r and s.

Notice that, when ranking is perfect, the quantity $\sigma^2_{k,p}$ does not depend on any unknowns, which is practically important as will be seen when the asymptotic normality is applied to develop inference procedures for the population quantile. In general, $F_{[r]}(x(p))$ depends on both the ranking mechanism and the unknown F, and needs to be estimated from the data.

As another immediate consequence of Theorem 2.7, we have a more general result as follows.

Theorem 2.9. *Let $0 < p_1 < \cdots < p_j \cdots < p_l < 1$ be l probabilities. Let $\boldsymbol{\xi} = (x(p_1), \ldots, x(p_l))^T$ and $\hat{\boldsymbol{\xi}} = (\hat{x}_n(p_1), \ldots, \hat{x}_n(p_l))^T$. Then*

$$\sqrt{n}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \to N_l(\mathbf{0}, \Sigma)$$

in distribution where, for $i \le j$, the (i, j)-th entry of Σ is given by

$$\sigma_{ij} = \frac{1}{k}\sum_{r=1}^{k} F_{[r]}(x(p_i))(1 - F_{[r]}(x(p_j)))/[f(x(p_i))f(x(p_j))].$$

We also derive some properties of the ranked-set order statistics. An order statistic $Z_{(k_n:n)}$ is said to be central if $\frac{k_n}{n}$ converges to some p such that $0 < p < 1$ as n goes to infinity. For central ranked-set order statistics, we have the following analogue of the results for simple random sample order statistics:

Theorem 2.10. *(i) If* $\frac{k_n}{n} = p + o(n^{-1/2})$ *then*

$$Z_{(k_n:n)} = x(p) + \frac{\frac{k_n}{n} - \hat{F}_{RSS}(x(p))}{f(x(p))} + R_n,$$

where, with probability 1,

$$R_n = O(n^{-3/4}(\log n)^{3/4}),$$

as $n \to \infty$.

(ii) If

$$\frac{k_n}{n} = p + \frac{c}{n^{1/2}} + o(n^{-1/2})$$

then

$$\sqrt{n}(Z_{(k_n:n)} - \hat{x}_n(p)) \to \frac{c}{f(x(p))}$$

with probability 1, and

$$\sqrt{n}(Z_{(k_n:n)} - x(p)) \to N\left(\frac{c}{f(x(p))}, \frac{\sigma^2_{k,p}}{f^2(x(p))}\right)$$

in distribution.

The results in this section can be proved in arguments parallel to the proof of the results on simple random sample quantiles. The details of the proof can be found in Chen [35].

2.6.2 Inference procedures for population quantiles based on ranked set sample

The results in Section 2.6.1 are applied in this section for inference procedures on quantiles such as confidence intervals and hypotheses testing.

(i) Confidence interval based on ranked-set order statistics. To construct a confidence interval of confidence coefficient $1 - 2\alpha$ for $x(p)$, we seek two integers l_1 and l_2 such that $1 \le l_1 < l_2 \le n$ and that

$$P(Z_{(l_1:n)} < x(p) < Z_{(l_2:n)}) = 1 - 2\alpha.$$

We restrict our attention to the intervals with equal tail probabilities, i.e., intervals satisfying

$$P(Z_{(l_1:n)} \leq x(p)) = 1-\alpha, \quad P(Z_{(l_2:n)} \leq x(p)) = \alpha.$$

Then the integers l_1 and l_2 can be found as follows. Let N_r denote the number of $X_{[r]i}$'s with fixed r which are less than or equal to $x(p)$. Let $N = \sum_{r=1}^{k} N_r$. We have

$$P(Z_{(l_1:n)} \leq x(p)) = P(N \geq l_1).$$

Note that the N_r's are independent binomial random variables with $N_r \sim Bi(m, p_r)$ where $p_r = F_{[r]}(x(p))$. Hence

$$P(N \geq l_1) = \sum_{j=l_1}^{n} \sum_{(j)} \prod_{r=1}^{k} \binom{m}{i_r} p_r^{i_r}(1-p_r)^{m-i_r},$$

where the summation $\sum_{(j)}$ is over all k-tuples of integers $(i_1, \ldots, i_k)$ satisfying $\sum_{r=1}^{k} i_r = j$. Then l_1 can be determined such that the sum on the right hand of the above equality is equal to or near $1-\alpha$. Similarly l_2 can be determined. Though not impossible, the computation will be extremely cumbersome. However, when n is large, l_1 and l_2 can be determined approximately as demonstrated below. Note that

$$EN = \sum_{r=1}^{k} mF_{[r]}(x(p)) = mkF(x(p)) = np,$$
$$Var(N) = \sum_{r=1}^{k} mF_{[r]}(x(p))[1-F_{[r]}(x(p))].$$

By the central limit theorem we have that, approximately,

$$\frac{N-np}{\sqrt{\sum_{r=1}^{k} mF_{[r]}(x(p))[1-F_{[r]}(x(p))]}} \sim N(0,1).$$

Hence

$$l_1 \approx np - z_\alpha \sqrt{\sum_{r=1}^{k} mF_{[r]}(x(p))[1-F_{[r]}(x(p))]},$$
$$l_2 \approx np + z_\alpha \sqrt{\sum_{r=1}^{k} mF_{[r]}(x(p))[1-F_{[r]}(x(p))]},$$

where z_α denotes the $(1-\alpha)$th quantile of the standard normal distribution. When ranking is perfect, $F_{[r]}(x(p)) = B(r, k-r+1, p)$, and the intervals above can be completely determined. However, in general, $F_{[r]}(x(p))$ is unknown and has to be estimated. We can take the estimate to be $\hat{F}_{[r]}(\hat{x}_n(p))$, where $\hat{F}_{[r]}(x) = (1/m)\sum_{i=1}^{m} I\{X_{[r]i} \leq x\}$.

For later reference, the interval $[Z_{(l_1:n)},\ Z_{(l_2:n)}]$ is denoted by $\tilde{I}_{S_n}$.

(ii) Confidence interval based on ranked-set sample quantiles. By making use of Theorem 2.8, another asymptotic confidence interval of confidence coefficient $1-2\alpha$ for $x(p)$ based on ranked-set sample quantiles can be constructed as:

$$\left[\hat{x}_n(p) - \frac{z_\alpha}{\sqrt{n}} \frac{\sigma_{k,p}}{f(x(p))},\ \hat{x}_n(p) + \frac{z_\alpha}{\sqrt{n}} \frac{\sigma_{k,p}}{f(x(p))}\right].$$

This interval is denoted by $\tilde{I}_{Q_n}$. Since $\tilde{I}_{Q_n}$ involves the unknown quantity $f(x(p))$, we need to replace it with some consistent estimate in practice. In the next section, we shall consider the estimation of f by the kernel method using RSS data. The RSS kernel estimate of f can well serve the purpose here. Let $\hat{f}_{\text{RSS}}$ denote the RSS kernel estimate of f. Then in $\tilde{I}_{Q_n}$ the unknown $f(x(p))$ can be replaced by $\hat{f}_{\text{RSS}}(\hat{x}_n(p))$. Note that the intervals $\tilde{I}_{S_n}$ and $\tilde{I}_{Q_n}$ are equivalent in the sense that the two intervals are approximately overlapping with each other while the confidence coefficients are the same. It follows from Theorems 2.7 and 2.10 that, with probability 1,

$$Z_{(l_1:n)} - [\hat{x}_n(p) - \frac{z_\alpha}{\sqrt{n}} \frac{\sigma_{k,p}}{f(x(p))}] = o(n^{-1/2}),$$

and

$$Z_{(l_2:n)} - [\hat{x}_n(p) + \frac{z_\alpha}{\sqrt{n}} \frac{\sigma_{k,p}}{f(x(p))}] = o(n^{-1/2}).$$

Noting that the length of the two intervals has order $O(n^{-1/2})$, the equivalence is established. In practice, either of these two intervals could be used.

(iii) Hypothesis testing using ranked-set sample quantiles. The joint asymptotic normality of the ranked-set sample quantiles, as stated in Theorem 2.9, can be used to test hypotheses involving population quantiles. Suppose the null hypothesis is of the form $\boldsymbol{l^T}\boldsymbol{\xi} = c$, where $\boldsymbol{\xi}$ is a vector of quantiles, say, $\boldsymbol{\xi} = (x(p_1), \ldots, x(p_q))^T$, and $\boldsymbol{l}$ and c are given vector and scalar of constants, respectively. The test statistic can then be formed as:

$$S_n = \frac{\sqrt{n}[\boldsymbol{l^T}\hat{\boldsymbol{\xi}} - c]}{\sqrt{\boldsymbol{l^T}\hat{\Sigma}\boldsymbol{l}}},$$

where $\hat{\boldsymbol{\xi}}$ is the vector of the corresponding ranked-set sample quantiles and $\hat{\Sigma}$ is the estimated covariance matrix of $\hat{\boldsymbol{\xi}}$ with its (i,j)-th $(i < j)$ entry given by

$$\hat{\sigma}_{ij} = \frac{1}{k} \sum_{r=1}^{k} p_{ir}(1 - p_{jr})/[\hat{f}(\hat{x}_n(p_i))\hat{f}(\hat{x}_n(p_j))].$$

By Theorem 2.9, the test statistic follows asymptotically the standard normal distribution under the null hypothesis. Hence, the decision rule can be made accordingly.

2.6.3 The relative efficiency of RSS quantile estimate with respect to SRS quantile estimate

We discuss the ARE of the RSS quantile estimate with respect to the SRS quantile estimate in this section. The counterpart of $\hat{x}_n(p)$ in SRS is the pth sample quantile, $\hat{\xi}_{np}$, of a simple random sample of size n. It can be found from any standard text book that $\hat{\xi}_{np}$ has an asymptotic normal distribution with mean $x(p)$ and variance $\frac{p(1-p)}{nf^2(x(p))}$. Hence, the ARE of $\hat{x}_n(p)$ with respect to $\hat{\xi}_{np}$ is given by

$$\text{ARE}(\hat{x}_n(p), \hat{\xi}_{np}) = \frac{p(1-p)}{\frac{1}{k}\sum_{r=1}^{k} F_{[r]}(x(p))[1 - F_{[r]}(x(p))]}.$$

By using the Bahadur representations of $\hat{x}_n(p)$ and $\hat{\xi}_{np}$ and applying Theorem 2.1 to the function

$$h(x) = \frac{p - I\{x \leq x(p)\}}{f(x(p))},$$

we obtain that

$$\text{ARE}(\hat{x}_n(p), \hat{\xi}_{np}) > 1,$$

provided that the ranking mechanism in RSS is consistent.

While $\text{ARE}(\hat{x}_n(p), \hat{\xi}_{np})$ is always greater than 1 for any p, the quantity can differ very much for different values of p. To gain more insight into the nature of the ARE, let us consider the case of perfect ranking. In this case, $F_{[r]}(x(p)) = B(r, k-r+1, p)$, and the relative efficiency depends only on p and k. For convenience, let it be denoted by $\text{ARE}(k, p)$. For fixed k, as a function of p, $\text{ARE}(k, p)$ is symmetric about $p = 0.5$. It achieves its maximum at $p = 0.5$ and damps away towards $p = 0$ and $p = 1$. For fixed p, $\text{ARE}(k, p)$ increases as k increases. For $k = 1, \ldots, 10$, $ARE(k, p)$ is depicted in Figure 2.1. The curves from the bottom to the top correspond to k from 1 to 10.

We can expect the largest gain in efficiency when we estimate the median of a population. The gain is quite significant even for small set sizes. For $k = 3, 4, 5$, the AREs are, respectively, 1.6, 1.83, 2.03 — but relatively poor compared with the RE's for mean for most distributions — see Table 2.1. In terms of sample sizes, we can reduce the sample size of an SRS by a half through RSS with set size $k = 5$ while maintaining the same accuracy. However, the efficiency gain in the estimation of extreme quantiles is almost negligible. To improve the efficiency for the estimation of extreme quantiles, other RSS procedures must be sought. We will get back to this point in Chapter 4.

2.7 Estimation of density function with ranked set sample

In the context of RSS, the need for density estimation arises in certain statistical procedures. For example, the confidence interval and hypothesis testing

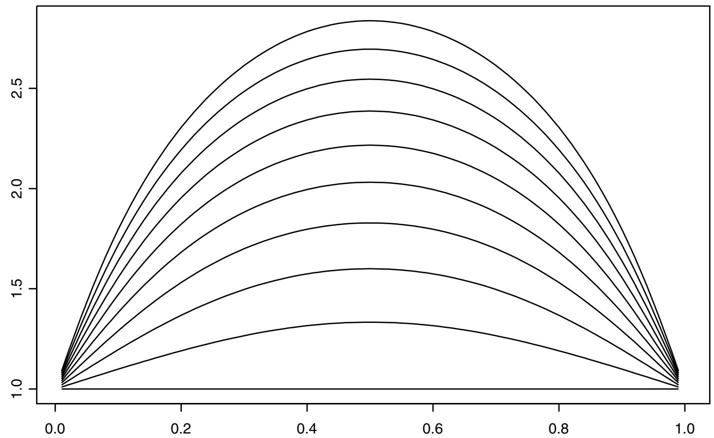

Fig. 2.1. The asymptotic relative efficiency of RSS quantile estimates

procedures based on ranked-set sample quantiles considered in Section 2.6 need to have an estimate of the values of the density function at certain quantiles. On the other hand, density estimation has its own independent interest. A density estimate can reveal important features such as the skewness and multimodality of the underlying distribution. A density estimate is an ideal tool for the presentation of the data back to the clients in order to provide explanations and illustrations of the conclusions that have been obtained. In this section, we take up the task for developing methods of density estimation using RSS data. In Section 2.7.1, the estimate of the density function f is given and its properties are investigated. In Section 2.7.2, the relative efficiency of the density estimate using RSS data with respect to its counterpart in SRS is discussed.

2.7.1 RSS density estimate and its properties

There is a vast literature on density estimation in SRS. A variety of methods have been proposed and developed including the nearest neighbor, the kernel, the maximum penalized likelihood and the adaptive kernel method, etc.. A good reference on the general methodology of density estimation is Silverman [153]. Each of the various methods has its own merits and drawbacks. There is no universal agreement as to which method should be used. We will focus our attention on the kernel method of density estimation and its ramifications. We choose to deal with the kernel method partly because it is a good choice in many practical problems and partly because its mathematical properties are well understood.

To facilitate our discussion that follows, before we deal with the properties of the RSS estimate, we give the definition of the SRS estimate along with some of its properties below. Based on a simple random sample $X_1, \ldots, X_n$, the kernel estimate of f is given by

$$\hat{f}_{\text{SRS}}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

The mean and variance of $\hat{f}_{\text{SRS}}(x)$ can be easily derived as

$$E\hat{f}_{\text{SRS}}(x) = \int \frac{1}{h} K\left(\frac{x-t}{h}\right) f(t)dt, \tag{2.17}$$

$$Var\hat{f}_{\text{SRS}}(x) = \frac{1}{n}\left\{\int \frac{1}{h^2} K\left(\frac{x-t}{h}\right)^2 f(t)dt - \left[\int \frac{1}{h} K\left(\frac{x-t}{h}\right) f(t)dt\right]^2\right\}. \tag{2.18}$$

To motivate the definition of the estimate with RSS data, note that, from the fundamental equality, we have

$$f = \frac{1}{k} \sum_{r=1}^{k} f_{[r]}, \tag{2.19}$$

where $f_{[r]}$ denotes the density function corresponding to $F_{[r]}$. The sub-sample, $X_{[r]i}, i = 1, \ldots, m$, is indeed a simple random sample from the distribution with pdf $f_{[r]}$. Hence $f_{[r]}$ can be estimated by the usual kernel method using the sub-sample. The kernel estimate, $\hat{f}_{[r]}$, of $f_{[r]}$ at x based on the sub-sample is defined as

$$\hat{f}_{[r]}(x) = \frac{1}{mh} \sum_{i=1}^{m} K\left(\frac{x - X_{[r]i}}{h}\right),$$

where K is a kernel function and h is the bandwidth to be determined. Thus a natural definition of the kernel estimate of f is given by

$$\hat{f}_{\text{RSS}}(x) = \frac{1}{k} \sum_{r=1}^{k} \hat{f}_{[r]}(x) = \frac{1}{kmh} \sum_{r=1}^{k} \sum_{i=1}^{m} K\left(\frac{x - X_{[r]i}}{h}\right).$$

It follows from (2.19) that

$$E\hat{f}_{\text{RSS}}(x) = \frac{1}{kh} \sum_{r=1}^{k} EK\left(\frac{x - X_{[r]i}}{h}\right)$$

$$
\begin{aligned}
&= \frac{1}{k}\sum_{r=1}^{k}\int K\left(\frac{x-t}{h}\right) f_{[r]}(t)dt \\
&= \int \frac{1}{h}K\left(\frac{x-t}{h}\right) f(t)dt = E\hat{f}_{\text{SRS}}(x), \\
Var\hat{f}_{\text{RSS}}(x) &= \frac{1}{mk^2}\sum_{r=1}^{k} Var\frac{1}{h}K\left(\frac{x-X_{[r]}}{h}\right) \\
&= \frac{1}{mk^2}\sum_{r=1}^{k}\left\{E\left[\frac{1}{h}K\left(\frac{x-X_{[r]}}{h}\right)\right]^2 - \left[E\frac{1}{h}K\left(\frac{x-X_{[r]}}{h}\right)\right]^2\right\} \\
&= \frac{1}{mk}\left\{E\left[\frac{1}{h}K\left(\frac{x-X}{h}\right)\right]^2 - \frac{1}{k}\sum_{r=1}^{k}\left[E\frac{1}{h}K\left(\frac{x-X_{[r]}}{h}\right)\right]^2\right\} \\
&= \text{Var}\hat{f}_{\text{SRS}}(x) \\
&\quad + \frac{1}{mk}\left\{\left[E\frac{1}{h}K\left(\frac{x-X}{h}\right)\right]^2 - \frac{1}{k}\sum_{r=1}^{k}\left[E\frac{1}{h}K\left(\frac{x-X_{[r]}}{h}\right)\right]^2\right\}.
\end{aligned}
$$

It follows again from (2.19) that

$$
E\frac{1}{h}K\left(\frac{x-X}{h}\right) = \frac{1}{k}\sum_{r=1}^{k} E\frac{1}{h}K\left(\frac{x-X_{[r]}}{h}\right).
$$

By the Cauchy-Schwarz inequality we have

$$
\left[E\frac{1}{h}K\left(\frac{x-X}{h}\right)\right]^2 < \frac{1}{k}\sum_{r=1}^{k}\left[E\frac{1}{h}K\left(\frac{x-X_{[r]}}{h}\right)\right]^2.
$$

Summarizing the argument above we conclude that $\hat{f}_{\text{RSS}}(x)$ has the same expectation as $\hat{f}_{\text{SRS}}(x)$ and a smaller variance than $\hat{f}_{\text{SRS}}(x)$. This implies that the RSS estimate has a smaller mean integrated square error (MISE) than the SRS estimate. The MISE of an estimate $\hat{f}$ of f is defined as $\text{MISE}(\hat{f}) = E\int[\hat{f}(x) - f(x)]^2 dx$. The conclusion holds whether or not ranking is perfect. In what follows we assume that f has certain derivatives and that K satisfies the conditions: (i) K is symmetric and (ii) $\int K(t)dt = 1$ and $\int t^2K(t)dt \neq 0$.

Lemma 2.11. *Under the above assumptions about f and K, for fixed k, as $h \to 0$,*

$$
\begin{aligned}
&\left[E\frac{1}{h}K\left(\frac{x-X}{h}\right)\right]^2 - \frac{1}{k}\sum_{r=1}^{k}\left[E\frac{1}{h}K\left(\frac{x-X_{[r]}}{h}\right)\right]^2 \\
&= [f^2(x) - \frac{1}{k}\sum_{r=1}^{k} f_{[r]}^2(x)] + O(h^2).
\end{aligned}
$$

Lemma 2.11 can be proved by a straightforward calculation involving Taylor expansions of f and $f_{[r]}$'s.

Let

$$\Delta(f,k) = \int [\frac{1}{k}\sum_{r=1}^{k} f_{[r]}^2(x) - f^2(x)]dx.$$

Note that $\Delta(f,k)$ is always greater than zero. We have the following result.

Theorem 2.12. *Suppose that the same bandwidth is used in both* $\hat{f}_{SRS}$ *and* $\hat{f}_{RSS}$*. Then, for fixed* k *and large* n*,*

$$MISE(\hat{f}_{RSS}) = MISE(\hat{f}_{SRS}) - \frac{1}{n}\Delta(f,k) + O\left(\frac{h^2}{n}\right).$$

We now consider the special case of perfect ranking and derive some asymptotic results that shed lights on the properties of the RSS density estimate. When ranking is perfect, $f_{[r]} = f_{(r)}$, the pdf of the rth order statistic. First, we have

Lemma 2.13. *If, for* $r = 1, \ldots, k$, $f_{(r)}$ *is the density function of the* r*-th order statistic of a sample of size* k *from a distribution with density function* f*, then we have the representation*

$$\frac{1}{k}\sum_{r=1}^{k} f_{(r)}^2(x) = kf^2(x)P(Y = Z),$$

where Y *and* Z *are independent with the same binomial distribution* $B(k-1, F(x))$*. Furthermore*

$$P(Y = Z) = \frac{1}{\sqrt{4\pi k F(x)[1-F(x)]}} + o\left(\frac{1}{k}\right).$$

Proof: When ranking is perfect, we have

$$f_{(r)}(x) = \frac{k!}{(r-1)!(k-r)!} F^{r-1}(x)[1-F(x)]^{k-r} f(x).$$

Thus, we can write

$$\begin{aligned}
&\frac{1}{k}\sum_{r=1}^{k} f_{(r)}^2(x) \\
&= \frac{1}{k}\sum_{r=1}^{k} [\frac{k!}{(r-1)!(k-r)!} F^{r-1}(x)(1-F(x))^{k-r} f(x)]^2 \\
&= kf^2(x)\sum_{j=0}^{k-1}\left[\binom{k-1}{j} F^j(x)(1-F(x))^{k-1-j}\right]^2.
\end{aligned}$$

The first part of the lemma is proved. The second part follows from the Edgeworth expansion of the probability $P(Y = Z)$.

Remark: Our computation for certain values of $F(x)$ has revealed that the approximation to the probability $P(Y = Z)$ is quite accurate for large or moderate k. For small k, $P(Y = Z)$ is slightly bigger than the approximation. However, the approximation can well serve our theoretical purpose.

In what follows we denote, for any function g, the integral $\int x^l g(x)dx$ by $i_l(g)$. Applying Lemmas 2.11 and 2.13, we have

Lemma 2.14. *If ranking is perfect, then, for a fixed large or moderate k, as $n \to \infty$, we have*

$$MISE(\hat{f}_{RSS}) = MISE(\hat{f}_{SRS}) - \frac{1}{n}[\sqrt{k}\delta(f) - i_0(f^2)] + O(\frac{h^2}{n}),$$

where

$$\delta(f) = \frac{1}{2\sqrt{\pi}} \int \frac{f^2(x)}{\sqrt{F(x)[1 - F(x)]}} dx.$$

Lemma 2.14 shows that the RSS estimate reduces the MISE of the SRS estimate at the order $O(n^{-1})$ by an amount which increases linearly in $\sqrt{k}$.

The results derived in this section can be extended straightforwardly to the adaptive kernel estimation described in Silverman ([153], p101). The ordinary kernel estimate usually suffers a slight drawback that it has a tendency of undersmoothing at the tails of the distribution. The adaptive kernel estimate overcomes this drawback and provides better estimates at the tails. We do not discuss the adaptive kernel estimate further. The reader is referred to Silverman ([153], Chapter 5) for details.

2.7.2 The relative efficiency of the RSS density estimate with respect to its SRS counterpart

In this section, we investigate the efficiency of the RSS estimate relative to the SRS estimate in terms of the ratio of the MISE's.

First, we derive an asymptotic expansion for the MISE of the SRS estimate. By Taylor expansion of the density function f at x under the integrals in (2.17) and (2.18) after making the change of variable $y = (x - t)/h$, we have

$$\text{bias}(\hat{f}_{\text{SRS}}(x)) = \frac{1}{2} i_2(K) f^{''}(x) h^2 + O(h^4),$$
$$\text{Var}(\hat{f}_{\text{SRS}}(x)) = \frac{1}{nh} i_0(K^2) f(x) - \frac{1}{n} f^2(x) + O(\frac{h^2}{n}).$$

Hence

$$\text{MISE}(\hat{f}_{\text{SRS}}) = \int [Var(\hat{f}_{\text{SRS}}(x)) + bias^2(\hat{f}_{\text{SRS}}(x))] dx$$

$$= \frac{1}{nh} i_0(K^2) + \frac{1}{4} i_2^2(K) i_0(f^{''2}) h^4 - \frac{1}{n} i_0(f^2) + O(\frac{h^2}{n}) + O(h^6). \tag{2.20}$$

Minimizing the leading terms with respect to h, we have that the minimum is attained at

$$h_{opt} = i_2(K)^{-2/5} \left[\frac{i_0(K^2)}{i_0(f''^2)} \right]^{1/5} n^{-1/5}. \tag{2.21}$$

Substituting (2.21) into (2.20) yields

$$\text{MISE}(\hat{f}_{\text{SRS}}) = \frac{5}{4} C(K) i_0(f^{''2})^{1/5} n^{-4/5} - i_0(f^2) n^{-1} + O(n^{-6/5}), \tag{2.22}$$

where $C(K) = i_2(K)^{2/5} i_0(K^2)^{4/5}$.

Combining (2.22) with Theorem 2.12, we have that the relative efficiency of the RSS estimate to the SRS estimate is approximated by

$$\frac{\text{MISE}(\hat{f}_{\text{SRS}})}{\text{MISE}(\hat{f}_{\text{RSS}})} \approx \left[1 - \frac{\Delta(f,k)}{(5/4)C(K) i_0(f''^2) n^{1/5} - i_0(f^2)} \right]^{-1}. \tag{2.23}$$

When ranking is perfect and k is large or moderate, the relative efficiency has the approximate expression:

$$\frac{\text{MISE}(\hat{f}_{\text{SRS}})}{\text{MISE}(\hat{f}_{\text{RSS}})} \approx \left[1 - \frac{\sqrt{k}\delta(f) - i_0(f^2)}{(5/4)C(K) i_0(f''^2) n^{1/5} - i_0(f^2)} \right]^{-1}. \tag{2.24}$$

We can conclude qualitatively from the approximation in (2.24) that (i) the efficiency of the RSS kernel estimate relative to the SRS kernel estimate increases as k increases at the rate $O(k^{1/2})$, (ii) the relative efficiency damps away as n gets large but the speed at which it damps away is very low (of order $O(n^{-1/5})$). Therefore, we can expect that for small or moderate sample size n the gain in efficiency by using RSS will be substantial. RSS can only reduce variance and the order $O(n^{-1})$ at which the variance is reduced is common in all the other statistical procedures such as the estimation of mean, variance and cumulative distribution, etc.. However, while the reduction in MISE is at order $O(n^{-1})$, the MISEs have order $O(n^{-4/5})$. When n is large, the component of the MISE at order $O(n^{-4/5})$ dominates. This explains the fact that the relative efficiency damps away as n goes to infinity.

A major application of the RSS density estimation is for estimating the density at certain particular points, e.g., certain quantiles. It is desirable to compare the performance of the RSS estimate and the SRS estimate at particular values of x. An argument similar to the global comparison leads to the following results. The MSEs of the two estimates at x have the equal components at order lower than $O(n^{-1})$. The components of the MSEs at order $O(n^{-1})$

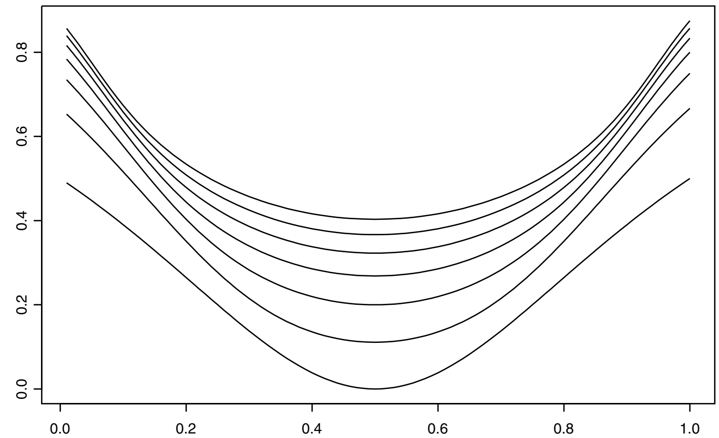

Fig. 2.2. The relative reduction in variance at order $O(n^{-1})$ of the RSS density estimate

of the SRS estimate and the RSS estimate are, respectively, $-(1/n)f^2(x)$ and $-(1/nk)\sum_{r=1}^{k} f_{[r]}^2(x)$. The relative reduction in MSE is then given by

$$\Lambda_f = [\frac{1}{k}\sum_{r=1}^{k} f_{[r]}^2(x) - f^2(x)]/\frac{1}{k}\sum_{r=1}^{k} f_{[r]}^2(x).$$

If ranking is perfect and $x = x(p)$, the pth quantile of the underlying distribution, then Λ_f becomes

$$\Lambda_f(p) = 1 - \left\{k\sum_{r=0}^{k-1}\left[\binom{k-1}{r} p^r(1-p)^{k-1-r}\right]^2\right\}^{-1}.$$

For $k = 2, \ldots, 8$, the relative reduction $\Lambda_f(p)$ is plotted against p in Figure 2.2. The curves from bottom to top correspond in turn to $k = 2, \ldots, 8$. It can be seen from the figure that the relative reduction increases symmetrically as p goes from 0.5 to 0 and 1. This indicates that the advantage of RSS is greater for estimating the densities at the tails of the distribution.

To end this section, we present some results of a simulation study. The following three distributions are used in the simulation: the standard normal distribution $N(0, 1)$, the gamma distribution with shape parameter $\beta = 3$ and the extreme value distribution with location parameter $\mu = 0$ and scale parameter $\sigma = 1$. The following combinations of set size k and cycle number m are considered: (1) $k = 6, m = 4$, (2) $k = 6, m = 8$, (3) $k = 8, m = 3$ and

Table 2.4. The comparison of the MISE between the RSS and SRS kernel estimates

Distribution	k	m	MISE.SRS	MISE.RSS	Rel. Eff.
Normal (0, 1)	6	4	0.0182(0.016)	0.0112(0.009)	1.625
	6	8	0.0102(0.008)	0.0067(0.005)	1.522
	8	3	0.0182(0.016)	0.0100(0.008)	1.820
	8	6	0.0102(0.008)	0.0062(0.004)	1.645
Gamma $\beta = 3$	6	4	0.0123(0.010)	0.0086(0.006)	1.430
	6	8	0.0074(0.006)	0.0053(0.003)	1.396
	8	3	0.0123(0.010)	0.0078(0.005)	1.577
	8	6	0.0074(0.006)	0.0049(0.003)	1.510
Extreme Value $\mu = 0$ $\sigma = 1$	6	4	0.0171(0.015)	0.0113(0.008)	1.513
	6	8	0.0098(0.007)	0.0068(0.004)	1.441
	8	3	0.0171(0.015)	0.0105(0.007)	1.629
	8	6	0.0098(0.007)	0.0065(0.004)	1.508

(4) $k = 8, m = 6$. For each combination of k and m and each distribution, 5000 SRS samples of size mk and 5000 RSS samples with set size k and cycle number m are generated using the random number generating functions in Splus 4.5. For each of these samples, a kernel estimate of the underlying density function is obtained. The Epanechnikov kernel, which is optimal in the sense that it minimizes $C(K)$ among certain class of kernels, is used in all the estimates. The Epanechnikov kernel is given by

$$K(x) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{x^2}{5}), & -\sqrt{5} \le x \le \sqrt{5}, \\ 0, & \text{otherwise.} \end{cases}$$

The bandwidth h is determined by

$$h = \frac{5}{4} C(K) \left[\frac{3}{8\sqrt{\pi}} \right]^{-1/5} A n^{-1/5},$$

where $A = \min\{$standard deviation of the sample, interquartile range of the sample $/1.34\}$. Since A differs from sample to sample, the bandwidths h used in the estimates are not exactly the same. However, they have the same order $O(n^{-1/5})$. For each estimate the integrated square error (ISE) $\int [\hat{f}(x) - f(x)]^2 dx$ is computed by numerical method. Then the average and the standard deviation of the ISEs of the 5000 RSS estimates are computed. The same is done to the 5000 SRS estimates. The averages of the ISEs are taken as the estimate of the MISEs and are compared between the RSS and the SRS estimates. The estimated MISEs are given in Table 2.4. The numbers in the parentheses are the standard deviations. The improvement on the MISE by using RSS is quite significant. The assertions we made from (2.24) manifest in the table. For fixed k, the relative efficiency decreases slowly as the sample size increases. On the other hand, for fixed sample size, the increment in relative efficiency is much faster when k increases.

The other methods of density estimation can also be investigated in a similar manner as we have done in this section. Because of the fact that a ranked-set sample indeed contains more information than a simple random sample of the same size, we expect that for all such methods the RSS version will provide a better estimate than the SRS version.

2.8 M-estimates with RSS data

The idea of M-estimates arises out of concern on the robustness of statistical procedures. For example, the usual estimate of the population mean, namely the sample mean, is not robust if the underlying distribution is heavily tailed. This problem does not go away in RSS. In this section, we investigate the properties of the RSS M-estimates including their asymptotic distribution and their efficiency relative to their counterparts with SRS. The RSS estimates are defined and their asymptotic properties are dealt with in Section 2.8.1. The relative efficiency of the RSS M-estimates is discussed in Section 2.8.2.

2.8.1 The RSS M-estimates and their asymptotic properties

Let $\psi(x)$ be an appropriate function. Define the functional $T(F)$ over all distribution functions as the solution of $\lambda_F(t) = \int \psi(x-t)dF(x) = 0$, if exists. In generic notation, if F is an unknown distribution function and $\hat{F}$ is an appropriate estimate of F, then $T(\hat{F})$ is called an M-estimate of $T(F)$. Let $\hat{T}_n = T(\hat{F}_{\text{RSS}})$, i.e., $\lambda_{\hat{F}_{\text{RSS}}}(\hat{T}_n) = 0$, which defines the RSS M-estimate of $T(F)$.

The following lemma is used later.

Lemma 2.15. *Suppose that $\psi(x)$ is an odd function and F is a symmetric location distribution, then the population mean μ is a solution of $\lambda_F(t) = 0$, i.e., $\mu = T(F)$, and, further, μ satisfies*

$$\int_{-\infty}^{+\infty} \psi(x-t)dF_{(r)}(x) + \int_{-\infty}^{+\infty} \psi(x-t)dF_{(k-r+1)}(x) = 0.$$

The following theorem gives conditions under which the RSS M-estimate exists and is also consistent.

Theorem 2.16. *Suppose that $\psi(x)$ is odd, continuous and either monotone or bounded, and that F is a symmetric location distribution. Then there is a solution sequence $\{\hat{T}_n\}$ of $\lambda_{\hat{F}_{RSS}}(t) = 0$ such that $\{\hat{T}_n\}$ converges to μ with probability 1.*

There are other conditions on ψ so that Theorem 2.16 holds. However, since the ψ's in practical applications satisfy the conditions in Theorem 2.16, we will concentrate on the ψ's satisfying these conditions. In particular, we will later consider the following two ψ functions. The first ψ is given by

$$\psi_1 = \begin{cases} -1.5, & x < -1.5, \\ x, & |x| \leq 1.5, \\ 1.5, & x > 1.5. \end{cases}$$

The corresponding M-estimator is a type of Winsorized mean. The other ψ is a smoothed "Hampel" given by

$$\psi_2 = \begin{cases} \sin(x/2.1), & |x| < 2.1\pi, \\ 0, & |x| \geq 2.1\pi. \end{cases}$$

Let

$$A_{(r)}(t) = \int \psi^2(x-t)dF_{(r)}(x) - [\int \psi(x-t)dF_{(r)}(x)]^2,$$
$$\lambda_F^{'}(T(F)) = \left.\frac{d\int \psi(x-t)dF(x)}{dt}\right|_{t=T(F)}.$$

In the following, we give three sets of conditions each of which, together with the conditions on F given in Theorem 2.16, guarantees the asymptotic normality of the sequence $\hat{T}_n$.

A1 $\psi(x)$ is odd and monotone; $\lambda_F(t)$ is differentiable at $t = \mu$, with $\lambda_F'(\mu) \neq 0$; $\int \psi^2(x-t)dF(x)$ is finite for t in a neighborhood of μ and is continuous at $t = \mu$.

A2 $\psi(x)$ is odd, continuous and satisfies $\lim_{t\to\mu} \parallel \psi(\cdot,t) - \psi(\cdot,\mu) \parallel_V = 0$; $\int \psi^2(x-t)dF(x) < \infty$ and $\lambda_F(t)$ is differentiable at $t = \mu$, with $\lambda_F'(\mu) \neq 0$.

A3 $\psi(x)$ is odd and uniformly continuous; $\int \partial\psi(x-t)/\partial t|_{t=\mu}dF(x)$ is finite and nonzero; $\int \psi^2(x-\mu)dF(x) < \infty$.

Theorem 2.17. *Assume that F is a symmetric location distribution. Then, under either (**A1**), (**A2**) or (**A3**),*

$$\sqrt{n}(\hat{T}_n - \mu) \to N(0, \sigma^2_{RSS}(F)), \tag{2.25}$$

*in distribution, where under (**A1**) and (**A2**),*

$$\sigma^2_{RSS}(F) = \frac{1}{k}\sum_{r=1}^{k} A_{(r)}(\mu) \Big/ \left[\frac{1}{k}\sum_{r=1}^{k}\lambda'_{F_{(r)}}(\mu)\right]^2,$$

*and under (**A3**),*

$$\sigma^2_{RSS}(F) = \frac{1}{k}\sum_{r=1}^{k} A_{(r)}(\mu) \Big/ \left[\frac{1}{k}\sum_{r=1}^{k}\int \partial\psi(x-t)/\partial t|_{t=\mu}dF_{(r)}(x)\right]^2.$$

The results in this section are straightforward extensions of the corresponding results in SRS. A sketch of the proof can be found in Zhao and Chen [175].

2.8.2 The relative efficiency of the RSS M-estimates

In this section, we deal with the ARE of RSS M-estimates with respect to SRS M-estimates. The SRS M-estimate of μ is given by $\tilde{T}_n = T(F_n)$ where F_n is the empirical distribution of a simple random sample of size n. The SRS M-estimate has asymptotically a normal distribution with mean μ and variance $\sigma^2_{\text{SRS}}(F)$ given, depending on the assumptions on $\psi(x)$, by either

$$\int \psi^2(x-\mu)dF(x)/[\lambda'_F(\mu)]^2$$

or

$$\int \psi^2(x-\mu)dF(x)/[\int \partial\psi(x-t)/\partial t|_{t=\mu}dF(x)]^2.$$

Note that, because of the fundamental equality,

$$\frac{1}{k}\sum_{r=1}^{k}\lambda'_{F_{(r)}}(\mu) = \lambda'_F(\mu);$$

$$\frac{1}{k}\sum_{r=1}^{k}\int \partial\psi(x-t)/\partial t|_{t=\mu}dF_{(r)}(x) = \int \partial\psi(x-t)/\partial t|_{t=\mu}dF(x);$$

$$\frac{1}{k}\sum_{r=1}^{k}\int \psi^2(x-t)dF_{(r)}(x) = \int \psi^2(x-t)dF(x).$$

Hence the ARE of the RSS M-estimate is given by

$$ARE(\hat{T}_n, \tilde{T}_n) = \frac{\int \psi^2(x-\mu)dF(x)}{\int \psi^2(x-\mu)dF(x) - \frac{1}{k}\sum_{r=1}^{k}[\int \psi(x-\mu)dF_{(r)}(x)]^2}.$$

It is obvious that the ARE is greater than 1. The ARE's of the RSS M-estimates with ψ_1 and ψ_2 for the $Cauchy(0,1)$, $N(0,1)$ and some contaminated normal distributions are given, respectively, in Table 2.5 and Table 2.6. It is evident that, as expected, RSS is much more efficient than SRS.

2.9 Appendix: Technical details for Section 2.4*

We first state and prove a general result. Let $Y_1, \cdots, Y_n$ be independent random variables with $E(Y_i) = \nu_i$, $\text{Var}(Y_i) = \eta_i^2$, $\gamma_i = E(Y_i - \nu_i)^3$, and $\kappa_i = E(Y_i - \nu_i)^4 - 3\eta_i^4$, $i = 1, \cdots, n$. Denote $\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$, $\Delta = diag(\eta_1^2, \cdots, \eta_n^2)$, $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_n)'$. Let Q be an $n \times n$ symmetric matrix.

Table 2.5. The ARE's of RSS w.r.t. SRS with ψ_1 for Cauchy, Normal and some contaminated Normal distributions

Dist.	Cauchy(0, 1)	N(0, 1)	0.9N(0, 1) +0.1N(0, 9)	0.7N(0, 1) +0.3N(0, 9)	0.5N(0, 1) +0.5N(0, 9)
k=2	1.4868	1.4949	1.4958	1.4939	1.4858
k=3	1.9651	1.9866	1.9888	1.9838	1.9625
k=4	2.4364	2.4757	2.4792	2.4700	2.4325
k=5	2.9016	2.9627	2.9673	2.9527	2.8975
k=6	3.3613	3.4481	3.4533	3.4321	3.3582
k=7	3.8162	3.9320	3.9372	3.9085	3.8154
k=8	4.2668	4.4145	4.4193	4.3823	4.2696
k=9	4.7136	4.8959	4.8998	4.8538	4.7212
k=10	5.1570	5.3762	5.3789	5.3232	5.1705

Table 2.6. The ARE's of RSS w.r.t. SRS with ψ_2 for Cauchy, Normal and some contaminated Normal distributions

Dist.	Cauchy(0, 1)	N(0, 1)	0.9N(0, 1) +0.1N(0, 9)	0.7N(0, 1) +0.3N(0, 9)	0.5N(0, 1) +0.5N(0, 9)
k=2	1.3038	1.4878	1.4767	1.4584	1.4380
k=3	1.5374	1.9676	1.9389	1.8922	1.8412
k=4	1.7327	2.4422	2.3886	2.3031	2.2157
k=5	1.9033	2.9131	2.8271	2.6930	2.5658
k=6	2.0570	3.3812	3.2554	3.0639	2.8952
k=7	2.1989	3.8471	3.6740	3.4176	3.2074
k=8	2.3321	4.3113	4.0836	3.7559	3.5051
k=9	2.4589	4.7741	4.4845	4.0805	3.7907
k=10	2.5807	5.2358	4.8772	4.3928	4.0662

Lemma 2.18.

$$E(\boldsymbol{Y}'Q\boldsymbol{Y}) = tr[Q(\Delta + \boldsymbol{\nu}\boldsymbol{\nu}^{'}] \tag{2.26}$$

$$\begin{aligned} Var((\boldsymbol{Y}'Q\boldsymbol{Y}) = &\sum_{i=1}^{n} Q_{ii}^2\kappa_i + 2tr[(Q\Delta)^2] + 4\boldsymbol{\nu}'Q\Delta Q\boldsymbol{\nu} \\ &+4\sum_{i=1}^{n} Q_{ii}\gamma_i(\sum_{l=1}^{n} Q_{il}\nu_l). \end{aligned} \tag{2.27}$$

Proof. (2.26) is obvious. To prove (2.27), write

$$\boldsymbol{Y}'Q\boldsymbol{Y} = (\boldsymbol{Y} - \boldsymbol{\nu})'Q(\boldsymbol{Y} - \boldsymbol{\nu}) + 2\boldsymbol{\nu}'Q(\boldsymbol{Y} - \boldsymbol{\nu}) + \boldsymbol{\nu}'Q\boldsymbol{\nu}.$$

Then

$$\mathrm{Var}(\boldsymbol{Y}'Q\boldsymbol{Y}) = \mathrm{Var}[(\boldsymbol{Y} - \boldsymbol{\nu})'Q(\boldsymbol{Y} - \boldsymbol{\nu})] + 4\mathrm{Var}[\boldsymbol{\nu}'Q(\boldsymbol{Y} - \boldsymbol{\nu})]$$

$$+4\text{Cov}[(\boldsymbol{Y}-\boldsymbol{\nu})'Q(\boldsymbol{Y}-\boldsymbol{\nu}),\boldsymbol{\nu}'Q(\boldsymbol{Y}-\boldsymbol{\nu})].$$

The first term can be simplified as

$$\begin{aligned}
&\text{Var}[(\boldsymbol{Y}-\boldsymbol{\nu})'Q(\boldsymbol{Y}-\boldsymbol{\nu})]\\
&=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}Q_{ij}Q_{kl}\text{Cov}[(Y_i-\nu_i)(Y_j-\nu_j),(Y_k-\nu_k)(Y_l-\nu_l)]\\
&=\sum_{i=1}^{n}Q_{ii}^2\text{Var}[(Y_i-\nu_i)^2]+2\sum_{i\neq j=1}^{n}Q_{ij}^2\text{Var}[(Y_i-\nu_i)(Y_j-\nu_j)]\\
&=\sum_{i=1}^{n}Q_{ii}^2(\kappa_i+2\eta_i^4)+2\sum_{i\neq j=1}^{n}Q_{ij}^2\eta_i^2\eta_j^2\\
&=\sum_{i=1}^{n}Q_{ii}^2\kappa_i+2\sum_{i=1}^{n}\sum_{j=1}^{n}Q_{ij}^2\eta_i^2\eta_j^2\\
&=\sum_{i=1}^{n}Q_{ii}^2\kappa_i+2tr[(Q\varDelta)^2].
\end{aligned}$$

The second term simplifies as

$$4\text{Var}[\boldsymbol{\nu}'Q(\boldsymbol{Y}-\boldsymbol{\nu})]=4\boldsymbol{\nu}'Q\varDelta Q\boldsymbol{\nu}.$$

Finally, the third term can be simplified as

$$\begin{aligned}
&4\text{Cov}[(\boldsymbol{Y}-\boldsymbol{\nu})'Q(\boldsymbol{Y}-\boldsymbol{\nu}),\boldsymbol{\nu}'Q(\boldsymbol{Y}-\boldsymbol{\nu})]\\
&=4\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}Q_{ij}Q_{kl}\text{Cov}[(Y_i-\nu_i)(Y_j-\nu_j),(Y_k-\nu_k)\nu_l]\\
&=4\sum_{i=1}^{n}\sum_{l=1}^{n}Q_{ii}Q_{il}\gamma_i\nu_l\\
&=4\sum_{i=1}^{n}Q_{ii}\gamma_i[\sum_{l=1}^{n}Q_{il}\nu_l].
\end{aligned}$$

The lemma then follows.

We now apply the lemma to the special case with Q given by the form

$$Q=A\otimes[I_m-\frac{\mathbf{1}_m\mathbf{1}_m'}{m}]+D\otimes\frac{\mathbf{1}_m\mathbf{1}_m'}{m},$$

where A and D are arbitrary symmetric matrices. For convenience, let $K_1=\mathbf{1}_m\mathbf{1}_m'/m$ and $K_0=I_m-K_1$, and, similarly, let $J_1=\mathbf{1}_k\mathbf{1}_k'/k$ and $J_0=I_k-J_1$. Let

$$\begin{aligned}
&\boldsymbol{\mu}=(\mu_{[1]},\mu_{[2]},\ldots,\mu_{[k]})',\ \varSigma=\text{diag}(\sigma_{[1]}^2,\sigma_{[2]}^2,\ldots,\sigma_{[k]}^2)\\
&\boldsymbol{\nu}=\boldsymbol{\mu}\otimes\mathbf{1}_m,\ \varDelta=\varSigma\otimes I_m.
\end{aligned}$$

Also, we denote by $[D\boldsymbol{\mu}]_r$ the rth element of the vector $D\boldsymbol{\mu}$. Then we have

Lemma 2.19.

$$E(\boldsymbol{X}'Q\boldsymbol{X}) = tr[((m-1)A + D)\Sigma + mD\boldsymbol{\mu}\boldsymbol{\mu}'], \tag{2.28}$$

$$\begin{aligned} Var(\boldsymbol{X}^{'}Q\boldsymbol{X}) = & \frac{1}{m}\sum_{r=1}^{k}[(m-1)a_{rr} + d_{rr}]^2\kappa^*_{[r]} \\ & +2[(m-1)tr(A\Sigma)^2 + tr(D\Sigma)^2] + 4m(D\boldsymbol{\mu})'\Sigma D\boldsymbol{\mu} \\ & +\frac{4}{m}\sum_{r=1}^{k}[(m-1)a_{rr} + d_{rr}]\gamma^*_{[r]}[D\boldsymbol{\mu}]_r. \end{aligned} \tag{2.29}$$

Proof. Using the special structure of $\boldsymbol{\nu}$, Q and Δ, we get

$$\begin{aligned} & \operatorname{tr}[Q(\Delta + \boldsymbol{\nu}\boldsymbol{\nu}')] \\ = & \operatorname{tr}[(A\otimes K_0 + D\otimes K_1)(\Sigma\otimes \boldsymbol{I}_m + \boldsymbol{\mu}\otimes \mathbf{1}_m(\boldsymbol{\mu}\otimes\mathbf{1}_m)')] \\ = & \operatorname{tr}[A\Sigma\otimes K_0 + D\Sigma\otimes K_1] + \boldsymbol{\mu}'A\boldsymbol{\mu}\mathbf{1}_m^{'}K_0\mathbf{1}_m + \boldsymbol{\mu}^{'}D\boldsymbol{\mu}\mathbf{1}_m^{'}K_1\mathbf{1}_m \\ = & (m-1)\operatorname{tr}(A\Sigma) + \operatorname{tr}(D\Sigma) + m\boldsymbol{\mu}^{'}D\boldsymbol{\mu}. \end{aligned}$$

since $\operatorname{tr}(A\otimes D) = \operatorname{tr}(A)\operatorname{tr}(D)$, $\mathbf{1}_m^{'}K_0\mathbf{1}_m = 0$ and $\mathbf{1}_m^{'}K_1\mathbf{1}_m = m$. Hence (2.28) is proved.

Note that, in view of the special structure of Q and $\boldsymbol{X}$,

$$\begin{aligned} & \kappa_j = \kappa^*_{[r]},\ \gamma_j = \gamma^*_{[r]},\ Q_{jj} = \frac{(m-1)a_{rr}}{m} + \frac{d_{rr}}{m}, \\ & j = (r-1)m+1,\cdots,rm, \quad r = 1,\cdots,k. \end{aligned} \tag{2.30}$$

Hence, we readily obtain

$$\sum_{j=1}^{n} Q_{jj}^2\kappa_j = m\sum_{r=1}^{k}[\frac{(n-1)a_{rr}}{m} + \frac{d_{rr}}{m}]^2\kappa^*_{[r]}, \tag{2.31}$$

where $n = mk$. Next, note that in view of idempotence of K_0 and K_1 and the fact that $K_0K_1 = \mathbf{0}$, we get

$$\begin{aligned} \operatorname{tr}(Q\Delta)^2 & = \operatorname{tr}[(A\otimes K_0 + D\otimes K_1)(\Sigma\otimes\boldsymbol{I}_m)]^2 \\ & = \operatorname{tr}[(A\Sigma\otimes K_0 + D\Sigma\otimes K_1)^2] \\ & = \operatorname{tr}[A\Sigma A\Sigma\otimes K_0 + D\Sigma D\Sigma\otimes K_1] \\ & = (m-1)\operatorname{tr}(A\Sigma)^2 + \operatorname{tr}(D\Sigma)^2. \end{aligned}$$

Moreover, since

$$Q\boldsymbol{\nu} = (A\otimes K_0 + D\otimes K_1)(\boldsymbol{\mu}\otimes\mathbf{1}_m) = D\boldsymbol{\mu}\otimes\mathbf{1}_m,$$

using the special structure of Δ, we get

$$(Q\boldsymbol{\nu})'\Delta(Q\boldsymbol{\nu}) = m(D\boldsymbol{\mu})'\Sigma(D\boldsymbol{\mu}).$$

Finally, using (2.30),

$$\sum_{j=1}^{n} Q_{jj}\gamma_j[\sum_{l=1}^{n} Q_{jl}\nu_l] = m\sum_{r=1}^{k}[\frac{(m-1)a_{rr}}{m} + \frac{d_{rr}}{m}]\gamma^*_{[r]}[D\boldsymbol{\mu}]_r.$$

Hence (2.29) is proved.

We are now in a position to prove Theorem 2.4.

Proof. (a) By (2.28),

$$E[\boldsymbol{X}'Q\boldsymbol{X}] = \mathrm{tr}[[(m-1)A + D]\Sigma + mD\boldsymbol{\mu}\boldsymbol{\mu}'].$$

On the other hand,

$$\sigma^2 = \frac{1}{k}\mathrm{tr}[\Sigma + J_0\boldsymbol{\mu}\boldsymbol{\mu}'].$$

Therefore, $\hat{\sigma}^2$ is unbiased for σ^2 if and only if $\mathrm{tr}[[(m-1)A + D - (1/k)\boldsymbol{I}]\Sigma] = 0$ for all diagonal Σ and $\boldsymbol{\mu}'[mD - (1/k)J_0]\boldsymbol{\mu} = 0$ for all possible values of $\boldsymbol{\mu}$. Hence the proof of (a).

(b) Using the condition of unbiasedness, it is clear that the only term in $\mathrm{Var}(\boldsymbol{X}'Q\boldsymbol{X})$ which depends on the off-diagonal elements of the matrix A is given by $\mathrm{tr}(A\Sigma)^2 = \sum_{r=1}^{k}\sum_{s=1}^{k} a_{rs}^2\sigma_{[r]}^2\sigma_{[s]}^2$. Obviously, a_{rr}'s are fixed from the unbiasedness condition, and the unique choice of a_{rs}'s for $r \neq s$ which makes $\mathrm{Var}(\boldsymbol{X}'Q\boldsymbol{X})$ a minimum is given by $a_{rs} = 0$ for all $r \neq s$. Hence the proof.

2.10 Bibliographic notes

The general framework of consistent ranked set sampling and an unified treatment on the estimation of means and smooth-function-of-means were given in Bai and Chen [6]. The minimum variance non-negative unbiased estimate of variance was developed by Perron and Sinha [132]. A similar estimate of variance was considered by MacEachern et al. [94]. The asymptotic pivot based on RSS for the tests and confidence intervals of a population mean was dealt with by Chen and Sinha [43]. The estimation of quantiles was treated in Chen [35]. Density estimation was considered in Chen [34]. The M-estimates were studied in full detail by Zhao and Chen [175]. The early research on ranked set sampling was concentrated on the non-parametric setting. Besides the seminal paper of McIntyre [96], earlier works include Halls and Dell [58], Takahasi and Wakimoto [167], Takahasi [163], [164], Dell and Clutter [50], Stokes [156],[157], [159], [158] etc.. Other aspects of ranked set sampling in the non-parametric setting were explored in the literature as well. Stokes and Sager [162] gave a characterization of ranked set sample and considered the estimation of distribution function. Kvam and Samaniego [83] considered the inadmissibility of empirical averages as estimators in ranked set sampling. Kvam and Samaniego

[85] and Huang [63] considered the nonparametric maximum likelihood estimation based on ranked set samples. Samawi and Muttlak [143] considered estimation of ratio using ranked set sample. Patil et al. [124] [129] dealt with ranked set sampling for finite populations. Presnell and Bohn [134] tackled U-statistics in ranked set sampling.