PAIRWISE PROBABILITIES IN PROBABILITY NONREPLACEMENT SAMPLING

James R. Chromy, Research Triangle Institute

1. INTRODUCTION

The problem of controlling the pairwise probabilities, P(ij), in probability nonreplacement sampling has existed since the development of a general theory of probability nonreplacement sampling in conjunction with the use of the Horvitz-Thompson estimator (ref. 1). A number of researchers have treated the subject directly or indirectly including Raj (ref. 2), Hanurav (ref. 3), Brewer (ref. 4), Durbin (ref. 5), Sampford (ref. 6), Jessen (ref. 7), Rao and Bayless (ref. 8), and Dodds and Fryer (ref. 9). Some of these approached the problem through a superpopulation model and examined such criteria as the superpopulation expectation of the variance of the Horvitz-Thompson estimator or the stability of the variance estimator. Jessen posed an intuitive criterion which relates most closely to the approach followed in this research, but he did not solve for optimum designs directly.

A general superpopulation model with size measures defined to potentially achieve the principal advantages of probability nonreplacement sampling is described. Under this model the superpopulation expectation of the variance is not a function of the pairwise probabilities, P(ij). The criterion employed in seeking optimum designs is the superpopulation variance of the finite population variance. A general solution for the P(ij) in terms of the variances of the error terms in the superpopulation model is developed. Although the solution can be generalized to any sample size, n (within certain restrictions), the problem of choosing three-wise to n-wise probabilities may still be difficult. A method for obtaining three-wise probabilities, P(ijk), consistent with the pairwise probabilities, P(ij), has been developed.

2. THE HORVITZ-THOMPSON ESTIMATOR

A theory for PPS nonreplacement sampling for within stratum samples of two or more elements was presented by Horvitz and Thompson (ref. 1). The unbiased estimator within Horvitz and Thompson's class two of linear estimators for probability nonreplacement (pnr) designs has come to be known as the Horvitz-Thompson estimator and is of the form

$$t(s|pnr) = \sum_{u=1}^{n} Y[s(u)]/P[s(u)]$$

where

s = [s(1),s(2),...,s(n)], a sample of n distinct elements from a sampling frame of N elements,

Y(i) = the variate of interest in the sampling investigation for the i-th element in the universe, and

P(i) = probability that the i-th element is included in the sample.

In practice, the P(i) are usually determined so that

$$P(i) = nX(i)/X(+),$$

where

X(i) = a known positive-valued variate associated with the i-th element in the universe, and

$$X(+) = \sum_{i=1}^{N} X(i).$$

Horvitz and Thompson presented a formula for the variance of t(s|pnr) under PPS sampling without replacement given by

$$V[t(s|pnr)] = \sum_{i=1}^{N} Y^2(i)[1-P(i)]/P(i)$$

$$+ \sum_{i \neq j}^{N} \sum^{N} Y(i)Y(j)[P(ij) - P(i)P(j)]/[P(i)P(j)],$$

where

P(ij) = the probability that the i-th and j-th units are both included in the sample (or the pairwise probability for the i-th and j-th units).

Yates and Grundy (ref. 10) presented an alternate expression for the variance of the Horvitz-Thompson estimator, namely

$$V[t(s|pnr)] = \sum_{i<j}^{N} \sum^{N} W(ij)D^2(ij)$$

where

$$W(ij) = P(i)P(j)-P(ij),$$

and

$$D(ij) = Y(i)/P(i)-Y(j)/P(j).$$

They also presented an alternate variance estimator which may be expressed as

$$v[t(s|pnr)] = \sum_{u<v}^{n} \sum^{n} w[s(u),s(v)]D^2[s(u),s(v)]$$

where

$$w(ij) = W(ij)/P(ij).$$

3. THE BASIC SUPERPOPULATION MODEL

The model considered most relevant for study of the Horvitz-Thompson estimator and probability nonreplacement sampling is

$$Y(i) = BX(i) + e(i)$$

where

B = an unknown constant, and

e(i) = the error term which measures the deviation from the model.

As pointed out by Horvitz and Thompson (ref. 1), t(s|pnr) will have zero variance if P(i) is exactly proportional to Y(i) for all i. The proposed model is designed to take advantage of this possibility. By making the P(i) proportional to X(i), a nonzero variance of t(s|pnr) occurs only if some of the e(i) are nonzero.

Suppose one believed that a linear model with a nonzero intercept was more realistic for the population to be sampled. Such a model may be written as

$$Y(i) = A + BZ(i) + e(i)$$

where A and B are both unknown but nonzero constants and $Z(i)$ is a measure of size. Use of the Horvitz-Thompson estimator and probability nonreplacement sampling with $P(i)$ proportional to $Z(i)$ would not provide the opportunity for a zero variance of $t(s|pnr)$ even if the $e(i)$ were all zero. A zero variance would occur with the $e(i)$ equal to zero only if the $P(i)$ were exactly proportional to $[A/B + Z(i)]$. Unless one wishes to rely upon an extremely fortuitous set of error terms, $e(i)$, to rescue the design and estimation plans from producing a large variance of $t(s|pnr)$, it appears more appropriate to use the no-intercept model

$$Y(i) = BX(i) + e(i)$$

where

$$X(i) = A/B + Z(i).$$

To define $X(i)$, it is not necessary to know A and B individually, but only their ratio. Even a poor guess should lead to a better design than one that sets the $P(i)$ in such a way that the principal advantage of PPS sampling cannot be achieved.

In summary, the premise put forth here is that if one uses the Horvitz-Thompson estimator in conjunction with probability nonreplacement sampling, the proper course to follow is to set $P(i)$ exactly proportional to some known $X(i)$ which is believed to be approximately proportional to $Y(i)$. If the no-intercept model does not present a good approximation of the $Y(i)$ to $X(i)$ relationship, the proper course to follow does not involve a magical choice of $P(ij)$ to compensate for this shortcoming. It does involve the selection of some other variate as a measure of size. It should be noted that $X(i)$ could be defined as a function of several other known variables.

Consideration of an intercept model can be defended in the instance of multi-purpose sample surveys for modeling the variance of estimates associated with some less major purposes of the survey. The choice of the $P(i)$ and $P(ij)$ should probably still be based on some known variate, $X(i)$, which fits the no-intercept model for the particular $Y(i)$ that is associated with the major purpose of the survey.

Suppose that the sampling statistician assumes the no-intercept model and utilizes as much information as is available to him to construct $X(i)$ approximately proportional to the unknown $Y(i)$. The following assumptions about the error terms appear reasonable in the predata collection stage of the sample design process:

(1) $Ee(i) = 0$ for $i = 1,2,\ldots,N$,

(2) $E[e(i)/X(i)]^2 = s^2(i)$ (finite), for $i = 1,2,\ldots,N$, and

(3) the error terms are independent.

The assumption of independence is not unreasonable if the $X(i)$ are properly chosen and if stratification is used to its fullest extent consistent with requirements for the estimability of the variance. For most applications where PPS sampling is applied, deep stratification is also utilized.

If one first assumes that the value of B is fixed and that the error terms sum to zero over the N population elements, then the independence assumption is errorneous. Prior to sampling, however, B is not known and cannot be estimated. In this _a priori_ sense, independence of error terms may be assumed as an indication of complete lack of further knowledge about the $Y(i)$ given the $X(i)$.

4. THE SUPERPOPULATION VARIANCE OF THE FINITE POPULATION VARIANCE

A particular outcome of the superpopulation model may be represented by an N-dimensional vector of error terms and designated as $\underline{e}$. The transpose of $\underline{e}$ may be written as

$$[e(1),e(2),e(3),\ldots,e(N)].$$

For brevity, the Horvitz-Thompson estimator will be denoted by $t(s)$ and its variance for a particular finite population (i.e., for a particular $\underline{e}$) will be denoted by $V[t(s)|\underline{e}]$. The $\underline{e}$ indicates a particular sample of N error terms selected from the hypothetical superpopulation. The s indicates a particular sample of n elements selected from the finite population of N elements.

In the finite population sense, the variance is computed over all possible samples s. The finite population variance is still a function of the error vector $\underline{e}$. It is reasonable under the model to consider taking the expectation of $V[t(s)|\underline{e}]$ over all possible outcomes $\underline{e}$. The finite population variance for the Horvitz-Thompson estimator may be expressed in terms of the superpopulation model by selectively substituting $nX(i)/X(+)$ for $P(i)$ and $BX(i) + e(i)$ for $Y(i)$ in the expression given in section 1. After some simplification this yields

$$n^2X^{-2}(+)V[t(s)|\underline{e}] = \sum_{i=1}^{N} [B+e(i)/X(i)]^2P(i)[1-P(i)]$$

$$+ \sum_{i\neq j}^{N}\sum^{N} [B+e(i)/X(i)][B+e(j)/X(j)][P(ij)-P(i)P(j)].$$

After collecting terms in B^2 and B, this expression becomes

$$B^2\{\sum_{i=1}^{N} P(i)[1-P(i)] + \sum_{i\neq j}^{N}\sum^{N} [P(ij)-P(i)P(j)]\}$$

$$+ B\{2\sum_{i=1}^{N} [e(i)/X(i)]P(i)[1-P(i)]$$

$$+ \sum_{i\neq j}^{N}\sum^{N} [e(i)/X(i)][P(ij)-P(i)P(j)]$$

$$+ \sum_{i\neq j}^{N}\sum^{N} [e(j)/X(j)][P(ij)-P(i)P(j)]\}$$

$$+ \sum_{i=1}^{N} [e(i)/X(i)]^2P(i)[1-P(i)]$$

$$+ \sum_{i\neq j}^{N}\sum^{N} [e(i)/X(i)][e(j)/X(j)][P(ij)-P(i)P(j)].$$

After noting that

$$\sum_{j\neq i}^{N} [P(ij)-P(i)P(j)] = -P(i)[1-P(i)],$$

the expression simplifies further to

$$n^2X^{-2}(+)V[t(s)|\underline{e}] = \mathop{S}_{i=1}^{N}[e(i)/X(i)]^2P(i)[1-P(i)]$$

$$+ \mathop{S\,S}_{i\neq j}^{N\ N}[e(i)/X(i)][e(j)/X(j)][P(ij)-P(i)P(j)].$$

Given the assumptions about the error terms, e(i), as stated above, the superpopulation expectation of the variance is simply

$$EV[t(s)|\underline{e}] = n^{-2}X^2(+)\mathop{S}_{i=1}^{N}P(i)[1-P(i)]s^2(i).$$

Under the assumed model, the expectation of the variance is not a function of the pairwise probabilities and does not, therefore, provide any basis for selection of the pairwise probabilities.

Since the particular choice of pairwise probabilities cannot be used to reduce the expected variance, another criterion must be developed. Intuitively, the sampling statistician should want to protect against an extremely large variance. To protect against a large variance occurring most of the time, the sample design should produce variances that have a tight superpopulation distribution about their expected value. This intuitive criterion suggests minimizing the superpopulation variance of the finite population variance; this variance may be written as

$$V\{V[t(s)|\underline{e}]\} = E\{V^2[t(s)|\underline{e}]\}-E^2\{V[t(s)|\underline{e}]\}.$$

Using the results obtained above, the square of the variance, $V^2[t(s)|\underline{e}]$, may be represented as

$$n^4X^{-4}(+)V^2[t(s)|\underline{e}] =\{\mathop{S}_{i=1}^{N}[e(i)/X(i)]^2P(i)[1-P(i)]\}^2$$

$$+ \{\mathop{S\,S}_{i\neq j}^{N\ N}[e(i)/X(i)][e(j)/X(j)][P(ij)-P(i)P(j)]\}^2$$

$$+ 2\{\mathop{S}_{i=1}^{N}[e(i)/X(i)]^2P(i)[1-P(i)]\}\ x$$

$$\{\mathop{S\,S}_{i\neq j}^{N\ N}[e(i)/X(i)][e(j)/X(j)][P(ij)-P(i)P(j)]\}.$$

The third term, or the cross product term can be seen to have a zero expectation under the assumed superpopulation model. After expanding the first two terms and noting which terms in the products have nonzero expectation, the superpopulation expectation of the squared variance, $EV^2[t(s)|\underline{e}]$, may be represented as

$$EV^2[t(s)|\underline{e}] =$$

$$n^{-4}X^4(+)\{\mathop{S}_{i=1}^{N}E[e(i)/X(i)]^4P^2(i)[1-P(i)]^2$$

$$+\mathop{S\,S}_{i\neq j}^{N\ N}s^2(i)s^2(j)P(i)P(j)[1-P(i)][1-P(j)]$$

$$+ 2\mathop{S\,S}_{i\neq j}^{N\ N}s^2(i)s^2(j)[P(ij)-P(i)P(j)]^2\}.$$

It must be further assumed that $E[e(i)/X(i)]^4$ exists and is finite.

The square of the superpopulation expectation of the variance, $E^2V[t(s)|\underline{e}]$, may be written as

$$E^2V[t(s)|\underline{e}] = n^{-4}X^4(+)\{\mathop{S}_{i=1}^{N}s^4(i)P^2(i)[1-P(i)]^2$$

$$+\mathop{S\,S}_{i\neq j}^{N\ N}s^2(i)s^2(j)P(i)P(j)[1-P(i)][1-P(j)]\}.$$

Combining the above results for $EV^2[t(s)|\underline{e}]$ and $E^2V[t(s)|\underline{e}]$ to obtain the superpopulation variance of the finite population variance produces

$$n^4X^{-4}(+)VV[t(s)|\underline{e}] =$$

$$\mathop{S}_{i=1}^{N}\{E[e(i)/X(i)]^4-s^4(i)\}P^2(i)[1-P(i)]^2$$

$$+ 2\mathop{S\,S}_{i\neq j}^{N\ N}s^2(i)s^2(j)[P(ij)-P(i)P(j)]^2.$$

To minimize this variance by choice of P(ij), it is only necessary to consider those terms in the expanded form that are a function of the P(ij). This approach reduces to minimizing

$$\mathop{S\,S}_{i\neq j}^{N\ N}[P(i)P(j) - P(ij)]^2s^2(i)s^2(j),$$

or, in more compact notation

$$\mathop{S\,S}_{i\neq j}^{N\ N}W^2(ij)s^2(i)s^2(j).$$

If the $s^2(i)$ are constant over all elements, i, this criterion reduces to the one suggested by Jessen (ref. 7).

5. THE ALGEBRAIC SOLUTION FOR OPTIMAL VALUES OF P(ij)

The solution for the P(ij) must satisfy the set of constraints

$$\mathop{S}_{j\neq i}^{N}P(ij) = (n-1)P(i).$$

In terms of the W(ij), these constraints are equivalent to requiring that

$$\mathop{S}_{j\neq i}^{N}W(ij) = P(i)[1-P(i)].$$

The method of Lagrange multipliers may be used to obtain the constrained solution for the W(ij) and subsequently for the P(ij). The function to be minimized may be written as

$$F\{W(ij),K(i):i\neq j\} = \mathop{S\,S}_{i\neq j}^{N\ N}W^2(ij)s^2(i)s^2(j)$$

$$- 4\mathop{S}_{i=1}^{N}K(i)\{\mathop{S}_{j\neq i}^{N}W(ij)-P(i)[1-P(i)]\}.$$

In the above, the $K(i)$ are the Lagrange multipliers. Taking derivatives with respect to the $W(ij)$ and setting them equal to zero as a necessary condition for obtaining a minimum produces $N(N-1)/2$ independent equations of the form

$$W(ij) = s^{-2}(i)s^{-2}(j)\{K(i)+K(j)\}.$$

Taking derivatives with respect to the $K(i)$ and setting them equal to zero yields N additional equations of the form

$$\sum_{j\neq i}^{N} W(ij) = P(i)[1-P(i)].$$

The entire process yields a set of $N(N+1)/2$ simultaneous equations. The number of equations to be solved initially can be reduced by applying the second set of equations to the first set to obtain N equations in $K(i)$ only:

$$P(i)[1-P(i)]s^{2}(i) = K(i)\sum_{j\neq i}^{N} s^{-2}(j) + \sum_{j\neq i}^{N} K(j)s^{-2}(j)$$

$$= K(i)\left[\sum_{j=1}^{N} s^{-2}(j) - 2s^{-2}(i)\right] + \sum_{j=1}^{N} K(j)s^{-2}(j).$$

This set of equations may be represented in matrix notation as

$$\underline{c} = M\,\underline{k}.$$

The N x 1 vector, $\underline{c}$, has elements $c(i)$ defined as

$$c(i) = P(i)[1-P(i)]s^{2}(i).$$

The N x 1 vector $\underline{k}$ is the vector of unknown Lagrange multipliers, $K(i)$. The N x N matrix, M, may be represented as

$$M = D + \underline{1}\,\underline{b}'$$

where D is a diagonal matrix. The diagonal elements of the matrix D may be expressed as

$$d(ii) = S^{(1)} - 2s^{-2}(i).$$

The term $S^{(1)}$ is defined as the following sum:

$$S^{(1)} = \sum_{k=1}^{N} s^{-2}(k).$$

The vector $\underline{1}$ consists of N ones. The vector $\underline{b}$ has elements

$$b(i) = s^{-2}(i).$$

If the two sums $S^{(2)}$ and $S^{(3)}$ are defined as

$$S^{(2)} = \sum_{k=1}^{N} b(k)d^{-1}(kk) \text{ and}$$

$$S^{(3)} = \sum_{k=1}^{N} b(k)c(k)d^{-1}(kk),$$

and if $S^{(2)} \neq -1$, then the simultaneous equations in $k(i)$ have solutions given by

$$k(i) = d^{-1}(ii)[c(i)-(1+S^{(2)})^{-1}S^{(3)}].$$

Having obtained algebraic solutions for the $K(i)$, the algebraic solutions for the $W(ij)$ may be obtained by substituting the solved $K(i)$ values in the first $N(N-1)/2$ equations, namely as

$$W(ij) = s^{-2}(i)s^{-2}(j)\{K(i) + K(j)\}.$$

The $P(ij)$ may be obtained as

$$P(ij) = P(i)P(j)-W(ij).$$

To be valid solutions in the sense of yielding a probability sampling design, all of the $P(ij)$ must satisfy the conditions

$$0 \le P(ij) \le \text{minimum}\{P(i),P(j)\}.$$

These may be called the "weak constraints".

An even stronger set of constraints on the $P(ij)$ may be considered, namely,

$$CP(i)P(j) \le P(ij) \le P(i)P(j)$$

for some small $C > 0$. These may be called the "strong constraints".

6. ADJUSTMENT PROCEDURES

Three alternative methods for adjusting the $P(ij)$ when the algebraic solutions fall outside of prescribed boundary values may be applied:

(1) constrained minimization,
(2) compensating additive adjustments, and
(3) sampling unit redefinition.

All three adjustment methods may be handled as methods for adjusting the $W(ij)$ rather than the $P(ij)$. Only method 2, compensating additive adjustments, is discussed in this paper.

Method 2 may be applied without accuracy limitations of computer matrix inversion procedures. The method treats one $W(ij)$ at a time as follows. Suppose $W(ij)$ exceeds the prescribed upper bound, i.e.,

$$W(ij) - U(ij) = a$$

and $a > 0$. Corrective adjustments may be made in the $W(ij)$ without violating the N constraints

$$\sum_{j\neq i}^{N} W(ij) = P(i)[1-P(i)].$$

The recommended procedure is:

(1) adjust $W(ij)$ by $-a$,
(2) adjust $W(ik)$ by $+a/(N-2)$ for $k\neq i$ or j,
(3) adjust $W(jk)$ by $+a/(N-2)$ for $k\neq i$ or j, and
(4) adjust $W(km)$ by $-2a/[(N-2)(N-3)]$ for $k\neq i$ or j and $m\neq i$ or j.

If several $W(ij)$ must be adjusted, they can be handled individually in an iterative fashion. Since the adjustments for different $W(ij)$ may tend to work against each other in the iterative process, it is recommended that the iteration continue until sufficient numerical accuracy is obtained.

7. SOME COMMENTS ON THE CHOICE OF CONSTRAINTS

Requiring that

$$P(ij) \le P(i)P(j)$$

guarantees that all terms in the Yates-Grundy form of the variance will be non-negative and the Yates-Grundy variance estimator also will be non-negative for all possible samples.

The initial form of the strong constraints was discussed in section 5 and was stated as

$$CP(i)P(j) \le P(ij) \le P(i)P(j)$$

for all $P(ij)$ and some small $C > 0$. If one considers the left-hand part of the inequalities and sums both sides over j not equal i, the result is

$$CP(i)[n-P(i)] \leq (n-1)P(i)$$

for all i, or after factoring out $P(i)$ and dividing by $[n-P(i)]$,

$$C \leq (n-1)/[n-P(i)]$$

for all i. If a single value of C is to be used, it must be less than the minimum value over all i of

$$(n-1)/[n-P(i)].$$

Setting C less than or equal to $(n-1)/n$ satisfies this requirement for an upper bound of C.

If the strong constraints are applied with C set at $(n-1)/n$, some immediate effects on the finite population variance and variance estimates may be noted.

Stating the constraints in terms of the $W(ij)$ gives

$$C \leq W(ij) \leq P(i)P(j)/n$$

The right hand term may also be expressed as $nX(i)X(j)/X^2(+)$. The variance of the Horvitz-Thompson estimator for designs employing this constraint will then be bounded above by

$$[n/X(+)]^2 \mathop{S}^{N} \mathop{S}^{N}_{i \neq j} X(i)X(j)D^2(ij)$$

where

$$D(ij) = Y(i)/P(i) - Y(j)/P(j)$$

or equivalently

$$D(ij) = [X(+)/n][Y(i)/X(i) - Y(j)/X(j)].$$

This upper bound is exactly equivalent to the variance of the total estimates for the standard Hansen-Hurwitz probability replacement design. This particular form of the variance for probability replacement sampling is developed as an analogue to the nonreplacement form by Kendall and Stuart (ref. 11).

The same constraint on the $P(ij)$ may be expressed in terms of coefficients, $w(ij)$, of the Yates-Grundy variance estimator by first noting that

$$w(ij) = P(i)P(j)/P(ij)-1.$$

Then the effects of the constraints on the $w(ij)$ can be noted in the following steps:

$$(n-1)P(i)P(j)/n \leq P(ij) \leq P(i)P(j),$$

$$1/[P(i)P(j)] \leq 1/P(ij) \leq n/[(n-1)P(i)P(j)],$$

$$1 \leq P(i)P(j)/P(ij) \leq n/(n-1),$$

and

$$0 \leq w(ij) \leq 1/(n-1).$$

The variance estimators for designs employing this constraint will be bounded above for each sample by

$$(n-1)^{-1} \mathop{S}^{n} \mathop{S}^{n}_{i<j} D^2(ij).$$

This expression is exactly equal to the variance estimate for the Hansen-Hurwitz probability replacement design when applied to samples of distinct elements (ref. 11).

It may then be concluded that application of the strong constraints with C equal to $(n-1)/n$ will produce designs with variances and variance estimates that are never greater than those that would be obtained under probability replacement designs.

8. SOME EMPIRICAL COMPARISONS

Two general questions cannot be answered directly by examination of the theoretical development in the preceding sections. The first question is concerned with the sensitivity of the superpopulation variance criterion to poor or wrong initial assumptions about the error structure of the population to be sampled. The second question is concerned with the effects of the choice of constraints on the superpopulation variance criterion and on other statistical properties such as the stability of the variance estimator as measured by the superpopulation expectation of the finite population variance of the Yates-Grundy variance estimator. These two questions were studied empirically in a small simulation experiment.

Two simple population simulation models based on independent Bernoulli trials were utilized. Since moments up to the fourth order could be computed for these models, theoretical superpopulation results could be obtained as well as results of repeated simulation trials. Only the theoretical results are presented in this paper.

Populations of size 10 were considered with size measures, $X(i)$, being the numbers one through 10. The observed values, $Y(i)$, were generated by the simulation models with the expected squared error, $Ee^2(i)$, proportional to $X(i)$ for the first population and proportional to $X^2(i)$ for the second population.

Twelve designs were studied based on four different assumptions about the expected squared error term in the superpopulation model and three successively more stringent sets of constraints on the algebraic solutions.

The results comparing the superpopulation variances of the finite population variance are presented in table 1. Little variation in the criterion function occurs due to erroneous assumptions about the true error model. The imposition of strong constraints increases the obtainable minimum slightly and decreases the variation due to error model assumed.

Table 2 shows the behavior of the superpopulation expectations of the finite population variance of the Yates-Grundy variance estimator. Among the four model assumptions studied the assumption of squared error proportional to $X(i)$ yields the minimum value of the function for either model and for any of the three constraints. Imposition of the strong constraints can be seen to be very effective in controlling the stability of the variance estimator.

9. EXTENSIONS

The solutions for $P(ij)$ may be obtained for any sample size, n. However, higher order probabilities must be obtained to completely specify the sample design if the sample size is greater than two. An extension to sample size three has been obtained (ref. 12). The problem remains

TABLE 1 - THEORETICAL SUPERPOPULATION VARIANCES OF THE FINITE POPULATION VARIANCE

Assumed model for design purposes	Constraints applied to the P(ij)		
	$.01P(i)(j) \le P(ij) \le \mathrm{Min}\{P(i),P(j)\}$	$.01P(i)P(j) \le P(ij) \le P(i)P(j)$	$.50P(i)P(j) \le P(ij) \le P(i)P(j)$
		True model: $Ee^2(i) = .16X(i)$	
$Ee^2(i) = k$	288.68	288.68	288.26
$Ee^2(i) = kX(i)$	287.95	287.95	287.96
$Ee^2(i) = kX^2(i)$	292.81	289.17	288.14
$Ee^2(i) = kX^3(i)$	305.58	291.20	288.28
		True model: $Ee^2(i) = .16X^2(i)$	
$Ee^2(i) = k$	9,686.55	9,686.55	9,687.80
$Ee^2(i) = kX(i)$	9,692.67	9,692.67	9,690.86
$Ee^2(i) = kX^2(i)$	9,664.04	9,671.54	9,683.12
$Ee^2(i) = kX^3(i)$	9,728.40	9,689.40	9,685.03

to extend the procedure to samples of size four or greater.

8. LIST OF REFERENCES

[1] D. G. Horvitz and D. J. Thompson, "A Generalization of Sampling from a Finite Universe," Journal of the American Statistical Association, 47:663-685, 1952.

[2] Des Raj, "A Note on the Determination of Optimum Probabilities in Sampling Without Replacement," Journal of the American Statistical Association, 51:269-284, 1956.

[3] T. V. Hanurav, "Optimum Utilization of Auxiliary Information: πPS Sampling of Two Units from a Stratum," J. R. Statistical Society, B29:371-391, 1967.

[4] R. K. W. Brewer, "A Model of Systematic Sampling with Unequal Probabilities," Australian Journal of Statistics, 5:5-13, 1963.

[5] J. Durbin, "Design of Multi-stage Surveys for the Estimation of Sampling Errors," Applied Statistics 16:152-164, 1967.

[6] M. R. Sampford, "On Sampling Without Replacement with Unequal Probabilities of Selection," Biometrika 54:499-513, 1967.

[7] R. J. Jessen, "Some Methods of Probability Nonreplacement Sampling," Journal of the American Statistical Association 64:175-193, 1969.

[8] J. N. K. Rao and D. L. Bayless, "An Empirical Study of the Stabilities of Estimators and Variance Estimators in Unequal Probability Sampling of Two Units Per Stratum," Journal of the American Statistical Association 64:540-559, 1969.

[9] D. J. Dodds and J. G. Fryer, "Some Families of Selection Probabilities for Sampling With Probability Proportional to Size," J. R. Statistical Society 33(2):263-274, 1971.

[10] F. Yates and P. M. Grundy, "Selection Without Replacement from Within Strata and with Probability Proportional to Size," J. R. Statistical Society B15:253-261, 1953.

[11] Maurice G. Kendall and Alan Stuart, The Advanced Theory of Statistics, Volume 3, Design and Analysis, and Time Series. Charles Griffin and Company Limited, London, 1966.

[12] James R. Chromy, Pairwise Probabilities in Probability Nonreplacement Sampling, Ph.d. Thesis, North Carolina State University, Raleigh, North Carolina, 1974.

TABLE 2 - THEORETICAL SUPERPOPULATION EXPECTATIONS OF THE FINITE POPULATION VARIANCE OF THE YATES-GRUNDY VARIANCE ESTIMATOR

Assumed model for design purposes	Constraints applied to the P(ij)		
	$.01P(i)P(j) \le P(ij) \le \mathrm{Min}\{P(i),P(j)\}$	$.01P(i)P(j) \le P(ij) \le P(i)P(j)$	$.50P(i)P(j) \le P(ij) \le P(i)P(j)$
		True model: $Ee^2(i) = .16X(i)$	
$Ee^2(i) = k$	4,647	4,647	3,979
$Ee^2(i) = kX(i)$	4,014	4,014	3,944
$Ee^2(i) = kX^2(i)$	35,252	7,735	4,410
$Ee^2(i) = kX^3(i)$	393,893	27,701	4,468
		True model: $Ee^2(i) = .16X^2(i)$	
$Ee^2(i) = k$	72,003	72,003	66,085
$Ee^2(i) = kX(i)$	62,815	62,815	62,826
$Ee^2(i) = kX^2(i)$	170,775	81,915	65,037
$Ee^2(i) = kX^3(i)$	1,549,399	166,684	66,517