THE REDUCTION OF BIAS IN PARAMETRIC ESTIMATION

by

William R. Schucany

Technical Report No. 93
Department of Statistics ONR Contract

April 17, 1970

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

Schucany, William Roger

B.A., University of Texas,
Austin, 1963
M.A., University of Texas,
Austin, 1965

The Reduction of Bias in Parametric Estimation

Adviser: Professor Donald B. Owen

Doctor of Philosophy degree conferred May 24, 1970

Dissertation completed April 17, 1970

A general class of transformations of estimators is introduced which induces a reduction in bias if any exists. The concept is related to that of the sequence to sequence transformations which are employed for convergence improvement in deterministic cases such as the evaluation of infinite series and improper integrals. The procedure introduced by Quenouille (1949), (1956) and later termed the "jackknife" by Tukey (1958) is seen to be a special case of these transformations. The general principles of the method provide insight into the applications where the ordinary jackknife is not trustworthy.

To illustrate the method and demonstrate its potential usefulness several examples are considered. For ratio estimation under a particular model a new unbiased estimator is produced which exhibits a favorable mean square error relative to existing adjusted estimators.

The existing notion of reapplication of such a procedure is shown to lack the property for which it was designed. Proper reapplication is proposed so as to conform to general principles. A higher order transformation is defined which provides an interesting algorithm for the corrected procedure. Possible extensions to nonlinear transformations are also mentioned.

ACKNOWLEDGEMENTS

I am indebted to Dr. D. B. Owen for his guidance and encouragement for my research. His help as my adviser and partial financial support through the THEMIS Contract are greatly appreciated. I also wish to thank each of the faculty members for the invaluable instruction which I have received. I am also extremely grateful to Dr. P. D. Minton for his encouragement and generous financial support.

I must also express my sincere gratitude to Dr. H. L. Gray for years of instruction, direction and research assistance. Many of the concepts which appear in this dissertation were developed in close association with him. My further thanks go to Mrs. Jimmie Roberts and Mrs. Judy Ham for their contributions to the preparation of this manuscript. LTV, Incorporated is also gratefully acknowledged for their generous Doctoral Fellowship program.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

One basic problem in mathematical statistics is, given a series of observations, $x_1, x_2, \ldots, x_n$, to find a function of these, $t_n = t_n(x_1, x_2, \ldots, x_n)$, which will provide an estimate of an unknown parameter θ. Properties of such estimators such as efficiency, sufficiency, consistency and unbiasedness are desirable or important in varying degrees depending upon the application. In many instances the estimators which result from common procedures, such as the method of maximum likelihood for example, are biased. Whenever this bias is small relative to an associated near minimal variance this deficiency is frequently acceptable. This situation exists since the mean square error of an estimator is a popular goodness criterion in many cases. However, if one can reduce or even eliminate the bias without incurring an appreciable increase in the variance, so as to leave the mean square error unchanged or even reduced, then normally there would be little basis for faulting the procedure.

The purpose of this dissertation is to introduce a general class of transformations of estimators which induces a reduction in bias. The concept is related to that of the sequence to sequence transformations which are employed for convergence improvement in deterministic cases such as the evaluation of infinite series and improper integrals [see Lubkin (1952) and Gray and Atchison (1968)]. The procedure introduced by Quenouille (1949), (1956) and later termed the "jackknife" by Tukey (1958) will be shown to be

a special case of these transformations. This can shed some light upon the problems which Miller (1964) exhibits as evidence that there are applications for which the ordinary jackknife is not trustworthy. A most recent use of the jackknife reported in the literature is for U-statistics with a specific application to variance component models [see Arvesen (1969)].

To illustrate the present method and demonstrate its potential usefulness several examples are considered. The problem of ratio estimation in sampling theory provides a fruitful area of application. One of the models which is usually adopted admits a complete correction for the bias in the ratio estimator with no increase in the mean square error over the biased classical estimator or its jackknifed counterpart, which is also biased. The estimation of a truncation point is examined and seen to lend itself to these transformations.

The notion of reapplication of the ordinary jackknife was introduced by Quenouille (1956) and reported by Kendall and Stuart (1961) and several others. The scheme which was given actually falls short of what must have been intended. This difficulty is discussed in the more general setting of the transformation introduced here. In this context a proper procedure for reapplication is proposed which has the desirable characteristics which were previously absent. An algorithm for the proper successive reapplication is given in a single determinantal expression for a large class of estimators. This introduces what shall be called a higher order transformation of a biased estimator.

CHAPTER II

A GENERAL TRANSFORMATION FOR BIAS REDUCTION

2.1 Definition and Properties

Let θ be an unknown parameter, and let $X_1, X_2, \ldots, X_n$ be n independent, identically distributed observations from the cumulative distribution function (cdf) F_θ. Suppose that the two functions t_1 and t_2 are defined over the n observations and are to be considered as two different estimators of the parameter θ. Further suppose that each of these estimators is biased such that

$$E[t_k(X_1, X_2, \ldots, X_n)] - \theta = b_k(n,\theta) \neq 0, \; k = 1,2. \tag{2.1}$$

The two estimators may be combined to produce a third estimator $\hat{\theta}$ which we may now define. As is customary we shall in some cases use only the symbol for a function to denote the value of that estimator evaluated at the observed points.

Definition 2.1

Let

$$R = \frac{b_1(n,\theta)}{b_2(n,\theta)}, \tag{2.2}$$

where the b_k are as given in equation (2.1). Then whenever $R \neq 1$, i.e. when the biases in the two estimators are unequal, define

$$\hat{\theta} = \frac{t_1 - Rt_2}{1-R} \quad . \tag{2.3}$$

An immediate result is the following:

Theorem 2.1

When R is known the quantity $\hat{\theta}$ given by (2.3) is an unbiased estimator for θ.

Proof:

Let $\underline{X}_n$ denote the n random variables, $X_1, X_2, \ldots, X_n$, and then

$$E[\hat{\theta}(\underline{X})_n] = \frac{1}{1-R} \; E[t_1(\underline{X}_n)] - \frac{R}{1-R} \; E[t_2(\underline{X}_n)]$$

$$= \theta + \frac{b_1(n,\theta) - R\, b_2(n,\theta)}{1-R}$$

$$= \theta.$$

The variance of this new estimator depends jointly upon the value of R and the covariance between the two estimators t_1 and t_2, as well as their separate variances. Since

$$\hat{\theta} - \theta = \frac{t_1 - Rt_2}{1-R} - \theta = \frac{(t_1 - \theta) - R(t_2 - \theta)}{1-R} \quad ,$$

we may write

$$\text{Var}\,[\hat{\theta}] = E\left[\frac{(t_1 - \theta) - R(t_2 - \theta)}{1-R}\right]^2$$

$$= \frac{1}{(1-R)^2} \left\{ E[(t_1-\theta)^2] + R^2 E[(t_2-\theta)^2] \right.$$

$$\left. - 2RE[(t_1-\theta)\;(t_2-\theta)\;] \right\} \quad ,$$

or alternatively when R is in fact the ratio of the biases,

$$\mathrm{Var}\,[\hat{\theta}] = \frac{1}{(1-R)^2}\left\{\mathrm{Var}\,(t_1) + R^2\mathrm{Var}(t_2) - 2R\mathrm{Cov}(t_1,t_2)\right\} .$$

It is desirable to consider the minimization of this quantity over all possible choices of the estimator t_2. However, since for a particular choice of t_2, its bias, variance and covariance with t_1 each affect the quantity, there seems to be little to be gained by a classical optimization procedure. Within the class of estimators for which b_2 has a set value and such that R is positive, it is clear that the estimator should have a high positive correlation with t_1. It is also apparent that if the two estimators are biased in opposite directions then R is negative and hence ideally the estimators should be negatively correlated as well. The covariance of the two estimators is an important consideration and principles underlying the concept of antithetic variables [see Hammersly and Mauldon (1956)] may prove useful in this setting. This however will not be considered in any depth in this paper.

Since θ is an unknown parameter, the values of $b_1(n,\theta)$, $b_2(n,\theta)$ and consequently of R will in many cases not be known to the statistician. There are many interesting situations where this is not true, at least for R; but when it is, the possibility exists that R be estimated with some fruitful consequences.

Before discussing these possibilities let us note the cases in which R need not be estimated. Suppose the functions $b_k(n,\theta)$ are separable functions of n and θ, i.e.

$$E[t_k(\underline{X}_n)] - \theta = f_k(n)b_k'(\theta) , \quad k = 1,2. \tag{2.4}$$

Next suppose that the estimator t_2 is derived in some fashion from the estimator t_1 such that

$$b_1'(\theta) = b_2'(\theta) = b(\theta). \tag{2.5}$$

This will be the case when for example t_2 is of the same functional form as t_1 but is defined over a proper subset of the observations for which t_1 is defined. If we omit the ith value from the sample and take $t_2 = t_1(\underline{X}_{n-1})$, where the definition of t_1 is adjusted to accomodate the subset of n-1 values, then $E[t_2] - \theta = f_2(n)b_1'(\theta)$. Thus when (2.4) and (2.5) hold, equation (2.2) simplifies to

$$R = \frac{f_1(n)}{f_2(n)} \quad , \tag{2.6}$$

a quantity which no longer requires estimation.

An illustration of the method at this point will serve to bring out several points concerning the procedure. If the first estimator, t_1, is a function of the minimal set of sufficient statistics, then the question arises as to whether the second estimator t_2 must of necessity introduce superfluous variability. In some instances this is the case and the decrease in bias is obtained only at the expense of a corresponding increase in variance. However in other cases the transformation will merely produce an altered function of the sufficient statistics which has less bias than t_1. Suppose that X_i (i=1, ..., n) are independent identically distributed (i.i.d.) as $N(\mu,\sigma^2)$ and that it is proposed that we estimate σ^2 by taking

$$t_1 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2 \quad ,$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad .$$

Now

$$E[t_1] = \frac{n-1}{n} \sigma^2 = \sigma^2 - \frac{1}{n} (\sigma^2)$$

and if t_2 is formed by successively deleting each observation from the sample, one at a time, and averaging the n results we would expect

$$E[t_2] = \sigma^2 - \frac{1}{n-1} (\sigma^2) \quad .$$

Hence the indicated value of R is (n-1)/n and

$$\hat{\theta} = \frac{t_1 - \frac{n-1}{n} t_2}{1 - \frac{n-1}{n}}$$

$$= nt_1 - (n-1)t_2 \quad .$$

So, upon deleting the ith observation the associated estimator for the particular subsample is

$$ {}_i t_{n-1} = \frac{1}{n-1} \sum_{j \neq i}^{n} x_j^2 - \left(\frac{n\bar{x} - x_i}{n-1} \right)^2 $$

$$ = \frac{1}{n-1} \sum_{j \neq i}^{n} x_j^2 - \frac{1}{(n-1)^2} (n^2\bar{x}^2 - 2nx_i\bar{x} + x_i^2) \quad , $$

and averaging over i we obtain

$$ t_2 = \frac{1}{n} \sum_{i=1}^{n} {}_i t_{n-1} $$

$$ = \frac{(n-1)^2 - 1}{n(n-1)^2} \sum_{j=1}^{n} x_j^2 + \frac{2n-n^2}{(n-1)^2} \bar{x}^2 $$

$$ = \frac{n^2-2n}{(n-1)^2} t_1 \quad . $$

Consequently

$$ E[t_2] = \frac{n^2-2n}{(n-1)^2} E[t_1] $$

$$ = \frac{n-2}{n-1} \sigma^2 \quad , $$

and indeed

$$ E[t_2] = \sigma^2 - \frac{1}{n-1} (\sigma^2) \quad . $$

We now combine the two functions of the sufficient statistics to obtain a third, namely

$$\hat{\theta} = nt_1 - (n-1)t_2$$

$$= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 - \frac{n^2-2n}{(n-1)n} \sum_{i=1}^{n} x_i^2 + \frac{n^2-2n}{n-1} \bar{x}^2$$

$$= \frac{n^2-n-n^2+2n}{(n-1)n} \sum_{i=1}^{n} x_i^2 - \frac{n^2-n-n^2+2n}{n-1} \bar{x}^2$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (x_i-\bar{x})^2 ,$$

which may be recognized as the uniformly minimum variance unbiased estimator of the variance of the normal distribution. Furthermore, it may be recalled that, with no more than the assumptions of uncorrelated X_i and each X_i having the same expectation and variance, the estimator $\hat{\theta}$ above is a distribution free unbiased estimator of the variance of X.

2.2 A Related Nonlinear Transformation

As defined in (2.3), $\hat{\theta}$ is a linear combination of two functions of the sample values. If an estimate of R is employed, say $\hat{R}$, then we obtain a different estimator

$$\tilde{\theta} = \frac{t_1 - \hat{R}t_2}{1-\hat{R}} \quad . \tag{2.7}$$

This has a nonlinear flavor which suggests an interesting connection with the nonlinear sequence-to-sequence transformations often used to increase the rate of convergence of a sequence to its limiting value. To facilitate this comparison the rationale behind these sequence-to-sequence transformations is included here. For a more extensive discussion see S. Lubkin

(1952) and D. Shanks (1955), two of the early contributors to the development of these transformations or P. Wynn (1966), who has written extensively concerning generalizations and algorithms for convergence acceleration.

Let A_m and B_n be any two sequences of partial sums which have the same limit, i.e.,

$$A_m = \sum_{k=1}^{m} a_k \quad , \quad B_n = \sum_{k=1}^{n} b_k \quad ,$$

and $\lim_{n\to\infty} B_n = \lim_{m\to\infty} A_m = S \neq \pm\infty$.

Now consider the possibility of constructing a third sequence

$$Z_n = \sum_{k=1}^{n} z_k$$

which also converges to S but does so more rapidly. Intuitively, we might expect Z_n to converge more rapidly than A_m or B_n if

$$\lim_{k\to\infty} \frac{z_k}{a_k} = \lim_{k\to\infty} \frac{z_k}{b_k} = 0. \tag{2.8}$$

That is, on a term by term comparison, the contributions to one sum are diminishing more rapidly than the contributions to either of the other sums. Thus, suppose we let

$$z_k = \frac{a_k - Rb_k}{1-R} \tag{2.9}$$

and consider R as a parameter to be selected to achieve the desired more rapid convergence. If

$$R = \lim_{k\to\infty} \frac{a_k}{b_k}$$

and is not equal to zero or one, then equation (2.8) is satisfied.

It was suggested in the previous section that the second estimator t_2 might be derived from the first by using a subset of the values from t_1. An analagous technique may be used in the present situation to remove the difficulty involved in producing a second sequence which is convergent to the same limit. Let $m = n+j$ and $a_k = b_k$, so that we are now using the same series twice by shifting the index. Then,

$$A_m = A_{n+j} = \sum_{k=1}^{n+j} b_k$$

$$= \sum_{k=1}^{j} b_k + \sum_{k=1}^{n} b_{k+j} \quad ,$$

and we may replace (2.9) with

(2.10)
$$z_k = \frac{b_{k+j} - Rb_k}{1-R} \quad .$$

In the discussion leading to the estimator in (2.7) it was mentioned that R might require estimation since we may be unable to produce the ratio of the biases. Similarly if there are difficulties in producing the limit R we may approximate this limit of the ratio of terms by the ratio itself. Hence let

$$R_n = \frac{b_{n+j}}{b_n} \quad ,$$

then by summing over k in equation (2.10) and then substituting R_n for R we have

(2.11)
$$Z_n = \frac{B_{n+j} - R_n B_n}{1 - R_n}$$

which is the desired "third sequence". When j=1 we have the expression of a sequence-to-sequence transformation which is usually referred to as the e_1 transform or Aitken's δ^2 process.

$$e_1(B_n) = \frac{B_{n+1} - \frac{b_{n+1}}{b_n} B_n}{1 - \frac{b_{n+1}}{b_n}}$$

$$= \frac{B_{n+1} \left(\frac{B_{n+1} - B_n}{B_n - B_{n-1}} \right) B_n}{1 - \left(\frac{B_{n+1} - B_n}{B_n - B_{n-1}} \right)}$$

$$= \frac{B_{n+1} B_{n-1} - B_n^2}{B_{n+1} + B_{n-1} - 2B_n} .$$

The nonlinear character of the transformation is quite apparent in this final form and the similarity of (2.11) and (2.7) was previously mentioned.

The motivation for defining an estimator $\tilde{\theta}$ as in (2.7) was that R, which is given by

$$R = \frac{\theta - E[t_1(\underline{X}_n)]}{\theta - E[t_2(\underline{X}_n)]} ,$$

may not be obtainable. One might therefore select $\hat{R}$ so that

$$E[\hat{R}] = R,$$

but this is not necessary. The only essential requirement is that

$$E[\tilde{\theta}] = \theta .$$

If one employs a development which parallels that used for deterministic sequences then instead of two estimators, t_1 and t_2, a sequence of estimators, $\{t(\underline{X}_n)\}$, whose index indicates increasing sample size, is required. With this sequence at hand one could take

$$\hat{R} = \frac{t(\underline{X}_n) - t(\underline{X}_{n-1})}{t(\underline{X}_{n-1}) - t(\underline{X}_{n-2})} .$$

This approach leads, as before, to the e_1 transformation of the sequence and (suppressing the arguments of the functions)

$$\tilde{\theta}_n = \frac{t_n t_{n-2} - t_{n-1}^2}{t_n + t_{n-2} - 2t_{n-1}} . \tag{2.12}$$

It has been suggested by Gray and Schucany (1968) that this transformation might be profitably applied to sequences of random variables. Partial motivation for this conjecture is the following.

<u>Theorem 2.2</u>

If $\{t_n\}_{n=m}^{\infty}$ is a sequence of random variables which converges in probability to θ_t with respect to a probability measure P, and if the functions $\{\tilde{\theta}_n\}_{n=m+1}^{\infty}$ as defined in (2.12) are measurable and finite almost everywhere

and the sequence converges in probability to θ_e, then $\theta_e=\theta_t$ a.e. (P).

The proof of this theorem will not be included here since no further use is made of the result. The primary drawback to this approach is that a sequence of estimators, which is derived from increasing the size of a set of independent variates, is necessarily the result of an arbitrary selection of the order in which the observations are considered to be included in the sample. This violates any notions of invariance or symmetry which are quite naturally associated with most estimators. However, there is a way to retain symmetry.

In practice one has a sample of size n and an estimator t_n defined over the n observation values. The problem is to construct a three member sequence associated with this estimate, in order that a new estimator be produced by means of the transformation (2.12). The ith observation can be deleted and the estimate, based on the remaining n-1 observations, denoted ${}_it_{n-1}$. Next the jth observation can be deleted from the remaining n-1 values (clearly $j\neq i$) and the estimate, based on the n-2 observations which remain, denoted ${}_{ij}t_{n-2}$. There are however n(n-1) ways in which this can be done and in many cases this gives rise to n(n-1) distinct three term sequences.

One possible procedure, which will restore symmetry in the resulting estimator, is to average the n(n-1) values of $\tilde{\theta}_{ij}$ which result from the transformations,

$$\tilde{\theta}_{ij} = \frac{{}_{ij}t_{n-2}\, t_n - {}_it_{n-1}^2}{{}_{ij}t_{n-2} + t_n - 2{}_it_{n-1}} ,$$

to obtain

$$\bar{\theta} = \frac{1}{n(n-1)} \sum_{\substack{i \\ i \neq j}} \sum_{j} \tilde{\theta}_{ij} .$$

Since each of the $\tilde{\theta}_{ij}$ is a consistent estimate of θ, the estimator $\bar{\theta}$ is also consistent. There is, however, little reason to suspect that these manipulations have been performed without incurring an increase in the variance of the transformed estimator. Furthermore there is no assurance that any significant reduction of bias results from this procedure. For practically all estimators the analytical determination of the distribution of the nonlinear combination of the correlated members of the sequence is a formidable task.

Nevertheless, these concepts are mentioned because of the proven success of the nonlinear sequence-to-sequence transformation in the deterministic case. Also the procedures which were suggested to restore invariance to the reuse of sample information are seen to be quite naturally applicable for the case of linear combinations of these correlated estimates.

Before leaving the realm of non linear transformations, another possible procedure should be noted. If the n values of ${}_it_{n-1}$ are averaged to obtain,

$$\bar{t}_{n-1} = \frac{1}{n} \sum_{i=1}^{n} {}_it_{n-1} , \tag{2.13}$$

then $\bar{t}_{n-1}$ and t_n are each estimators of θ and could be considered as adjacent members of a sequence of estimators. A third member produced in a parallel fashion would be

$$\bar{t}_{n-2} = \frac{1}{n(n-1)} \sum_{\substack{i \\ i \neq j}} \sum_{j} {}_{ij}t_{n-2} \quad .$$

At this point the potential of the e_1 transformation may again be entertained and still another estimate defined by

$$\theta^* = \frac{\bar{t}_{n-2}t_n - \bar{t}^2_{n-1}}{\bar{t}_{n-2} + t_n - 2\bar{t}_{n-1}} \quad .$$

Again the properties of the new estimator are difficult to establish but the principle employed to obtain $\bar{t}_{n-1}$ and $\bar{t}_{n-2}$ are useful. This technique was introduced in connection with the jackknife [see Quenouille (1956)] and shows equal promise for the more general transformation defined in equation (2.3).

2.3 Relationship to the Jackknife

Suppose that

$$t_1 = t_n(X_1, X_2, \ldots, X_n),$$

and, as mentioned in Section 2.1, that the bias is separable, viz.,

$$E[t_1] - \theta = f(n)\, b(\theta) \quad .$$

Further suppose that the second estimator is formed according to rules discussed in Section 2.2 which led to equation (2.13). Consequently

$$t_2 = \frac{1}{n} \sum_{i=1}^{n} {}_{i}t_{n-1}\,(\underline{X}_{n-1})$$

and

$$E[t_2] - \theta = f(n-1)\ b(\theta) \quad .$$

If it is assumed that the bias in t_1 is inversely proportional to the sample size then

$$R = \frac{f(n)}{f(n-1)} = \frac{n-1}{n} \quad ,$$

and $\hat{\theta}$ becomes

$$\hat{\theta} = nt_1 - (n-1)t_2 \quad ,$$

which is the form of the jackknife introduced by Quenouille. This technique will eliminate the 1/n term from a bias and the justification for the procedure which was given by Quenouille (1956) is that the bias in t_1 is often expressible as a power series in (1/n). Indeed if

$$E[t_1] = \theta + \frac{a_1}{n} + \frac{a_2}{n^2} + \ldots \quad ,$$

then it can be easily shown that

$$E[\hat{\theta}] = \theta - \frac{a_2}{n(n-1)} + \ldots \quad .$$

If some intermediate quantities $\hat{\theta}_i$, $i=1,\ldots,n$ (these are called "pseudo-values" by Tukey for the special case $R = (n-1)/n$) are defined by

$$\hat{\theta}_i = \frac{1}{1-R}\, t_1 - \frac{R}{1-R}\, {}_it_{n-1}(\underline{X}_{n-1}) \quad , \tag{2.14}$$

then

$$\hat{\theta} = n^{-1} \sum_{i=1}^{n} \hat{\theta}_i \quad .$$

Moreover, Tukey has proposed that in many instances these "pseudo-values" are approximately independently, identically distributed. If this proposal holds then

$$\frac{1}{n(n-1)} \sum_{i=1}^{n} (\hat{\theta}_i - \hat{\theta})^2$$

should be an (approximate) estimate of the variance of $\hat{\theta}$ and

$$\sqrt{n}(\hat{\theta}-\theta) \left[\frac{1}{(n-1)} \sum_{i=1}^{n} (\hat{\theta}_i - \hat{\theta})^2 \right]^{-\frac{1}{2}} \tag{2.15}$$

should be approximately distributed as a Student's t variate with n-1 degrees of freedom.

This technique has been used to good advantage by Miller (1968) to construct approximate tests for scale parameters in the two sample problem. Miller publishes the results of a Monte Carlo study giving observed power functions for the jackknife test and the classical F test among others for various distributions. In a more recent work Shorack (1969) compares several more robust alternatives to the F test on the basis of Pitman asymptotic relative efficiency and Monte Carlo studies of power functions. Shorack's approximate permutation test and the jackknife procedure are found to be most satisfactory.

Criticisms of the jackknife procedure which have been voiced for particular applications may be found to be due to the restrictive use of R=(n-1)/n when another choice is more appropriate and is indicated by some inspection. The difficulties for estimating truncation points will be taken up in the next chapter. The more general definition of the pseudo-values should improve the procedure in most specific situations. It is

clear that the approximation for Var $[\hat{\theta}]$ will be improved according to the degree of improvement of the bias afforded by the proper choice of R.

Durbin (1959) exhibited a class of problems in ratio estimation where the jackknife appears to be appropriate. Both the bias and the mean square error were shown to decrease under a reasonable model. The use of the more general procedure can produce still further improvements under various models.

A point that is brought out by Mantel (1967) is that if "... an estimator had a more serious bias, say one inversely proportional to the root of sample size, it is not eliminated ... but it is approximately halved." Indeed if

$$E[t_1] - \theta = \frac{b(\theta)}{\sqrt{n}}$$

and

$$E[t_2] - \theta = \frac{b(\theta)}{\sqrt{n-1}}$$

and the ordinary jackknife is employed to obtain

$$\hat{\theta} = nt_1 - (n-1)t_2$$

then

$$E[\hat{\theta}] = \theta + (\sqrt{n} - \sqrt{n-1})b(\theta) \quad .$$

Therefore

$$\frac{\text{Bias } (\hat{\theta})}{\text{Bias } (t_1)} = \sqrt{n^2} - \sqrt{n(n-1)}$$

$$= n - n\left(1 - \frac{1}{n}\right)^{\frac{1}{2}}$$

$$= n - n(1 - \frac{1}{2n} - \frac{1}{8n^2} - \frac{1}{16n^3} - \cdots$$

$$= \frac{1}{2} + \frac{1}{8n} + \frac{1}{16n^2} + \cdots \quad .$$

However had the proper value of R,

namely

$$R = \sqrt{\frac{n-1}{n}}$$

been used in the more general transformation then $\hat{\theta}$ would be completely unbiased.

One further aspect in the evolution of the ordinary jackknife which has an interesting relationship with the present general procedure is Tukey's graphical method. The assumption that the bias in t_1 and t_2 as estimators of θ is inversely proportional to their basic sample sizes, or nearly so, implies that the points $(0,\theta)$, $(\frac{1}{n}, E[t_1])$ and $(1/(n-1), E[t_2])$ lie on or close to a straight line. The two point formula for a straight line yields the following expression for the intercept of the ordinate axis, which would be the unbiased estimate of θ:

$$t_1 - \frac{1}{n}\left(\frac{t_2 - t_1}{\frac{1}{n-1} - \frac{1}{n}}\right)$$

$$= t_1 - (n-1)(t_2 - t_1)$$

$$= nt_1 - (n-1)t_2 \quad .$$

The method described above suggests the possibility of fitting polynomials in 1/n to the averages of estimates computed from sub-samples of various sizes. This possibility, successively omitting one, two, three, ... sampling units, has been explored by Burdick (1961). This led Jones (1963) to speculate that:

> "In some applications of the jackknife method, as when the population parameter to be estimated is the reciprocal of the mean, it would seem that graphing the polynomial might suggest expansion in powers of some function of n other than 1/n in order to obtain adjusted estimates with minimum variance or minimum mean square error in the future applications to situations of the same kind."

This borders upon the essence of the present technique and approximations thereto.

CHAPTER III

APPLICATIONS TO BIASED ESTIMATORS

3.1 Ratio Estimators

In sampling theory there is a greater emphasis placed upon the use of auxiliary information than there is in most other branches of statistics. One method of interest is the use of auxiliary information to improve the precision of estimates through consideration of a population ratio, $\rho = Y/X$. The use of lower case letters for random variables and capitals for population values is conventional in sampling theory. Frequently a situation exists where the ratio of a variable y to another variable x is believed to be less variable than the y variable alone. Suppose for example one were interested in the value of Y (population total). Rather than estimate this total directly from the sample it may be better to estimate ρ from the sample and then multiply it by the known total of x to estimate the total for y. This is called ratio estimation. An instructive hypothetical example is given by Raj (1968) as follows:

Suppose it is desired to estimate the total agricultural area, Y, in a region containing N communes. There are very big communes and very small communes which causes y to vary tremendously over the region. But the ratio, y/x, of agricultural area to the population of the commune (per capita area) would be less variable. If the population figures are known for each commune it would be preferable to estimate the ratio of agricultural area to the census population from the sample of communes and multiply this figure by the known census population total of all the communes in the

region. If a random sample of n communes gives $\sum_{i=1}^{n} y_i$ and $\sum_{i=1}^{n} x_i$ as the totals for y and x, respectively, the total of Y for the region is estimated by

$$\hat{Y} = X \sum_{i=1}^{n} y_i / \sum_{i=1}^{n} x_i \quad ,$$

where X is the known total for x for the region. If the census data on x is not employed, then

$$\hat{Y} = N \sum_{i=1}^{n} y_i \quad /n \ .$$

The above ratio estimator is biased, though in many situations only negligibly so. On the other hand the bias may be considerable in surveys with many strata and small or moderate samples within strata if it is deemed appropriate to use separate ratio estimators. When it is considered to be important that proper confidence statements be made, it is often necessary that the bias of an estimator be negligibly small. Consequently, in recent years, considerable attention has been given to the development of unbiased or approximately unbiased ratio estimators.

The following theorem shows that

$$r = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{y}}{\bar{x}}$$

is usually a biased estimator of ρ. (Goodman and Hartley, 1958).

Theorem 3.1

In simple random sampling, the bias of the ratio estimator, r, is given by

$$E[r]-\rho = -[E(\bar{x})]^{-1}\mathrm{Cov}(r,\bar{x})$$

Note that the bias associated with the estimate

$$Y = X(\bar{y}/\bar{x})$$

for the total of y is

$$\mathrm{Bias}(\hat{Y}) = X\mathrm{Bias}(r).$$

Also note that practically speaking r is never unbiased since this occurs only if r and $\bar{x}$ are uncorrelated, a situation in which the ratio estimator is not employed.

The decision to use a ratio estimator in hopes of improving the precision is ordinarily based on consideration of the coefficients of variation for the variables x and y and upon the correlation believed to be present between the two. In general the ratio estimator is useful if the characters x and y have a correlation coefficient which exceeds 1/2. The variable y must be nearly proportional to x or other schemes such as difference estimators or regression estimators may be dictated.

After the decision to use a ratio estimator has been made, the evaluations of the various modifications to the classical estimator which exist will, of necessity, depend upon the assumed model for the relationship between y and x and for the family to which the distribution of x belongs. Several authors have examined under two general models the bias and approximations to the mean square error (MSE) of various estimators.

Durbin (1959) examined ratio estimators of the form $r=y/x$ where the regression of y on x is linear and where x is normally distributed. He considers an application of Quenouille's method which splits the sample into two halves to yield

$$r_1 = \frac{y_1}{x_1} \text{ and } r_2 = \frac{y_2}{x_2} \quad ,$$

where

$$y = \frac{1}{2}(y_1+y_2)$$

$$x = \frac{1}{2}(x_1+x_2).$$

Then the new estimate, $\hat{R}_2$, of $\rho = E(y)/E(x)$ is

$$\hat{R}_2 = 2r - \frac{1}{2}(r_1+r_2).$$

Secondly Durbin investigates r and $\hat{R}_2$ assuming x has a gamma distribution. These two models form the basis for most of the work done in the area of bias reduction for ratio-type estimation.

Suppose

$$y = a+bx+u,$$

where the Var $(u)=\delta$, a constant of $O(n^{-1})$, and $E[u|x]=0$. Hence,

$$\rho = \frac{E(y)}{E(x)} = b+a/E(x)$$

and

$$E(r) = aE(x^{-1})+b. \tag{3.1}$$

3.1.1 Normal Auxiliary

Suppose that x is a normal variable with variance h, which is $O(n^{-1})$, and the units of measurement are chosen so that $E(x)=1$, and then let $x=1-\xi$; for sufficiently large n we have

$$E(x^{-1}) = E(1+\xi+\xi^2+\xi^3+\ldots) \quad .$$

Taking the first four non-vanishing terms we find

$$E(x^{-1}) = 1+h+3h^2+15h^3+O(n^{-4}). \tag{3.2}$$

Similarly

$$\begin{aligned} E(x^{-2}) &= E(1+2\xi+3\xi^2+4\xi^3+\ldots) \\ &= 1+3h+15h^2+105h^3+O(n^{-4}). \end{aligned} \tag{3.3}$$

At this point we see that if (3.2) is substituted in (3.1) the bias in r may be determined as

$$\begin{aligned} E(r)-\rho &= aE(x^{-1})+b-(a+b) \\ &= a(h+3h^2+15h^3), \end{aligned}$$

neglecting terms of $O(n^{-4})$. Further, since $\mathrm{Var}\,(x_1)=\mathrm{Var}(x_2)=2h$, we may replace h by 2h in (3.2) and (3.3) and obtain

$$E(x_i^{-1}) = 1+2h+12h^2+120h^3$$

and

$$E(x_i^{-2}) = 1+6h+60h^2+840h^3 \ , \qquad i=1,2.$$

Thus if $t_2=(r_1+r_2)/2$ then its bias is given by

$$E(t_2)-\rho = a(2h+12h^2+120h^3),$$

Hence if the principles introduced in previous chapter are employed to obtain a proper combination of

$$t_1 = r$$

and

$$t_2 = \frac{1}{2}(r_1+r_2) ,$$

of the form

$$\frac{t_1-Rt_2}{1-R} ,$$

then to eliminate the bias to terms of order $O(n^{-4})$ the appropriate choice of the combining parameter is

$$R = \frac{1 + 3h + 15h^2}{2(1+6h+60h^2)} . \tag{3.4}$$

Selecting $R=\frac{1}{2}$ leads to the estimator $\hat{R}_2$ given by Durbin. For small h the above expression for R is quite near 1/2. The importance of (3.4) becomes clear when it is recalled that the auxiliary character x is being used because its distribution is known. Therefore h is not an unknown and this improved value for R may be used to yield

$$\hat{R}_3 = \frac{2(1 + 6h + 60h^2)r - (1 + 3h + 15h^2)\ (r_1 + r_2)}{1 + 9h + 105h^2}$$

Using (3.3) Durbin has shown that, not only is the bias of $\hat{R}_2$ smaller than that of r, but $Var(\hat{R}_2) < Var(r)$. The estimator $\hat{R}_3$ combines the same

two quantities, t_1 and t_2 , in the same fashion with a further improvement in the bias. J. N. K. Rao (1956) has shown that the bias and variance of the jackknifed classical ratio estimator are both decreasing functions of the number (g) of subsets into which the sample is split. In other words, let the sample of pairs (y_i, x_i) $(i=1,\ldots,n)$ be split at random into g groups each of size m, then we get the estimator

$$\hat{R}_j = \bar{y}_i'/\bar{x}_j'$$

from the sample after omitting the jth group, where

$$\bar{y}_j' = (n\bar{y}-m\bar{y}_j)/(n-m)$$

$$\bar{x}_j' = (n\bar{x}-m\bar{x}_j)/(n-m) \quad ,$$

and $\bar{y}_j$ and $\bar{x}_j$ are the sample means for the jth group. Then Quenouille's estimator is

$$\hat{R}_Q = gr - \frac{g-1}{g} \sum_1^g \hat{R}_j \tag{3.5}$$

and Bias $(\hat{R}_Q)$ and Var$(\hat{R}_Q)$ are both decreasing functions of g. For g=2 we have $\hat{R}_2$ given and studied by Durbin. Consequently, the indicated optimal choice of t_2 is (corresponding to g=n)

$$t_2 = \frac{1}{n} \sum_{j=1}^{n} \hat{R}_j \quad .$$

However, the appropriate combining parameter should not be (n-1)/n, contrary to standard practice. Since the value of h is assumed to be known and h is $O(n^{-1})$, we shall consider the case in which h=c/n for a known constant c. This requires us to choose

$$R = \frac{a[h + 3h^2 + 15h^3]}{a\left[\frac{c}{n-1} + 3\frac{c^2}{(n-1)^2} + 15\frac{c^3}{(n-1)^3}\right]}$$

$$= \left(\frac{n-1}{n}\right)^3 \left[\frac{n^2 + 3cn + 15c^2}{n^2 + (3c-2)n + (15c^2-3c + 1)}\right] ,$$

as the proper parameter in the estimator

$$\hat{R}_4 = \left[\frac{t_1 - Rt_2}{1 - R}\right] .$$

Nevertheless, because the purpose of this section is one of illustrating a procedure and its potential usefulness rather than introducing the optimal ratio estimator, we shall avoid the complications of $\hat{R}_4$ and give the variance of $\hat{R}_3$, which can be compared with the variances of r and $\hat{R}_2$. Durbin (1959) gives the following:

$$\begin{aligned}
\text{Bias } (r) &= a(h+3h^2+15h^3) = aB(r) \\
\text{Var } (r) &= a^2(h+8h^2+69h^3) + \delta(1+3h+15h^2+105h^3) \\
&= a^2S_1(r) + \delta\, S_2(r) \\
\text{Bias } (\hat{R}_2) &= a(6h^2+90h^3) = aB(\hat{R}_2) \\
\text{Var } (\hat{R}_2) &= a^2(h+4h^2+12h^3) + \delta(1+2h+8h^2+108h^3) \\
&= a^2S_1(\hat{R}_2) + \delta\, S_2(\hat{R}_2) \quad .
\end{aligned}$$

The estimator $\hat{R}_3$ is unbiased to $O(n^{-4})$. In order to derive the variance of $\hat{R}_3$ let

$$\hat{R}_3 = cr - d\left(\frac{r_1+r_2}{2}\right) ,$$

where $c = \frac{1}{1-R}$, $d = \frac{R}{1-R}$; $c-d=1$.

(Note that for $\hat{R}_2$; $R=\frac{1}{2}$, $c=2$, $d=1$.)

Using the linear model introduced previously and splitting the sample as before we have $u=\frac{1}{2}(u_1+u_2)$, $y_i=a+bx_i+u_i$ and $r_i=y_i/x_i$, $i=1,2$, and further that $E(u_i|x_i)=0$ and $E(u_i^2|x_i)=2\delta$.

Now we may write

$$\hat{R}_3 = cb+\frac{c}{x}(a+u)-db-\frac{d}{2x_1}(a+u_1) - \frac{d}{2x_2}(a+u_2)$$

$$= b + a\left\{\frac{c}{x} - \frac{d}{2}\left(\frac{1}{x_1}+\frac{1}{x_2}\right)\right\} + \frac{cu}{x} - \frac{d}{2}\left(\frac{u_1}{x_1}+\frac{u_2}{x_2}\right).$$

Hence

$$E(\hat{R}_3-b) = aE\left\{\frac{c}{x} - \frac{d}{2}\left(\frac{1}{x_1}+\frac{1}{x_2}\right)\right\}, \tag{3.6}$$

and when R is given by (3.4),

$$c = \frac{2(1+6h+60h^2)}{1+9h+105h^2}, \qquad d = \frac{1+3h+15h^2}{1+9h+105h^2}$$

and therefore

$$E(\hat{R}_3-b) = a+0(n^{-4}).$$

Next consider

$$E\left\{\frac{c}{x} - \frac{d}{2}\left(\frac{1}{x_1}+\frac{1}{x_2}\right)\right\}^2$$

$$= E\left\{\frac{c^2}{x^2} + \frac{d^2}{4}\left(\frac{1}{x_1}+\frac{1}{x_2}\right) + \left(\frac{d^2}{2} - cd\right)\left(\frac{1}{x_1x_2}\right)\right\}$$

$$= c^2(1+3h+15h^2+105h^3) + \frac{d^2}{2}(1+6h+60h^2+840h^3)$$

$$+ \left(\frac{d^2}{2} - 2cd\right)(1+2h+12h^2+120h^3)^2$$

and

$$(3.8) \qquad E_x E_{u|x} \left[\left\{ \frac{c(u_1+u_2)}{2x} - \frac{d}{2}\left(\frac{u_1}{x_1} + \frac{u_2}{x_2}\right)\right\}^2 \middle| x_1, x_2 \right]$$

$$= E_x 2\delta \left[\left(\frac{c}{2x} - \frac{d}{2x_1}\right)^2 + \left(\frac{c}{2x} - \frac{d}{2x_2}\right)^2 \right]$$

$$= \frac{\delta}{2} E\left[\frac{2c^2}{x^2} + d^2\left(\frac{1}{x_1^2} + \frac{1}{x_2^2}\right) - 4cd\left(\frac{1}{x_1 x_2}\right)\right]$$

$$= \delta \left\{ c^2(1+3h+15h^2+105h^3) + d^2(1+6h+60h^2+840h^3) - 2cd(1+2h+12h^2+120h^3) \right\} .$$

Hence substituting approximate expressions for c^2, d^2 and cd we obtain, after some algebra,

$$E(\hat{R}_3-b)^2 = a^2 \left[\frac{1+26h+435h^2+4098h^3}{1+18h+291h^2+1890h^3}\right] + \delta \left[\frac{1+28h+471h^2+4680h^3}{1+18h+291h^2+1890h^3}\right] .$$

Consequently

$$\mathrm{Var}(\hat{R}_3) = \mathrm{Var}(\hat{R}_3-b) = a^2 \left[\frac{1+8h+117h^2+2046h^3}{1+18h+291h^2+1890h^3}\right] + \delta \left[\frac{1+28h+471h^2+4680h^3}{1+18h+291h^2+1890h^3}\right] .$$

Since direct comparisons of the variance expressions are difficult the values of the coefficients, S, of a^2 and δ have been tabulated for several values of h. The coefficients of a in the expressions for the bias are also given here in Table I. No bias was calculated for $\hat{R}_3$ since all terms containing the fourth power and higher in h have been neglected in the original approximations and $\hat{R}_3$ is corrected for bias to this degree. Because of the approximation the entries in the lower half of the table are subject to considerable error. For instance the error in B(r) for h=0.50 is greater than 6.0. In spite of this, the large values of h have been included for two reasons. First of all these larger values are not unusual in ordinary practical applications. Evidence in support of this contention may be found in Rao (1969). He lists 16 natural populations and includes the coefficient of variation of the auxiliary variable x. Recall that h is the square of this coefficient. Six of 16 populations exhibit a value of h greater than 1.0 and only two have a value of h less than 0.2. Secondly, note from Table I that even at h=0.1 the bias in $\hat{R}_2$ is greater than that of r; a disturbing result since $\hat{R}_2$ was proposed from a bias reduction standpoint. The biases continue to exhibit this behavior to a greater extent for increasingly larger values of h.

The breakdown of the necessary approximation for the normal model for the realistically large coefficients of variation is one inducement to examine a different model. Furthermore, since practically all auxiliary random variables, x, which are used in real problems are positive the normal model is not realistic when h is near 1. The gamma distribution has received more attention in the recent literature; possibly due to some of these arguments.

TABLE I: VARIANCE COMPARISONS FOR NORMAL MODEL

h	r Bias B	r Variance S_1	r Variance S_2	$\hat{R}_2$ Bias B	$\hat{R}_2$ Variance S_1	$\hat{R}_2$ Variance S_2	$\hat{R}_3$ Bias B	$\hat{R}_3$ Variance S_1	$\hat{R}_3$ Variance S_2
.01	0.010	0.011	1.032	0.001	0.010	1.021	0.0	0.077	1.100
.05	0.058	0.079	1.201	0.016	0.061	1.133	0.0	0.331	1.454
.10	0.145	0.249	1.555	0.150	0.152	1.388	0.0	0.528	1.736
.15	0.268	0.563	2.142	0.439	0.280	1.844	0.0	0.646	1.900
.20	0.440	1.072	3.040	0.960	0.456	2.584	0.0	0.722	2.005
.25	0.672	1.828	4.328	1.781	0.687	3.687	0.0	0.776	2.078
.30	0.975	2.883	6.085	2.97	0.984	5.236	0.0	0.815	2.130
.40	1.84	6.096	11.320	6.72	1.808	9.992	0.0	0.870	2.203
.50	3.125	11.125	19.375	12.75	3.000	17.500	0.0	0.906	2.250
.75	8.765	34.359	55.984	41.34	8.062	52.562	0.0	0.958	2.318
1.00	19.0	78.0	124.0	96.0	17.0	119.0	0.0	0.987	2.355

*In each expression the terms of the fourth power in h and higher have been neglected.

3.1.2 Gamma Auxiliary

Let (y_i, x_i), $i=1, \ldots, n$, denote a simple random sample from a population assumed to be infinite and let $\bar{y}$ and $\bar{x}$ denote the sample means. Let $y_i = a + bx_i + u_i$ where $E(u_i|x)=0$ and $E(u_i^2|x)=n\delta$ where as before δ is a constant of $O(n^{-1})$. Let the variates x_i/n have the gamma distribution with parameter h so that $\bar{x}$ has the gamma distribution with parameter m=nh, i.e. the density of $\bar{x}$ is

$$\bar{x}^{m-1}\exp(-\bar{x})/\Gamma(m), \quad \bar{x}>0.$$

It follows that

$$E\left(\frac{1}{\bar{x}}\right) = \frac{1}{m-1} \quad \text{and } E\left(\frac{1}{\bar{x}^2}\right) = \frac{1}{(m-1)(m-2)} \quad .$$

Proceeding as in the previous section we assume that n is even and the sample is randomly split in half. Let $\bar{x}_j$ and $\bar{y}_j$ $(j=1,2)$ denote the sample means within the two groups. Hence $\frac{1}{2}\bar{x}_1$ and $\frac{1}{2}\bar{x}_2$ are independent gamma variables with parameters m/2, and

$$E\left(\frac{1}{\bar{x}_j}\right) = \frac{1}{m-2} \quad \text{and } E\left(\frac{1}{\bar{x}_j^2}\right) = \frac{1}{(m-2)(m-4)} \quad , \quad j=1,2 \quad .$$

Here we have $\rho = b + a/m$ and since

$$E\left(\frac{\bar{y}}{\bar{x}} - b\right) = \frac{a}{m-1} \quad ,$$

the bias in $r = \bar{y}/\bar{x}$ is

$$\text{Bias}(r) = \frac{a}{m-1} - \frac{a}{m} = \frac{a}{m(m-1)} \quad .$$

Furthermore the MSE of r is

$$M(r) = a^2 \left[\frac{m+2}{m^2(m-1)(m-2)} \right] + \delta \left[\frac{1}{(m-1)(m-2)} \right] .$$

Now let $t_2 = \frac{1}{2}(r_1+r_2) = \frac{1}{2}\left(\frac{\bar{y}_1}{\bar{x}_1} + \frac{\bar{y}_2}{\bar{x}_2}\right)$ as before and then

$$E(t_2-b) = \frac{a}{2}\left(\frac{1}{m-2} + \frac{1}{m-2}\right) = \frac{a}{m-2} .$$

Thus

$$\text{Bias}(t_2) = \frac{a}{m-2} - \frac{a}{m} = \frac{2a}{m(m-2)} .$$

Consequently, instead of choosing $R = \frac{1}{2}$ which leads to $\hat{R}_2$ studied by Durbin and more recently by P.S.R.S. Rao (1969) it is obvious that the proper choice is

$$R = \frac{\text{Bias } (r)}{\text{Bias } (t_2)} = \frac{m-2}{2(m-1)} .$$

As in the previous section the result is

$$\hat{R}_3 = cr-d\left(\frac{r_1+r_2}{2}\right) ,$$

where $c = \frac{2(m-1)}{m}$ and $d = \frac{m-2}{m}$.

It follows that $\hat{R}_3$ is unbiased for ρ and using the development of equations (3.7) and (3.8) we have

$$E(\hat{R}_3-b)^2 = a^2E\left\{\frac{c^2}{\bar{x}^2} + \frac{d^2}{4}\left(\frac{1}{\bar{x}_1^2} + \frac{1}{\bar{x}_2^2}\right) + \left(\frac{d^2}{2} - 2cd\right)\left(\frac{1}{\bar{x}_1\bar{x}_2}\right)\right\}$$

$$+ \frac{\delta}{2} E\left\{\frac{2c^2}{\bar{x}} + d^2\left(\frac{1}{\bar{x}_1^2} + \frac{1}{\bar{x}_2^2}\right) - 4cd\left(\frac{1}{\bar{x}_1\bar{x}_2}\right)\right\}$$

$$= a^2\left[\frac{c^2}{(m-1)(m-2)} + \frac{d^2}{2(m-2)(m-4)} + \frac{d^2/2-2cd}{(m-2)^2}\right]$$

$$+ \frac{\delta}{2}\left[\frac{2c^2}{(m-1)(m-2)} + \frac{2d^2}{(m-2)(m-4)} - \frac{4cd}{(m-2)^2}\right]$$

$$= \frac{a^2}{2m^2}\left[\frac{8(m-1)}{(m-2)} + \frac{(m-2)}{(m-4)} - \frac{7m^2-20m+12}{(m-2)^2}\right]$$

$$+ \frac{\delta}{2m^2}\left[\frac{8(m-1)}{(m-2)} + \frac{2(m-2)}{m-4} - \frac{8m^2-24m+16}{(m-2)^2}\right] .$$

After more algebra we obtain

$$E(\hat{R}_3-b)^2 = \frac{a^2(m-3)}{m^2(m-4)} + \frac{\delta(m-2)}{m^2(m-4)} ,$$

and therefore the MSE of $\hat{R}_3$ is

$$M(\hat{R}_3) = E(\hat{R}_3-b)^2 - \frac{a^2}{m^2}$$

$$= \frac{a^2}{m^2(m-4)} + \frac{\delta(m-2)}{m^2(m-4)} .$$

Durbin compares the MSE of r to that of $\hat{R}_2$ and determines that

$$M(\hat{R}_2) < M(r)$$

provided that $m > 16$ and that the inequality might hold true for some values of m between 10 and 16. In order to attach some meaning to these values recall that

$$E(x) = \mathrm{Var}\ (x) = m$$

and hence the coefficient of variation of x is $m^{-1/2}$. Let us make this same comparison using the new estimator $\hat{R}_3$. Now

$$M(r)-M(\hat{R}_3) = \frac{a^2}{m^2}\left[\frac{(m+2)}{(m-1)(m-2)} - \frac{1}{(m-4)}\right]$$

$$+ \delta\left[\frac{1}{(m-1)(m-2)} - \frac{(m-2)}{m^2(m-4)}\right]$$

$$= \frac{a^2}{m^2}\left[\frac{(m-10)}{(m-1)(m-2)(m-4)}\right] + \frac{\delta}{m^2}\left[\frac{m^2-8m+4}{(m-1)(m-2)(m-4)}\right]$$

which is certainly positive for $m > 10$. Moreover, since the roots of (m^2-8m+4) are real, positive and less than 8, the inequality

$$M(\hat{R}_3) < M(r)$$

may hold for some values of m between 8 and 10, and is surely valid for $m > 10$.

Further evidence of the superiority of $\hat{R}_3$ over $\hat{R}_2$ is apparent in the comparison,

$$M(\hat{R}_2)-M(\hat{R}_3) = \frac{a^2}{m^2}\left[\frac{m^3-5m^2+12m+16}{(m-1)(m-2)^2(m-4)} - \frac{1}{(m-4)}\right]$$

$$+ \delta\left[\frac{m^2-7m+18}{(m-1)(m-2)^2(m-4)} - \frac{(m-2)}{m^2(m-4)}\right]$$

$$= \frac{a^2(4m+20)}{m^2(m-1)(m-2)^2(m-4)} + \frac{\delta(20m-8)}{m^2(m-1)(m-2)^2(m-4)} .$$

which is positive over the entire range of reasonable values for m. Therefore, utilizing the general procedure introduced in the previous chapter we have realized a strict improvement in the bias and MSE of the ratio estimator under the assumed model.

3.2 Estimation of a Truncation Point

In many practical applications of statistical theory the random variables which enter the mathematical model are restricted to a finite range; even though the distribution, which has been satisfactorily assumed, may ordinarily be defined upon unbounded random variables. In these instances the estimation of the bounds may be of definite practical importance.

Suppose an estimate of a truncation point, θ, is desired; and we have a random sample of size n from the distribution F(x) which has been truncated to $x \leq \theta$, i.e.

$$X \sim F_\theta(x) = F(x)/F(\theta) .$$

A natural choice for the estimator is the largest of the n observations, that is let

$$t_1 = \max\left\{X_1, X_2, \ldots, X_n\right\} = X_{(n)} .$$

Following Robson and Whitlock (1964) the bias in t_1 may be computed if we first make the probability transformation

$$Y = F_\theta(X_{(n)}) = F(X_{(n)})/F(\theta)$$

so that

$$X_{(n)} = H_{F_\theta}(Y) \quad .$$

Secondly, the function $H_{F_\theta}(Y)$ may be expanded in a Taylor series about the point Y=1, to yield (omitting the subscript for H)

$$(3.9) \qquad X_{(n)} = \theta + (Y-1)H'(1) + \frac{(Y-1)^2}{2!}H''(1) + \frac{(Y-1)^3}{3!}H'''(1) + \ldots,$$

where

$$H(1) = \theta \quad ,$$

$$H'(1) = \frac{F(\theta)}{F'(\theta)} \quad ,$$

$$H''(1) = -\left[\frac{F(\theta)}{F'(\theta)}\right]^2 \frac{F''(\theta)}{F'(\theta)} \quad ,$$

$$H'''(1) = \left[\frac{F(\theta)}{F'(\theta)}\right]^3 \left\{3\left[\frac{F''(\theta)}{F'(\theta)}\right]^2 - \frac{F'''(\theta)}{F'(\theta)}\right\} \quad ,$$

and so forth.

Since (1-Y) is distributed as Y_1 the smallest of a random sample of size n from the uniform distribution over (0,1) we have

$$E(Y-1)^k = (-1)^k E(Y_1^k) = (-1)^k \frac{k!\ n!}{(n+k)!} \quad ,$$

and therefore

$$(3.10) \qquad E(X_{(n)}) = \theta - \left(\frac{1}{n+1}\right)H'(1) + \frac{1}{(n+1)(n+2)} H''(1) - \frac{n!}{(n+3)!} H'''(1) + \ldots$$

Miller (1964) draws upon this general problem to illustrate the pitfalls of universal application of the ordinary jackknife. Let

$$t_2 = \frac{1}{n} \sum_{i=1}^{n} {}_i t_1$$

as in the previous chapter, then recall that the jackknife always employs $R=(n-1)/n$ to yield

$$\hat{\theta} = nt_1 - (n-1)t_2 \quad .$$

When $t_1 = X_{(n)}$ this procedure produces

$$\hat{\theta} = X_{(n)} + \frac{n-1}{n} (X_{(n)} - X_{(n-1)}) \ .$$

Miller considers three classes of distributions as to their behaviour near the point of truncation in order to demonstrate various departures from the desired asymptotic properties of a Student-t-like statistic.

Aside from these considerations it should be slightly discouraging that $\hat{\theta}$ remains biased even when X has a uniform distribution over $(0,\theta)$. In this case where

$$F_\theta(x) = \frac{x}{\theta} \quad , \qquad 0<x<\theta \quad ,$$

since $H'_{F_\theta}(1)=\theta$ is the only nonvanishing derivative in equations (3.9) and so (3.10) becomes

$$E(X_{(n)}) = \theta - \left(\frac{1}{n+1}\right)\theta = \frac{n}{n+1}\theta$$

as is well known.

Hence if we take R to be the ratio of the biases in t_1 and t_2 then

$$R = \left(\frac{\theta}{n+1}\right) \Big/ \left(\frac{\theta}{n}\right) = \frac{n}{n+1}$$

and

$$\hat{\theta} = (n+1)X_{(n)} - n\left(\frac{n-1}{n} X_{(n)} + \frac{1}{n} X_{(n-1)}\right) \tag{3.11}$$

$$= 2X_{(n)} - X_{(n-1)}$$

which is completely unbiased for θ in this uniform example.

Interestingly, Robson and Whitlock (1964), almost concurrently with Miller's study, through consideration of (3.10) proposed the estimator given in (3.11) as being "modified to fit the factorial series rather than the power series in 1/n". This appears to be one instance in the literature where the principle of the transformation is used instead of the rule. The results of this are generally good and Miller (1968) later refers to this alteration as "satisfactory" performance of the "jackknife", although, strictly speaking, it is no longer the jackknife.

Note that $\hat{\theta}$ as given in (3.11) should be used for all distributions in the class under consideration and the bias will be given by

$$E(\hat{\theta}-\theta) = -\frac{1}{(n+1)(n+2)} H''(1) + \frac{2n!}{(n+3)!} H'''(1) - \frac{3n!}{(n+4)!} H^{(4)}(1) + \ldots$$

An interesting reapplication of the technique for this problem is given at the end of the next chapter.

Using the proper R the estimate for the variance which we obtain from the "psuedo-values" is

$$\frac{1}{n^2} \sum_{i=1}^{n} (\hat{\theta}_i - \hat{\theta})^2 = (X_{(n)} - X_{(n-1)})^2 ,$$

and Robson and Whitlock suggest a good approximate confidence bound for θ of the form

$$X_{(n)}+C_{\alpha}(X_{(n)}-X_{(n-1)}) \ .$$

Also note that for the uniform example it may be recalled that

$$MSE(t_1) = E[(X_{(n)}-\theta)^2] = \frac{2\theta^2}{(n+1)(n+2)}$$

and, even though the inner order statistic $X_{(n-1)}$ is introduced in $\hat{\theta}$ and hence the new estimator is not based on the sufficient statistic alone, we have

$$MSE\ (\hat{\theta}) = Var(\hat{\theta}) = \frac{2\theta^2}{(n+1)(n+2)} \quad .$$

Furthermore we may use the Rao-Blackwell theorem to obtain

$$\hat{\theta}_{RB} = E[\hat{\theta}|X_{(n)}] = \frac{n+1}{n}\ X_{(n)}$$

which is the unique minimum variance unbiased estimate of θ for this simple example.

CHAPTER IV

HIGHER ORDER TRANSFORMATIONS

4.1 Reapplication and Motivation

In one of the earliest articles on the ordinary jackknife procedure Quenouille (1956) suggests a reapplication of the method in order to eliminate the $O(n^{-2})$ term which remains after an initial application. When the assumed model for the bias is

$$E[t_1]-\theta = \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \dots \quad ,$$

$$t_2 = \bar{t}_{n-1} \qquad \text{and} \qquad \hat{\theta} = nt_1-(n-1)t_2 \quad ,$$

then $\hat{\theta}$ is biased by terms of order $O(n^{-2})$. Specifically

$$E[\hat{\theta}]-\theta = -\frac{a_2}{n(n-1)} - \frac{(2n-1)a_3}{n^2(n-1)^2} - \dots \ ,$$

which, because

$$\frac{1}{n(n-1)} = \frac{1}{n^2}\left(1 + \frac{1}{n} + \frac{1}{n^2} + \dots\right) \quad ,$$

Quenouille and later Kendall and Stuart (1961) choose to write as

$$E[\hat{\theta}]-\theta = -\frac{a_2}{n^2} + O(n^{-3}) \quad .$$

Hence the form given in their works for a second application is

$$\hat{\theta}^{(2)} = \frac{n^2\,\hat{\theta} - (n-1)^2\,\bar{\hat{\theta}}}{n^2-(n-1)^2}$$

which, as they state, is unbiased to terms of order $O(n^{-3})$. The serious flaw in this particular rule is that if $a_k=0$ for all $k > 2$ then $\hat{\theta}^{(2)}$ is <u>not</u> exactly unbiased as one would have desired.

In order to clarify the symbols employed in the discussion of re-applied transformations recall that for $R = (n-1)/n$

$$\hat{\theta} = nt_n - \frac{n-1}{n} \sum_{i=1}^{n} {}_it_{n-1}$$

so that if the jth unit is omitted from the sample the corresponding statistic is

$$\hat{\theta}_j = (n-1)\ {}_jt_{n-1} - \frac{n-2}{n-1} \sum_{\substack{i=1\\ i\neq j}}^{n} {}_{ji}t_{n-2}\ .$$

Hence averaging over j yields

$$\bar{\hat{\theta}} = \frac{1}{n} \sum_{j=1}^{n} \hat{\theta}_j \tag{4.1}$$

$$= \frac{n-1}{n} \sum_{j=1}^{n} {}_jt_{n-1} - \frac{n-2}{n(n-1)} \sum\sum_{i\neq j} {}_{ji}t_{n-2}$$

$$= (n-1)\ \bar{t}_{n-1} - (n-2)\ \bar{t}_{n-2}\ .$$

Therefore the reapplication suggested by Quenouille may be written

$$\hat{\theta}^{(2)} = \frac{n^2\left[nt_n - (n-1)\bar{t}_{n-1}\right] - (n-1)^2\left[(n-1)\bar{t}_{n-1} - (n-2)\bar{t}_{n-2}\right]}{n^2 - (n-1)^2}$$

$$= \frac{n^3 t_n - (2n^2 - 2n + 1)\ (n-1)\bar{t}_{n-1} + (n-1)^2(n-2)\bar{t}_{n-2}}{2n-1} .$$

It may now be seen that

$$E[\hat{\theta}^{(2)}] = \frac{n^3-2n^3+4n^2-3n+1+n^3-4n^2+5n-2}{2n-1}\ \theta$$

$$+ \frac{n^2-2n^2+2n-1+n^2-2n+1}{2n-1}\ a_1$$

$$+ \frac{(n^2-n)\ (n-2) - (2n^2-2n+1)\ (n-2) + (n-1)^3}{(n-1)\ (n-2)\ (2n-1)}\ a_2$$

$$+ O(n^{-3})$$

$$= \theta + 0.a_1$$

$$+ \frac{n^3-3n^2+2n-2n^3+6n^2-5n+2+n^3-3n^2+3n-1}{(n-1)\ (n-2)\ (2n-1)}\ a_2$$

$$+ O(n^{-3}) .$$

And as mentioned previously, if a_1 and a_2 are the only two nonzero terms in the power series expansion in 1/n for the bias, it would be desirable for $\hat{\theta}^{(2)}$ to be unbiased. However, in this event

$$E[\hat{\theta}^{(2)}] - \theta = \frac{a_2}{(n-1)\ (n-2)\ (2n-1)}$$

when the reapplication is carried out according to Quenouille.

A more recent attempt to prescribe the reapplication technique was made by Mantel (1967). In this instance the author elects to retain the value $R = (n-1)/n$ for each step of the procedure. Hence

$$\hat{\theta} = nt_n - (n-1)\ \bar{t}_{n-1}$$

and

$$\hat{\theta}^{(2)} = n\hat{\theta} - (n-1)\bar{\hat{\theta}}$$

$$= n^2 t_n - (2n-1)(n-1)\bar{t}_{n-1} + (n-1)(n-2)\bar{t}_{n-2} \quad .$$

Clearly both of these procedures are lacking because the essential worth of the parameter R is overlooked or ignored.

After a single application of the ordinary jackknife we have,

$$E[\hat{\theta}] - \theta = -\frac{a_2}{n(n-1)} - \frac{(2n-1)a_3}{n^2(n-1)^2} - \cdots ,$$

which indicates a proper value of

$$R = \frac{1}{n(n-1)} \Big/ \frac{1}{(n-1)(n-2)} = \frac{n-2}{n}$$

if the combining second estimator is to be obtained as in (4.1). Hence

$$\hat{\theta}^{(2)} = \frac{n}{2}\hat{\theta} - \frac{(n-2)}{2}\bar{\hat{\theta}}$$

$$= \frac{n^2}{2} t_n \quad \frac{n(n-1)}{2}\bar{t}_{n-1} - \frac{(n-1)(n-2)}{2}\bar{t}_{n-1} + \frac{(n-2)^2}{2}\bar{t}_{n-2} \tag{4.2}$$

$$= \frac{n^2 t_n - 2(n-1)^2\bar{t}_{n-1} + (n-2)^2\bar{t}_{n-2}}{2} \quad .$$

In this instance, if a_1 and a_2 are the only two nonzero coefficients in the power series, it may be shown that

$$E[\hat{\theta}^{(2)}] = \frac{n^2 - 2(n-1)^2 + (n-2)^2}{2}\theta$$

$$+ \frac{n - 2n + 2 + n-2}{2} a_1$$

$$+ \frac{1-2+1}{2} a_2 = \theta \ .$$

Since this formulation of the reapplied transformation appears to be consistent with the basic aim of the procedure, the subsequent reapplications must not be ignored. An algorithm for computation of the kth iterate of the procedure would be advantageous if the evaluation of the appropriate R for each step is avoided. Consideration of this question leads to what will be called a higher order transformation.

Recalling the straight line formulation of the problem within Tukey's graphical method, suggests the possible use of a determinantal equation. Upon examining the use of determinants by Shanks (1955) or Gray, Atchison and McWilliams (1970) an alternative form of the ordinary jackknife becomes apparent which furthermore is suggestive for a logical extension of the method. Note that

$$\hat{\theta} = \frac{\begin{vmatrix} t_1 & t_2 \\ 1/n & 1/(n-1) \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ 1/n & 1/(n-1) \end{vmatrix}}$$

or more generally, if

$$E[t_1] - \theta = f_1(n)b_1(\theta)$$

and

$$E[t_2] - \theta = f_2(n)b_1(\theta) \quad ,$$

then

(4.3)
$$\hat{\theta} = \frac{\begin{vmatrix} t_1 & t_2 \\ f_1 & f_2 \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ f_1 & f_2 \end{vmatrix}}$$

$$= \frac{f_2 t_1 - f_1 t_2}{f_2 - f_1}$$

$$= \frac{1}{1-R} t_1 - \frac{R}{1-R} t_2$$

since $R = f_1/f_2$. This is sufficient motivation to consider an extended version of the transformation.

4.2 Definition and Properties

If there exist (k+1) estimators of θ ($k<n$) with distinct non zero biases based on n observations such that

$$E[t_j] - \theta = \sum_{i=1}^{\infty} f_{ij}(n) b_i(\theta), \quad j=1,\ldots,k+1$$

then in many instances the following quantity will exist and prove useful in estimation problems, at least for small k.

Definition 4.1

The determinant in the denominator of the following is nonzero, define the kth order generalized jackknife $\hat{\theta}^{(k)}$ by

$$(4.4) \qquad \hat{\theta}^{(k)} = \frac{\begin{vmatrix} t_1 & t_2 \cdots & t_{k+1} \\ f_{11} & f_{12} \cdots & f_{1,k+1} \\ f_{21} & f_{22} \cdots & f_{2,k+1} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ f_{k1} & f_{k2} & f_{kk+1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \cdots & 1 \\ f_{11} & f_{12} \cdots & f_{1,k+1} \\ f_{21} & f_{22} \cdots & f_{2,k+1} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ f_{k1} & f_{k2} \cdots & f_{k,k+1} \end{vmatrix}} .$$

Examination of $\hat{\theta}^{(1)}$ will produce the conclusion that $\hat{\theta}$ as given in equation (4.3) is special case of $\hat{\theta}^{(k)}$ and therefore we denote $\hat{\theta} = \hat{\theta}^{(1)}$. Also, if C_k^j denotes the cofactor of t_j in the determinant of the numerator then we may write

$$\hat{\theta}^{(k)} = \frac{\sum_{j=1}^{k+1} C_k^j \, t_j}{\sum_{j=1}^{k+1} C_k^j} .$$

One indication of the propriety of this definition may be obtained by examining $\hat{\theta}^{(2)}$ when

$$f_{i,j+1} = \frac{1}{(n-j)^i} .$$

In other words, consider

$$t_1 = t_n,\ t_2 = \bar{t}_{n-1} \text{ and } t_3 = \bar{t}_{n-2}$$

so that

$$\hat{\theta}^{(2)} = \frac{\begin{vmatrix} t_n & \bar{t}_{n-1} & \bar{t}_{n-2} \\ \frac{1}{n} & \frac{1}{(n-1)} & \frac{1}{(n-2)} \\ \frac{1}{n^2} & \frac{1}{(n-1)^2} & \frac{1}{(n-2)^2} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ \frac{1}{n} & \frac{1}{(n-1)} & \frac{1}{(n-2)} \\ \frac{1}{n^2} & \frac{1}{(n-1)^2} & \frac{1}{(n-2)^2} \end{vmatrix}} .$$

Here

$$C_2^1 = \left[\frac{1}{(n-1)(n-2)^2} - \frac{1}{(n-1)^2(n-2)}\right] = \frac{1}{(n-1)^2(n-2)^2}$$

$$C_2^2 = -\left[\frac{1}{n(n-2)^2} - \frac{1}{n^2(n-2)}\right] = \frac{-2}{n^2(n-2)^2}$$

$$C_2^3 = \left[\frac{1}{n(n-1)^2} - \frac{1}{n^2(n-1)}\right] = \frac{1}{n^2(n-1)^2} ,$$

and thus the determinant of the denominator is

$$\sum_{j=1}^{3} c_2^j = \frac{n^2-2(n-1)^2+(n-2)^2}{n^2(n-1)^2(n-2)^2}$$

$$= \frac{2}{n^2(n-1)^2(n-2)^2} \quad .$$

Therefore

$$\hat{\theta}^{(2)} = \frac{n^2 t_n - 2(n-1)^2\ \bar{t}_{n-1} + (n-2)^2\ \bar{t}_{n-2}}{2}$$

which is identical to the expression obtained in (4.2) for a proper second application of the general procedure. This fact and the discussion of the previous section indicate the validity of the following:

<u>Theorem 4.1</u>

If $\hat{\theta}^{(k)}$ exists and either

i) $f_{ij}(n)=0$, $j=1, \ldots, k+1$, for all $i>k$

or

ii) $b_i(\theta) = 0$, for all $i>k$

then

$$E\left[\hat{\theta}^{(k)}\right] = \theta \ .$$

In other words, if

$$E[t_j]-\theta = \sum_{i=1}^{k} f_{ij}(n) b_i(\theta), \ j=1, \ldots, k+1$$

and $\hat{\theta}^{(k)}$ exists then it is unbaised

<u>Proof:</u>

Since

$$\hat{\theta}^{(k)} = \frac{\begin{vmatrix} t_1 & t_2 & \cdots & t_{k+1} \\ f_{11} & f_{12} & \cdots & f_{1,k+1} \\ f_{21} & f_{22} & \cdots & f_{2,k+1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ f_{k1} & f_{k2} & \cdots & f_{k,k+1} \end{vmatrix}}{\sum_{j=1}^{k+1} C_k^j},$$

it is clear that

$$E\left[\hat{\theta}^{(k)}\right] = \theta + \begin{vmatrix} \sum_{i=1}^{k} f_{i1}b_i & \sum_{i=1}^{k} f_{i2}b_i & \cdots & \sum_{i=1}^{k} f_{i,k+1}b_i \\ f_{11} & f_{12} & \cdots & f_{1,k+1} \\ f_{21} & f_{22} & \cdots & f_{2,k+1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ f_{k1} & f_{k2} & \cdots & f_{k,k+1} \end{vmatrix} \cdot \left[\sum_{j=1}^{k+1} C_k^j\right]^{-1}.$$

Now the determinant in the numerator of the second term is zero since the first row is obviously a linear combination of the other k rows. The $(k+1)^{st}$ row being multiplied by a negative b_i and the sum over $i(i=1,\ldots,k)$ added

to the first row produces a row of all zeroes. Therefore

$$E[\hat{\theta}^{(k)}] = \theta$$

as was to be shown.

We may give a general expression for the bias in $\hat{\theta}^{(k)}$ by employing the argument offered in the above proof to obtain

$$E[\hat{\theta}^{(k)}] - \theta = \frac{C_k^1 \sum_{i=k+1}^{\infty} f_{i1}b_i + C_k^2 \sum_{i=k+1}^{\infty} f_{i2}b_i + \dots + C_k^{k+1} \sum_{i=k+1}^{\infty} f_{i,k+1}b_i}{\sum_{j=1}^{k+1} C_k^j}$$

(4.5)

$$= \frac{\sum_{j=1}^{k+1} \sum_{i=k+1}^{\infty} C_k^j f_{ij}b_i}{\sum_{j=1}^{k+1} C_k^j} .$$

This expression suggests a special result when there exists a power series expansion in (1/n) for the bias in $t_1 = t_n(\underline{X}_n)$ as a consistent estimator of θ. And furthermore when the required accompanying estimators t_k have been obtained by averaging the analogous estimators evaluated over all subsamples of set size, i.e.,

$$t_{k+1} = \bar{t}_{n-k}(\underline{X}), \quad k<n.$$

In this case the following theorem applies.

Theorem 4.2

If $f_{i,j+1} = (n-j)^{-i}$ then the higher order transformation reduces the bias to terms of order k+1 in 1/n, i.e.

$$E[\hat{\theta}^{(k)}]-\theta = O(n^{-(k+1)}).$$

Proof:

From equation (4.5) above

$$E[\hat{\theta}^{(k)}] - \theta = \frac{\sum_{j=1}^{k+1} C_k^j \sum_{i=k+1}^{\infty} (n-j+1)^{-i} b^i}{\sum_{j=1}^{k+1} C_k^j}$$

(4.6)

$$= \frac{\sum_{j=1}^{k+1} C_k^j \, (n-j+1)^{-k-1} \sum_{i=1}^{\infty} b_{k+1}/(n-j+1)^{i-1}}{\sum_{j=1}^{k+1} C_k^j} .$$

Because all of the elements of the cofactors are positive and less than unity, the C_k^j are bounded above by one. Furthermore the C_k^j can each be written as a Vandermonde determinant,

$$C_k^j = \left[\prod_{\substack{i=0 \\ i\neq j-1}}^{k} \frac{1}{(n-i)}\right] \cdot \begin{vmatrix} 1 & 1 & \cdots & 1 & 1 & \cdots & 1 \\ \frac{1}{n} & \frac{1}{n-1} & & \frac{1}{(n-j+2)} & \frac{1}{(n-j)} & \cdots & \frac{1}{(n-k)} \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \frac{1}{n^{k-1}} & \frac{1}{(n-1)^{k-1}} & \cdots & \frac{1}{(n-j+2)^{k-1}} & \frac{1}{(n-j)^{k-1}} & \cdots & \frac{1}{(n-k)^{k-1}} \end{vmatrix} \cdot (-1)^{j+1}$$

$$= \left[\prod_{\substack{i=0\\ i\neq j-1}}^{k} \frac{1}{(n-i)}\right]\left[\prod_{\substack{\ell > m\\ \ell,m\neq j-1}}^{k} \frac{1}{n-\ell} - \frac{1}{n-m}\right] \cdot (-1)^{j+1}$$

$$= 0(n^{-k}) \cdot 0\left(n^{-k(k-1)}\right)$$

$$= 0\left(n^{-k^2}\right) .$$

The series

$$\sum_{i=1}^{\infty} b_{k+i}/(n-j+1)^{i-1}$$

is clearly of order 0(1) and therefore, consideration of the last expression within the algebra of order relations allows one to conclude that

$$E[\hat{\theta}^{(k)}]-\theta = 0(n^{-k-1}),$$

as was to be shown.

To illustrate the use of this higher order transformation consider the estimation of the parameter μ^4 from a random sample of size n for the normal distribution $N(\mu,\sigma^2)$. That is, suppose the

$$X_i \text{ are i.i.d. as } N(\mu,\sigma^2) \quad i=1,\ldots,n$$

and $\theta=\mu^4$ and it is proposed that $t_1 = \bar{x}^4$. We know that

$$E[t_1] = \mu^4 + \frac{6\mu^2\sigma^2}{n} + \frac{3\sigma^4}{n^2}$$

and hence if

$$t_2 = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{(n-1)^4} (n\bar{x}-x_i)^4 ,$$

$$t_3 = \frac{2}{n(n-1)} \sum_{i > j}\sum \frac{1}{(n-2)^4} (n\bar{x}-x_i-x_j)^4 ,$$

and

$$\hat{\theta} = \frac{1}{2}\ (n^2 t_1 - 2\ (n-1)^2 t_2 + (n-2)^2 t_3),$$

then

$$E[\hat{\theta}] = \theta.$$

Also recall the estimator for a truncation point and the expression for its bias as given in equation (3.10). For this problem we have

$$t_1 = \max\{X_1, \ldots, X_n\} = X_{(n)} \quad ,$$

and again omitting subsamples of size one and averaging,

$$t_2 = \frac{n-1}{n}\ X_{(n)} + \frac{1}{n}\ X_{(n-1)} \quad ,$$

and then all possible subsamples of size two,

$$t_3 = \frac{2}{n(n-1)} \left[\frac{(n-1)(n-2)}{2}\ X_{(n)} + (n-2)\ X_{(n-1)} + X_{(n-2)} \right] \quad .$$

In the notation of this chapter

$$f_{11}(n) = \frac{1}{n+1}\ , \quad f_{21}(n) = \frac{1}{(n+1)(n+2)}$$

and thus

$$f_{12}(n) = \frac{1}{n}\ , \quad f_{22}(n) = \frac{1}{n(n+1)} \quad ,$$

$$f_{13}(n) = \frac{1}{n-1}, \ f_{23}(n) = \frac{1}{n(n-1)} \quad .$$

Hence

$$\hat{\theta}^{(2)} = \begin{vmatrix} t_1 & t_2 & t_3 \\ \frac{1}{n+1} & \frac{1}{n} & \frac{1}{n-1} \\ \frac{1}{(n+1)(n+2)} & \frac{1}{n(n+1)} & \frac{1}{n(n-1)} \end{vmatrix} \cdot \left[\sum_{j=1}^{3} C_2^j \right]^{-1}$$

where

$$C_2^1 = \left[\frac{1}{n^2(n-1)} - \frac{1}{n(n-1)(n+1)} \right] = \frac{1}{n^2(n-1)(n+1)} ,$$

$$C_2^2 = \left[\frac{1}{(n-1)(n+1)(n+2)} - \frac{1}{n(n-1)(n+1)} \right] = \frac{-2}{n(n-1)(n+1)(n+2)}$$

and

$$C_2^3 = \left[\frac{1}{n(n+1)^2} - \frac{1}{n(n+1)(n+2)} \right] = \frac{1}{n(n+1)^2(n+2)} .$$

Now

$$\sum_{j=1}^{3} C_2^j = \frac{2}{n^2(n+1)^2(n-1)(n+2)}$$

so

$$\begin{aligned} \hat{\theta}^2 &= \frac{1}{2} \left[(n+1)(n+2)t_1 - 2n(n+1)t_2 + n(n-1)t_3 \right] \\ &= \frac{1}{2} \left[(n+1)(n+2) - 2(n+1)(n-1) + (n-1)(n-2) \right] X_{(n)} \\ &\quad + \frac{1}{2} [2(n-2) - 2(n+1)] X_{(n-1)} + X_{(n-2)} \\ &= 3X_{(n)} - 3X_{(n-1)} + X_{(n-2)} . \end{aligned}$$

This is the same estimator which Robson and Whitlock (1964) derive by adhering to the proper principles for reapplication of a bias reduction scheme. In the notation of section (3.2) the bias of $\hat{\theta}^{(2)}$ is

$$E[\hat{\theta}^{(2)}]-\theta = - \frac{n!}{(n+3)!} H''' (1) + \frac{3n!}{(n+4)!} H^{(4)} (1) - \ldots$$

which is clearly $O(n^{-3})$.

CHAPTER V

SUMMARY AND DISCUSSION

The general transformation of an estimator, which may be classified as the re-use of sample information, has been shown to be helpful in several applied problems. In the cases where the ordinary jackknife is appropriate the present procedure will incorporate it as a special case. For those applications where Quenouille's method has been used and the more general approach is appropriate, a definite improvement has been demonstrated, e.g. ratio estimation and truncation points.

The fact that some applications may be found for which the bias reduction is accompanied by a decrease in MSE or even in variance suggests the desirability of a characterization theorem. In other words it would be helpful if one could characterize the distribution of the estimator t_1, say, for a fixed method of producing t_2 such that

$$\text{MSE } (\hat{\theta}) \leq \text{MSE } (t_1) \ .$$

If the problem of linear combination of two correlated estimators is examined from the viewpoint of minimizing the variance of the combined estimate then the proper coefficients are well known. To minimize the variance of $\hat{\theta}$ without regard to the restriction that R be the ratio of the biases we take

$$(5.1) \quad \hat{\theta} = \frac{[\text{Var}(t_2)-\text{Cov}(t_1,t_2)]t_1+[\text{Var}(t_1)-\text{Cov}(t_1,t_2)]t_2}{\text{Var}(t_1-t_2)} \ .$$

Next suppose that for some constant $c > 0$

$$\mathrm{Var}(t_1) \cong \frac{c}{n^2} \quad , \quad \mathrm{Var}(t_2) \cong \frac{c}{(n-1)^2}$$

and

$$\mathrm{Cov}(t_1, t_2) \cong \frac{c}{n(n-1)} \quad .$$

An estimator, t_1, exhibiting this character may be said to be superconsistent, a phenomenon occasionally associated with estimates of a range parameter. The correlation of t_1 and t_2 is near unity when t_2 is the same estimator as t_1 but evaluated for subsets of n-1 as has been previously discussed. Regardless of how such variances might arise the equation (5.1) becomes

$$\hat{\theta} = \frac{\frac{n-(n-1)}{(n-1)^2 n} t_1 + \frac{(n-1)-n}{n^2(n-1)} t_2}{\frac{1}{n^2(n-1)^2}}$$

$$= nt_1 - (n-1)t_2 \quad ,$$

the ordinary jackknife. The example mentioned as a possible source of superconsistency is one in which the value of $R = (n-1)/n$ has been shown to be inappropriate.

The major objections which have been raised against the special case (jackknife), of the present technique are mainly concerned with asymptotic properties. The asymptotic distribution of a test statistic derived from the pseudo-values is not crucial if one is primarily concerned with improved point estimates. The difficulty for median estimation reported by

Miller (1968) which is due to Lincoln Moses is again one which is related to the asymptotic distribution. Moses (1970) states that in his study; "the estimated variance turned out to be the reciprocal of an estimate of the squared density at the median (as it should be) but multiplied by the wrong constant." Even though the more general parameter R may not remove this difficulty, the difficulty is not one which pertains directly to the improved point estimate.

BIBLIOGRAPHY

1. Arvesen, J. N. (1969). Jackknifing U-statistics, Ann. Math. Statist., 40, pp. 2076-2100.

2. Burdick, D. S. (1961). Stage by stage modification of polynomial estimators by the jackknife method, Microfilmed Ph. D. Dissertation, Princeton University.

3. Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios, Biometrika, 46, pp. 477-480.

4. Goodman, L. A. and Hartley, H. O. (1958). The precision of unbiased ratio-type estimators, J. Am. Statist. Assoc., 53, pp. 491-508.

5. Gray, H. L. and Atchison, T. A. (1968). The generalized G-transform, Math. Comp., 22, pp. 595-605.

6. Gray, H. L., and Atchison, T. A. and McWilliams, G. V. (1970). Higher order G-transformations, SIAM J. Numer. Anal., to appear.

7. Gray, H. L. and Schucany, W. R. (1968). On the evaluation of distribution functions, J. Am. Statist. Assoc., 63, pp. 715-720.

8. Hammersly, J. M. and Mauldon, J. G. (1956). General principles of antithetic variates, Proc. Camb. Phil. Soc., 52, pp. 476-481.

9. Jones, H. L. (1963). The jackknife method, Proceedings of the IBM Scientific Computing Symposium on Statistics, pp. 185-197.

10. Kendall, M. G. and Stuart, A. (1961). The Advanced Theory of Statistics, Vol. 2, Hafner Publishing Company, New York.

11. Lubkin, S. (1952). A method of summing infinite series, J. Res. Nat. Bur. Standards, 48, pp. 228-254.

12. Mantel, N. (1967). Assumption-free estimators using U-statistics and a relationship to the jackknife method, Biometrics, 23, pp. 567-571.

13. Miller, R. G., Jr. (1964). A trustworthy jackknife, Ann. Math. Statist, 35, pp. 1594-1605.

14. Miller, R. G., Jr. (1968). Jackknifing variances, Ann. Math. Statist., 39, pp. 567-582.

15. Moses, L. E. (1970). Personal communication.

16. Quenouille, M. H. (1949). Approximate tests of correlation in time-series, J. Roy. Statist. Soc. Ser. B., 11, pp. 68-84.

17. Quenouille, M. H. (1956). Notes on bias in estimation, Biometrika, 43, pp. 353-360.

18. Raj, D. (1968). Sampling Theory, McGraw-Hill Book Company, New York.

19. Rao, J. N. K. (1965). A note on the estimation of ratios by Quenouille's method, Biometrika, 52, pp. 647-649.

20. Rao, J. N. K. (1969). Ratio and regression estimators, New Developments in Survey Sampling, N. L. Johnson and Harry Smith, Eds., Wiley-Interscience, New York, pp. 213-234.

21. Rao, J. N. K. and Webster, J. T. (1966). On two methods of bias reduction in the estimation of ratios, Biometrika, 53, pp. 571-577.

22. Rao, P. S. R. S. (1969). Comparison of four ratio-type estimates under a model, J. Am. Statist. Assoc., 64, pp. 574-580.

23. Robson, D. S. and Whitlock, J. H. (1964). Estimation of a truncation point, Biometrika, 51, pp. 33-39.

24. Shanks, D. (1955). Non-linear transformations of divergent and slowly convergent sequences, J. Math. Phys., 34, pp. 1-42.

25. Shorack, G. R. (1969). Testing and estimating ratios of scale parameters, J. Am. Statist. Assoc., 64, pp. 999-1013.

26. Tukey, J. W. (1958). Bias and confidence in not-quite large samples, Am. Math. Statist. (abstract), 29, p. 614.

27. Wynn, P. (1966). On the convergence and stability of the epsilon algorithm, SIAM J. Numer. Anal., 3, pp. 91-122.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* SOUTHERN METHODIST UNIVERSITY	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP UNCLASSIFIED

3. REPORT TITLE

The Reduction of Bias in Parametric Estimation

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*

William R. Schucany

6. REPORT DATE April 17, 1970	7a. TOTAL NO. OF PAGES 63 pages	7b. NO. OF REFS 27
8a. CONTRACT OR GRANT NO. N00014-68-A-0515 b. PROJECT NO. NR 042-260 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 93 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research

13. ABSTRACT

A general class of transformations of estimators is introduced which induces a reduction in bias if any exists. The concept is related to that of the sequence to sequence transformations which are employed for convergence improvement in deterministic cases such as the evaluation of infinite series and improper integrals. The procedure introduced by Quenouille (1949), (1956) and later termed the "jackknife" by Tukey (1958) is seen to be a special case of these transformations. The general principles of the method produce insight into the applications where the ordinary jackknife is not trustworthy.

To illustrate the method and demonstrate its potential usefulness several examples are considered. For ratio estimation under a particular model a new unbiased estimator is produced which exhibits a favorable mean square error relative to existing adjusted estimators.

The existing notion of reapplication of such a procedure is shown to lack the property for which it was designed. Proper reapplication is proposed so as to conform to general principles. A higher order transformation is defined which provides an interesting algorithm for the corrected procedure. Possible extensions to nonlinear transformations are also mentioned.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED
Security Classification

A-31408