K.R.W. Brewer
c/o Commonwealth Schools Commission
P.O. Box 34
Woden, Canberra
A.C.T. 2606
Australia

Muhammad Hanif
Department of Statistics
El-Fateh University
Tripoli
Libya (S.P.L.A.J.)

AMS Subject Classification: 62D05

Library of Congress Cataloging in Publication Data
Brewer, K. R. W.
Monograph on sampling with unequal probabilities.
(Lecture notes in statistics; v. 15)
Bibliography: p.
Includes indexes.
1. Sampling (Statistics) 2. Estimation theory.
I. Hanif, Muhammad. II. Title. III. Series: Lecture notes in statistics (Springer-Verlag); v. 15.
QA276.6.B74 1982 519.5'2 82-19256

With 9 Illustrations

Printed and bound by R.R. Donnelley & Sons, Harrisonburg, VA.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-90807-2 Springer-Verlag New York Heidelberg Berlin
ISBN 3-540-90807-2 Springer-Verlag Berlin Heidelberg New York

PREFACE

Work for this monograph on sampling with unequal probabilities was started when Muhammad Hanif was a visitor to the then Commonwealth Bureau of Census and Statistics, Canberra, in 1969. It remained in abeyance until he again visited Canberra, this time the Australian National University's Survey Research Centre in 1978 as Visiting Fellow. The work was substantially completed when K.R.W. Brewer visited El-Fateh University during January 1980 as Visiting Professor. Finally, in 1982 the Bibliography was revised and corrected, and a number of references added which do not appear in the text. These are indicated by an asterisk (*).

The authors are indebted to Mr. E.K. Foreman and the sampling staff (past and present) at the Australian Bureau of Statistics for their help and encouragement and to Mrs Barbara Geary for her excellent mathematical typing.

Canberra K.R.W. Brewer

May 1982. Muhammad Hanif

CONTENTS

CHAPTER 1

AN INTRODUCTION TO SAMPLING WITH UNEQUAL PROBABILITIES

1.1 SOME BASIC CONCEPTS

Most survey work involves sampling from finite populations. There are two parts to any sampling strategy. First there is the selection procedure, the manner in which the sample units are to be selected from the finite population. Then there is the estimation procedure which prescribes how inferences are to be made from the sample to the population as a whole. These inferences may be either enumerative or analytical.

Enumerative inference seeks only to describe the particular finite population under study; analytical inference in some sense to explain it. Thus in dealing with a population of households we might attempt to enumerate the mean number of persons per household, the total number of persons in the population, the proportion of adults with tertiary education, the mean household income, and the ratio of total income to total number of persons (that is, per capita income). Enumerative inference typically concerns itself with such means, totals, proportions and ratios.

Viewing the same population analytically we might seek to regress household income on such variables as number of employed adults, educational level of household head, and dummy variables indicative of geographical location. Regression implies an explanatory model, and analytical inference consists of specifying what model is appropriate, estimating its parameters, and (at least in principle) checking to see whether the fitted model adequately describes the sample, and hence the population from which it was selected. Estimation of simple and multiple correlation coefficients also implies a regression model.

It is customary to distinguish between enumerative and analytical inference in terms of the complexity of the population characteristics being estimated. Estimating a mean is regarded as an enumerative problem; estimating a regression or correlation coefficient as an analytical one. But if the mean in

question is a parameter of a simple explanatory model, the inference is in fact analytical. Similarly if the regression or correlation coefficient is being used as a purely descriptive measure without any explanatory significance, the inference is enumerative. (One may question the usefulness of estimating such coefficients in these circumstances, but not the fact that such estimation is often carried out.)

It is therefore the use of a statistical model for explanatory purposes which separates analytical from enumerative inference. (Such models may be used, especially at the sample design stage, to ensure that a particular piece of enumerative inference can be carried out with reasonable efficiency, but unless that enumerative inference is robust against model breakdown, it will not receive general acceptance.)

Analytical and enumerative inferences therefore proceed along entirely different paths. For analytical inference the model used provides its own probability structure, in terms of which the inference can proceed. This is the same methodology as is used in almost every other area of statistics. For enumerative inference a quite different probability structure is generally used, which depends on the manner in which the sample is selected. This is the classical finite population sampling inference deriving from Bowley (1913) and developed by Neyman (1934).

The co-existence of these two methods of inference is a matter that impinges very sharply on samples drawn with unequal probability. As long as all samples are drawn with equal probabilities the two forms of inference are equivalent (apart from the finite population correction) or can be made equivalent. Thus, for instance, in a stratified sampling situation where the sampling fractions differ from stratum to stratum, the two inferences can be made equivalent by considering each stratum as a separate population. A self-weighting sample drawn in two stages can also be considered, from the standpoint of analytical inference, as though it were an experimental design problem with the block effects being random variables.

But when each unit in a population has its own individual probability of inclusion in sample, there is no simple way in which the two methods of inference can be made to coincide. For analytical purposes the probabilities of selection are irrelevant. For enumerative inference they are crucial. This monograph is basically concerned with the situation where the probabilities of selection are relevant, and therefore with enumerative inference. Population models will be called on for comparative purposes for help in optimal design, and to shed light on the estimation process; but they will not be central to the problem, and any estimation of their parameters will be a means to an end rather than an end in itself.

The use of unequal probabilities in sampling was first suggested by Hansen and Hurwitz (1943). Prior to that date there had been substantial developments in sampling theory and practice, but all these had been based on the assumption that the probabilities of selection within each stratum would be equal. Hansen and Hurwitz demonstrated, however, that the use of unequal selection probabilities within a

stratum frequently made for more efficient estimators of total than did equal probability sampling. They proposed a two-stage sampling scheme. The first-stage selection took place in independent draws. At each draw a single first-stage unit was selected with probabilities proportional to a size measure (the number of second-stage units within each first-stage unit). At the second stage, the same number of second-stage units was selected from each sample first-stage unit.

Because it was possible for the same first-stage unit to be selected more than once, this type of unequal probability sampling (*ups*) is generally known as *sampling with replacement*. Since, however, the independence of the draws is not a necessary condition for the units to have a non-zero probability of being selected more than once, another name first suggested by Hartley and Rao (1962) will be used instead. This is *multinomial sampling*, a term justified by the multinomial distribution of the number of units in the sample.

This scheme compared favourably with two other two-stage sampling schemes; these used unequal probabilities of selection at the first stage, and then took either a fixed number or a constant proportion of sub-sampling units from each selected first stage unit.

This first suggestion for the use of *ups* was thus already associated with the technique of multistage sampling and sampling with probability proportional to size (also known as *pps* or sometimes as πps sampling). Unequal probability sampling can, however, be used in single stage designs, and need not necessarily be with probability exactly proportional to size, although some kind of size measure is almost always used as a starting off point for assigning selection probabilities. A number of without replacement selection procedures result in probabilities of inclusion in sample which are only approximately proportional to size, and sometimes it is actually desirable to select with probability proportional to a power of size.

Practically all the published literature on *ups* to date has been concerned either with *sampling with replacement* (*multinomial sampling*) or *sampling without replacement*. It is these two cases which will be considered in the main portion of this Chapter. Chapters 2, 3 and 4 will be taken up with a detailed discussion of procedures for *sampling without replacement*. A full description of each procedure will be found in Chapter 2. Those procedures specifically designed for use with the Horvitz-Thompson (1952) estimator will be compared in Chapter 3, and other procedures in Chapter 4. In Chapter 5, *ups* is considered in the context of multistage sampling, and an algorithm is presented for estimating the components of variance for any sampling method that uses unbiased estimation. Finally, in Chapter 6, robust estimation is examined in the context of *ups*, and in particular the simultaneous optimization of the estimator and the selection probabilities in a large single stage sample.

1.2 NOTATION AND ABBREVIATIONS

The following is a summary of the principal notation to be used in this monograph. It is intended for reference purposes only, and further notation will also be introduced as required.

Capital letters and subscripts will be used for population values, and lower case for sampling values. Thus Y_I is the value of the Y-variate taken by the Ith population unit and y_i is the corresponding value taken by the ith sample unit. This departure from the usual practice is needed to avoid ambiguities in multistage sampling and in any case makes for greater clarity. The adoption of this notation is also useful for teaching purposes.

N, n, ... will denote the number of units in the population and sample respectively. Wherever the number of sample units is a random variable it will be denoted by n, and its expectation by ν.

Y, y, ... will denote values of the current or estimand variable.

X, x, ... will denote values of a benchmark variable used in ratio estimation.

Z, z, ... will denote measures of size.

P, p, ... will denote normed measures of size, that is, $P_I = Z_I/Z$ where $Z = \sum_{I=1}^{N} Z_I$.

π_I ... will denote the probability of inclusion in sample of the Ith unit and π_i will denote its sample value.

π_{IJ} ... will denote the joint probability of inclusion in sample of the Ith and Jth population units and π_{ij} will denote its sample value.

E ... is the sampling expectation operator (expectation over all possible samples).

V ... is the sampling variance operator, and v will denote a sample estimator of the corresponding variance.

The following abbreviations will also be used:

srswr ... simple random sampling with replacement.
srswor ... simple random sampling without replacement.

πpswor ... probability of inclusion in sample proportional to size without replacement.
ppswr ... probability of selection proportional to size with replacement (multinomial sampling).
ups ... unequal probability sampling.
upswor ... unequal probability sampling without replacement.
upswr ... unequal probability sampling with replacement.

The last three abbreviations are more general in their application than the others in that they do not require the probabilities of selection to be proportional to a given size measure. They therefore embrace such schemes as sampling with probability proportional to a function of size, and the special selection procedures devised by Raj, by Das, and by Rao, Hartley and Cochran, mentioned in Section 1.5 below.

1.3 MULTINOMIAL SAMPLING (*Sampling with Replacement*)

As already mentioned, the sampling procedure of Hansen and Hurwitz (1943) used sampling *ppswr*. One unit was selected at each of n draws. The probability of selection of the Ith population unit at any of these draws was $P_I = Z_I/Z$, where Z_I was the measure of size of that unit and $Z = \sum_{I=1}^{N} Z_I$ was the total measure of size of the N units in the population.

The Hansen-Hurwitz estimator y'_{HH} of the population total, $Y = \sum_{I=1}^{N} Y_I$ is

$$y'_{HH} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{P_i} . \tag{1.3.1}$$

The estimator is unbiased and its variance may be expressed variously as

$$V(y'_{HH}) = \frac{1}{n} \left(\sum_{I=1}^{N} \frac{Y_I^2}{P_I} - Y^2 \right) , \tag{1.3.2}$$

$$= \frac{1}{n} \sum_{I=1}^{N} P_I \left(\frac{Y_I}{P_I} - Y \right)^2 , \text{ or} \tag{1.3.3}$$

$$= \frac{1}{2n} \sum_{\substack{I,J=1 \\ J \neq I}}^{N} \sum P_I P_J \left(\frac{Y_I}{P_I} - \frac{Y_J}{P_J} \right)^2 . \tag{1.3.4}$$

An unbiased estimator of this variance is

$$v(y'_{HH}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{y_i}{p_i} - y'_{HH} \right)^2 , \tag{1.3.5}$$

$$= \frac{1}{2n^2(n-1)} \sum_{\substack{i,j=1 \\ j \neq i}}^{n} \sum \left(\frac{y_i}{p_i} - \frac{y_j}{p_j} \right)^2 . \tag{1.3.6}$$

This scheme of *sampling with replacement* is less efficient than certain alternative schemes which have since been proposed for sampling without replacement. In spite of this some people prefer to use *sampling with replacement* for the following reasons:

(i) Selection of the sample is simple.

(ii) The method can be used for any finite predetermined (but not necessarily distinct) number of units in sample.

(iii) The unbiased estimator of variance is simple.

(iv) It is also comparatively easy to obtain unbiased estimators of total variance and components of variance in multistage designs.

1.4 *πpswor* METHODS USING THE HORVITZ-THOMPSON ESTIMATOR

Six years after Hansen and Hurwitz's paper, Madow (1949) proposed the use of systematic sampling with unequal probabilities so as to avoid the possibility of units being selected more than once. This suggestion was followed up by a large number of alternative selection procedures commencing with that devised by Narain (1951), which will be considered in detail in Chapter 2.

Horvitz and Thompson (1952) produced a general theory of sampling with unequal probabilities without replacement based on the use of the estimator

$$y'_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} , \tag{1.4.1}$$

where y_i is the value of the ith sample unit. The important properties of the Horvitz-Thompson estimator are as follows:

(i) It is the only unbiased estimator of the class in which the same weight is attached to a particular population unit whenever it is selected (Horvitz and Thompson 1952).

(ii) It is admissible in the class of all homogeneous linear unbiased estimators of population total Y ; that is, there does not exist any member of that class which has a smaller variance than y'_{HT} (Roy and Chakravarti, 1960; Godambe, 1960).

(iii) If the Y_I are all exactly proportional to the corresponding π_I and the number of units in sample is fixed, the variance of y'_{HT} is zero. This is a property usually associated with ratio estimators and will be referred to as the *ratio estimator property*.

If Z_I values are known for all units in the population, and the Y_I are approximately proportional to the Z_I , the variance of y'_{HT} can be made small by setting the π_I proportional to the Z_I . This is one important reason why selection with probability proportional to size has assumed a central importance in *ups*.

The following is a summary of the general estimation theory for selection with probabilities to size without replacement (*πpswor*) based on the papers by Horvitz and Thompson (1952) and by Hanurav (1962a, 1967).

$$\sum_{I=1}^{N} \pi_I = E(n) = \nu , \tag{1.4.2}$$

$$\sum_{J \neq I}^{N} \pi_{IJ} = (\nu-1)\pi_I , \tag{1.4.3}$$

and

$$\sum_{\substack{I,J=1 \\ J \neq I}}^{N} \sum \pi_{IJ} = \nu(\nu-1) + V(n) . \tag{1.4.4}$$

The variance of the unbiased estimator y'_{HT} is:

$$V(y'_{HT}) = \sum_{I=1}^{N} \frac{1-\pi_I}{\pi_I} Y_I^2 + \sum_{\substack{I,J=1 \\ J \neq I}}^{N} \sum \frac{\pi_{IJ}-\pi_I\pi_J}{\pi_I\pi_J} Y_I Y_J . \tag{1.4.5}$$

The following alternative expression, which was derived independently by Sen (1953) and by Yates and Grundy (1953) (it will be attributed to Sen-Yates-Grundy), is valid only if the number of units in sample is fixed:

$$V(y'_{HT}) = \sum_{\substack{I,J=1 \\ J>I}}^{N} \sum (\pi_I\pi_J - \pi_{IJ}) \left(\frac{Y_I}{\pi_I} - \frac{Y_J}{\pi_J} \right)^2 . \tag{1.4.6}$$

The following estimator of $V(y'_{HT})$ is unbiased if $\pi_{IJ} \neq 0$ for all $J \neq I$, that is if all possible pairs of distinct population units have a non-zero probability of inclusion in sample. (Such estimators as these will be referred to as *conditionally unbiased*.)

$$v\left(y'_{HT}\right) = \sum_{i=1}^{n} \frac{1-\pi_i}{\pi_i^2} y_i^2 + \sum_{\substack{i,j=1 \\ j\neq i}}^{n} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}\pi_i\pi_j} y_i y_j \ . \qquad (1.4.7)$$

This estimator suffers from the disadvantage that it is not always zero when the variance is zero. The following alternative conditionally unbiased estimator was suggested by Sen (1953) and by Yates and Grundy (1953) for use when the number of sample units is fixed:

$$v_{SYG}\left(y'_{HT}\right) = \sum_{\substack{i,j=1 \\ j>i}}^{n} \frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_i}{\pi_j}\right)^2 \ . \qquad (1.4.8)$$

Both (1.4.7) and (1.4.8) can assume negative values, but (1.4.8) rarely seems to do so in practice. It has also performed much better than (1.4.7) in a number of empirical comparisons, commencing with that in Yates and Grundy's (1953) paper. For $n = 2$ it is the only possible non-negative variance estimator (Vijayan, 1975).

Sen (1953) also compared the efficiency of (1.4.7) and (1.4.8) taking a population of five units and selecting all possible samples of two units under the following two schemes:

(a) The first unit is drawn with *pps* and the second unit with *pps* without replacement.

(b) The first unit is drawn with *pps* and the second unit with equal probabilities without replacement.

He demonstrated that the expression (1.4.8) took positive values for all these samples, but that (1.4.7) was negative for some of the pairs. He further showed that $\pi_{ij} > \pi_i\pi_j$; for all $i \neq j$ for $n = 2$, and hence that when selection is made with π*pswor* using the Horvitz-Thompson estimator, (1.4.8) is always positive.

Raj (1956a) proved further that the expression (1.4.8) was always positive under the schemes (a) and (b) above.

Rao (1963a) also proved under two other well-known procedures for *upswor* that (1.4.8) was always positive. These procedures (that of Midzuno - reported by Horvitz and Thompson (1952) - and the Yates and Grundy (1953) rejective procedure) will be described in Chapter 2.

Rao and Singh (1973) used Brewer's (1963) π*pswor* selection procedure to compare (1.4.7) and (1.4.8) for the case $n = 2$, employing a wide variety of populations. Their empirical evidence also indicates that (1.4.8) is more stable than (1.4.7). A similar result was obtained by Lanke (1974) using Hajek's "Method I" (1964b).

1.5 *upswor* METHODS USING OTHER ESTIMATORS

Das (1951) suggested the following strategy. A sample of n units is selected. At each draw the selection is made among those units not already selected with probabilities proportional to size. The estimator t'_{mean} of the population total Y is the arithmetic mean of the n unbiased estimators:

$$\left.\begin{aligned} t'_1 &= \frac{y_1}{p_1} \ , \\ t'_2 &= \frac{1-p_1}{p_1 p_2} \frac{y_2}{N-1} \ , \\ &\cdots \\ t'_r &= \prod_{i=1}^{r-1} \left(1 - \sum_{j=1}^{i} p_j\right) y_r \div \prod_{i=1}^{r} p_i \prod_{i=1}^{r-1} (N-i) \ , \quad r = 1, 2, \ldots, n \ . \end{aligned}\right\} \qquad (1.5.1)$$

Raj (1956a) modified Das's strategy as follows. The estimator t_{mean} of the population total Y is the arithmetic mean of the following set of unbiased and uncorrelated estimators:

$$\left.\begin{aligned} t_1 &= \frac{y_1}{p_1} \ , \\ t_2 &= y_1 + \frac{y_2}{p_2} \left(1-p_1\right) \ , \\ &\cdots \\ t_k &= y_1 + y_2 + \ldots + y_{n-1} + \frac{y_n}{p_n} \left(1 - p_1 - p_2 - \ldots - p_{n-1}\right) \ . \end{aligned}\right\} \qquad (1.5.2)$$

(The estimator t_1 thus depends only on the first unit selected, t_2 on the first two units, and so on.)

Murthy (1957) considered all possible permutations of the sample order which led to different values of t_{mean} . He proved that the symmetrized estimator, which gave weight to each possible value of t_{mean} in proportion to the *a priori* probability of the observed sample units being selected in that order, had a smaller variance than t_{mean} .

Rao, Hartley and Cochran's (1962) sampling strategy is as follows. The population units are divided randomly into n groups of N_J units, $J = 1, 2, 3, \ldots, n$, where the N_J are predetermined. One unit is selected from

each group, the probabilities of selection being the normed measures of size within the group. Their estimation procedure is to form the Horvitz-Thompson estimator for each group separately and to add these over the groups.

1.6 LIST OF PROCEDURES FOR SAMPLING WITHOUT REPLACEMENT

The following is a list of 50 *πpswor* procedures. A mnemonic title is suggested for each, and a basic reference is given. Descriptions of these procedures will be given in Chapter 2.

Procedure 1: Ordered Systematic Procedure; Madow (1949).
Procedure 2: Random Systematic Procedure; Goodman and Kish (1950).
Procedure 3: Grundy's Systematic Procedure; Grundy (1954).
Procedure 4: Yates-Grundy Draw-by-Draw Procedure; Yates and Grundy (1953).
Procedure 5: Yates-Grundy Rejective Procedure; Yates and Grundy (1953).
Procedure 6: Midzuno's Procedure; reported by Horvitz and Thompson (1952).
Procedure 7: Narain's Procedure; Narain (1951).
Procedure 8: Brewer's Procedure; Brewer (1963, 1975).
Procedure 9: Durbin's "Method I"; Durbin (1967).
Procedure 10: Durbin's "Grouped Method"; Durbin (1967).
Procedure 11: Rao-Sampford Procedure; Rao (1965), Sampford (1967).
Procedure 12: Durbin-Sampford Procedure; Sampford (1967).
Procedure 13: Fellegi's Procedure; Fellegi (1963).
Procedure 14: Carroll-Hartley Rejective Procedure; Carroll and Hartley (1964).
Procedure 15: Carroll-Hartley Draw-by-Draw Procedure; Carroll and Hartley (1964).
Procedure 16: Carroll-Hartley Whole Sample Procedure; Carroll and Hartley (1964).
Procedure 17: Durbin-Hanurav Procedure; Durbin (1953b); Hanurav (1966, 1967).
Procedure 18: Hanurav's Scheme B-A'; Hanurav (1967).
Procedure 19: Hanurav-Vijayan Procedure; Hanurav (1967); Vijayan (1968).
Procedure 20: Raj's Variance Minimization Procedure; Raj (1956b).
Procedure 21: Hanurav's Simple Junctional Procedure; Hanurav (1962a).
Procedure 22: Hanurav's Modified Junctional Procedure; Hanurav (1962a).
Procedure 23: Hanurav's Double Junctional Procedure; Hanurav (1962a).

Procedure 24: Hanurav's Sequential Procedure; Hanurav (1962a).
Procedure 25: Rao-Hartley-Cochran Procedure; Rao, Hartley and Cochran (1962).
Procedure 26: Stevens' Procedure; Stevens (1958).
Procedure 27: Poisson Sampling; Hajek (1964b).
Procedure 28: Hajek's "Method I"; Hajek (1964b).
Procedure 29: Hajek's "Method II"; Hajek (1964b).
Procedure 30: Hajek's "Method III"; Hajek (1964b).
Procedure 31: Hajek's "Method IV"; Hajek (1964b).
Procedure 32: Deming's Systematic Procedure; Deming (1960).
Procedure 33: Variance Estimator Optimization Procedure; Brewer and Hanif (1969a).
Procedure 34: Jessen's "Method 1"; Jessen (1969).
Procedure 35: Jessen's "Method 2"; Jessen (1969).
Procedure 36: Jessen's "Method 3"; Jessen (1969).
Procedure 37: Jessen's "Method 4"; Jessen (1969).
Procedure 38: Modified Poisson Sampling; Ogus and Clark (1971).
Procedure 39: Collocated Sampling; Brewer, Early and Hanif (1980).
Procedure 40: Das-Mohanty Procedure; Das and Mohanty (1973).
Procedure 41: Mukhopadhyay's Procedure; Mukhopadhyay (1972).
Procedure 42: Sinha's Extension Procedure; Sinha (1973).
Procedure 43: Sinha's Reduction Procedure; Sinha (1973).
Procedure 44: Chaudhuri's Procedure; Chaudhuri (1976).
Procedure 45: Lahiri's Procedure; Lahiri (1951).
Procedure 46: Ikeda-Midzuno Procedure; Midzuno (1952).
Procedure 47: Fuller's "Scheme B"; Fuller (1971).
Procedure 48: Singh's Procedure; Singh (1978).
Procedure 49: Choudhry's Procedure; Choudhry (1979).
Procedure 50: Chromy's Procedure; Chromy (1979).

1.7 CLASSIFICATION OF PROCEDURES

The fifty procedures listed above may be classified in a number of ways. This section deals with three of the most useful and instructive classifications.

1.7.1 Classification by Manner of Selection

The classification by manner of selection set out below is based on that of Carroll and Hartley (1964), which although not entirely unambiguous is nevertheless useful for expository purposes.

(i) Draw-by-Draw Procedures

At each successive draw one unit is selected, usually from among those population units not previously selected. Probabilities of selection are defined for each draw and (since the selection is without replacement) always depend on which units are already selected. If the probabilities of selection at a given draw are (apart from a normalizing factor) independent of which units were selected at previous draws, they are sometimes referred to as *working probabilities*. The draw-by-draw procedures listed above are Procedures 4, 6, 7, 8, 9, 10, 12, 13, 15, 18, 19, 21, 22, 23, 24, 25, 26, 41, 44, 46, 47, 49 and 50.

(ii) Systematic Procedures

Systematic selection involves an ordering of the population and the cumulation of inclusion probabilities. The order of units may or may not be random. A random number r $(0 < r \leq 1)$ is chosen and the n units selected are those whose cumulated values of π_I (the desired probability of inclusion) are the smallest equal to or greater than each of r, $r+1$, $r+2$, ..., $r+n-1$. The systematic procedures listed above are Procedures 1, 2, 3, 32 and 48.

(iii) Rejective Procedures

The term *rejective* has been employed by Hajek (1964b) and is somewhat wider in its connotation than the term *mass draw* used by Carroll and Hartley (1964). Rejective procedures resemble draw-by-draw procedures in that only a single unit is selected at each of n successive draws. They differ from ordinary draw-by-draw procedures in that the selection at a given draw may give rise to the selection of an already selected unit, in which case the partial sample is abandoned and the selection recommenced. The rejective procedures listed above are Procedures 5, 11, 14, 17, 28, 29, 30 and 31.

(iv) Whole Sample Procedures

In these procedures the units are not individually drawn: a probability is specified for each possible sample of n distinct units and one selection using these probabilities selects the whole sample. The whole sample procedures listed above are Procedures 16, 20, 33, 34, 35, 36, 37, 40 and 45.

(v) Other Selection Procedures

Other selection procedures not listed in the above four categories are as follows:

Procedure 27: Poisson Sampling.
Procedure 38: Modified Poisson Sampling.
Procedure 39: Collocated Sampling.
Procedure 42: Sinha's Extension Procedure.
Procedure 43: Sinha's Reduction Procedure.

1.7.2 Classification by Equivalence Class

Two procedures belong to the same equivalence class when the joint probabilities of inclusion of all possible combinations of units are identical. It is obvious that each systematic, draw-by-draw and rejective procedure has an equivalent whole sample procedure. Godambe (1955) pointed out that any whole sample procedure also has a draw-by-draw equivalent. Hence it is possible, for a number of the procedures described in this monograph, to devise different selection procedures in the same equivalence class in a straightforward fashion.

Procedures 9, 11 and 12 belong to an equivalence class possessing the characteristic that the joint inclusion probabilities π_{ij} can be stated explicitly in analytic form, thus making the variance formulae comparatively simple. This will be referred to as Equivalence Class A. Procedure 8 belongs to this class for $n = 2$.

Procedures 14, 15 and 16 belong to a second equivalence class. Since they include the *symmetric mass draw* procedures, they may be designated as *symmetric* procedures, or Equivalence Class B. For the case $n = 2$, Procedures 13 and 49 also belong to this equivalence class.

1.7.3 Classification by Type of Estimator Appropriate

The first estimator suggested for use with sampling with unequal probabilities without replacement was that of Horvitz and Thompson (1.4.1). Any selection procedure may be used with the Horvitz-Thompson estimator, but it is only those for which the sample number is fixed and the probabilities of inclusion in sample can be made exactly proportional to an already known measure of size for which this estimator has the *ratio estimator property* (see Section 1.4). A further important property which applies under the same conditions is that the value of the Sen-Yates-Grundy (1953) variance estimator (1.4.8) is then also zero. Procedures possessing this property exactly are all those listed except Procedures 4, 5, 25-31, 38-39 and 45-46. Most of these exceptions were initially put forward as approximations to an ideal procedure which would confer the ratio estimator property on the Horvitz-Thompson estimator. Procedure 25 however was suggested explicitly for use with a special estimator. Certain other special estimators, notably t'_{mean}, t_{mean} and t_{symm} have since been suggested for use with Procedure 4 by Das (1951), Raj (1956a) and Murthy (1957). The special estimators t_{mean} and t_{symm}, when used in conjunction with Procedure 4 also possess the ratio estimator property.

All published selection procedures may therefore be classified into three groups:

(i) Those for which the Horvitz and Thompson estimator possesses the ratio estimator property exactly. These are Procedures 1-3, 6-24, 32-37, 40-44 and 47-50.

(ii) Those for which no estimator possesses this property exactly but for which the Horvitz and Thompson estimator possesses it approximately. These are Procedures

5, 26 and 28-31.

(iii) Those for which some other estimator or estimators possess this property.

Procedure 4: Yates-Grundy Draw-by-Draw Procedure - with the estimator t_{mean} suggested by Raj (1956a) or the estimator t_{symm} suggested by Murthy (1957).

Procedure 25: Rao-Hartley-Cochran Procedure - with RHC estimator.

Procedure 27: Poisson Sampling - with a ratio estimator (Brewer, Early and Joyce 1972).

Procedure 38: Modified Poisson Sampling - with a ratio estimator (Brewer, Early and Hanif 1980).

Procedure 39: Collocated Sampling - with a ratio estimator (Brewer, Early and Hanif 1980).

Procedure 45: Lahiri's Procedure - with the conventional ratio estimator.

Procedure 46: Ikeda-Midzuno Procedure - with the conventional ratio estimator.

1.8 SOME ASYMPTOTIC VARIANCE FORMULAE FOR *πpswor*

The variance (1.4.5) of the Horvitz-Thompson estimator involves the quantities π_{IJ}. Considerable difficulties are involved in the determination of these quantities for most of the procedures listed in Section 1.6. Hartley and Rao (1962) obtained an approximate solution to this problem for the Random Systematic Procedure with the help of an asymptotic theory which assumed that N was large and that $n/N \to 0$ as $N \to \infty$. They obtained an expression for the asymptotic value of π_{IJ} under these assumptions and substituted this in expression (1.4.5) to obtain the following asymptotic variance formula for any n :

$$V(y'_{HT}) \doteq \sum_{I=1}^{N} \pi_I \left(1 - \frac{n-1}{n}\pi_I\right)\left(\frac{Y_I}{\pi_I} - \frac{Y}{n}\right)^2 - \frac{n-1}{n^2}\sum_{I=1}^{N}\left(2\pi_I^3 - \frac{\pi_I^2}{2}\sum_{J=1}^{N}\pi_J^2\right)\left(\frac{Y_I}{\pi_I} - \frac{Y}{n}\right)^2 + \frac{2(n-1)}{n^3}\left(\sum_{I=1}^{N}\pi_I Y_I - \frac{Y}{n}\sum_{J=1}^{N}\pi_J^2\right)^2 . \quad (1.8.1)$$

This formula is correct to order N^0 .

Rao (1963a) showed that with $n = 2$ the variances of the Horvitz-Thompson estimator for three selection procedures were given asymptotically to order N^0 by

$$V(y'_{HT}) \doteq \sum_{I=1}^{N} \pi_I \left(1 - \frac{\pi_I}{2}\right)\left(\frac{Y_I}{\pi_I} - \frac{Y}{2}\right)^2 - \tfrac{1}{2}\sum_{I=1}^{N}\left(\pi_I^3 - \frac{\pi_I^2}{4}\sum_{J=1}^{N}\pi_J^2\right)\left(\frac{Y_I}{\pi_I} - \frac{Y}{2}\right)^2 + \lambda\left(\sum_{I=1}^{N}\pi_I Y_I - \frac{Y}{2}\sum_{J=1}^{N}Y_J^2\right)^2 . \quad (1.8.2)$$

The value of λ in this formula is 3/32 for Narain's Procedure, 1/8 for the Carroll-Hartley Rejective Procedure and 1/4 for the Random Systematic Procedure. Rao (1965) later showed that $\lambda = 0$ for Brewer's Procedure. Since Fellegi's Procedure and the three Carroll-Hartley Procedures are in the same equivalence class, the value of λ is 1/8 for all these Procedures. Similarly, since the Rao-Sampford Procedure, Durbin's "Method I" and Brewer's Procedure are in the same equivalence class, $\lambda = 0$ for all these procedures.

All the procedures mentioned so far in this Section have the same variance formula to order N^1 for $n = 2$, namely:

$$V(y'_{HT}) \doteq \sum_{I=1}^{N} \pi_I \left(1 - \frac{\pi_I}{2}\right)\left(\frac{Y_I}{\pi_I} - \frac{Y}{2}\right)^2 . \quad (1.8.3)$$

From (1.8.1) it follows that for the Random Systematic Procedure, to order N^1 ,

$$V(y'_{HT}) \doteq \sum_{I=1}^{N} \pi_I \left(1 - \frac{n-1}{n}\pi_I\right)\left(\frac{Y_I}{\pi_I} - \frac{Y}{n}\right)^2 . \quad (1.8.4)$$

Equation (1.8.4) was also shown by Rao (1963b) to be asymptotically valid for Narain's Procedure and the Carroll-Hartley Rejective Procedure.

There are almost certainly other *πpswor* procedures for which (1.8.4) is also asymptotically valid, but it seems necessary to derive it for each one separately. It is shown in this Section that under the assumption of a linear stochastic model, (1.8.4) is asymptotically valid for all *πpswor* procedures (Hanif, 1974).

The model is the same as that employed by Cochran (1953), which appears to have been originated by Smith (1938). The same linear model has been employed by several other authors, for example, Rao (1966), T.J. Rao (1967), Hanurav (1967), Vijayan (1967), Foreman and Brewer (1971), Hanif and Ahmad (1977), Cassel, Särndal and Wretman (1977).

The assumption is that the observed population can be treated as a sample of one from an infinite set of hypothetical populations generated by a stochastic model. In this particular application of the model it is assumed that the probabilities of inclusion in sample π_I are exactly proportional to size and constant from population to population. The model specification then is

$$\left.\begin{aligned} Y_I &= \beta Z_I + \epsilon_I \ , \\ E^*(\epsilon_I) &= 0 \ , \quad E^*(\epsilon_I \epsilon_J) = \begin{cases} \sigma_I^2 \ , & J = I \\ 0 \ , & \text{otherwise} \end{cases} \\ \sigma_I^2 &= \sigma^2 Z_I^{2\gamma} \quad \text{where} \quad \tfrac{1}{2} \le \gamma \le 1 \ , \end{aligned}\right\} \tag{1.8.5}$$

where β, σ^2 are constants and E^* denotes the expectation over all possible hypothetical populations. The model based variance of y' regarded as an estimator of Y will be written as

$$V^*(y'_{HT}) = E^*(y'_{HT}-Y)^2 \ . \tag{1.8.6}$$

The design expectation of expression (1.8.6) when $\pi_I \propto Z_I$ is

$$EE^*(y'_{HT}-Y)^2 = \sum_{I=1}^{N} \sigma_I^2 \left(\frac{1}{\pi_I} - 1\right) \ ; \tag{1.8.7}$$

but the same expression may be obtained asymptotically from the model expectation of the right side of the expression (1.8.4). This model expectation is

$$E^*\left[\sum_{I=1}^{N} \pi_I \left(1 - \frac{n-1}{n}\pi_I\right)\left(\frac{Y_I}{\pi_I} - \frac{Y}{n}\right)^2\right]$$

$$= \sum_{I=1}^{N} \sigma_I^2 \left(\frac{1}{\pi_I} - 1\right) + \frac{n-1}{n^2}\left(2\sum_{I=1}^{N}\sigma_I^2 \pi_I - \sum_{I=1}^{N}\frac{\pi_I^2}{n} - \sum_{I=1}^{N}\sigma_I^2\right) \ . \tag{1.8.8}$$

Now the second term of the expression (1.8.8) is of order N^0 while the leading term contains only expressions of order N^2 and N^1. Hence asymptotically only the leading term is left, which is the same as (1.8.7). Hence under model (1.8.5) the asymptotic variance formula (1.8.4) is valid for all *πpswor* procedures.

1.9 A GENERAL THEORY OF ESTIMATORS POSSESSING THE RATIO ESTIMATOR PROPERTY

Vijayan (1975), Rao and Vijayan (1977), and Rao (1979) have presented some general results which enable the mean square error of any linear estimator of total possessing the ratio estimator property in a straightforward fashion, and also exhibit the necessary form of any non-negative quadratic unbiased estimators of that mean square error. These results are essentially contained in the following theorem from Rao (1979).

Let the general linear estimator of the population total Y, based on a sample

s with associated probability of selection $p(s)$ be written

$$\hat{Y} = \sum_{I=1}^{N} d_{Is} Y_I \ , \tag{1.9.1}$$

where the weights d_{Is} do not depend on the Y_I, and $d_{Is} = 0$ if $I \notin s$. Suppose $\hat{Y}$ is such that $MSE(\hat{Y})$ becomes zero when $Y_I = cZ_I$ where $c \neq 0$ is an arbitrary constant (the ratio estimator property). Then

(a) $MSE(\hat{Y})$ reduces to

$$MSE(\hat{Y}) = -\sum_{\substack{I,J=1 \\ J>I}}^{N} d_{IJ} Z_I Z_J (R_I - R_J)^2 \ , \tag{1.9.2}$$

where $R_I = Y_I/Z_I$ and

$$d_{IJ} = E(d_{Is}-1)(d_{Js}-1) \ ; \tag{1.9.3}$$

(b) a non-negative quadratic unbiased estimator of $MSE(\hat{Y})$ is necessarily of the form

$$mse(\hat{Y}) = -\sum_{\substack{I,J=1 \\ J>I}}^{N} d_{IJ}(s) Z_I Z_J (R_I - R_J)^2 \ , \tag{1.9.4}$$

where the coefficients $d_{IJ}(s)$ do not depend on the Y_I, $d_{IJ}(s) = 0$ if s does not contain both units I and J and the following unbiasedness condition is satisfied:

$$\sum_{s \ni I,J} p(s) d_{IJ}(s) = d_{IJ} \ , \quad J > I \ . \tag{1.9.5}$$

The proof of this theorem depends on the matrix lemma that if $B = (b_{IJ})$ is a positive semidefinite matrix and $\sum_{I,J=1}^{N} b_{IJ} = 0$, then $\sum_{J=1}^{N} b_{IJ} = 0$ for all I.

(a) Using (1.9.3) we can write

$$MSE(\hat{Y}) = E(\hat{Y}-Y)^2 = \sum_{I,J=1}^{N} d_{IJ} Z_I Z_J R_I R_J \ .$$

$$\equiv \sum_{I=1}^{N} d_{II}^2 Z_I^2 R_I^2 + \sum_{\substack{I,J=1 \\ J \neq I}}^{N} d_{IJ} Z_I Z_J R_I R_J \ . \tag{1.9.6}$$

Since $MSE(\hat{Y}) = 0$ when all $R_I = c$, we get $\sum_{I,J=1}^{N} d_{IJ} Z_I Z_J = 0$ and by the

lemma $\sum_{J=1}^{N} d_{IJ}Z_IZ_J = 0$ for each I, that is, $d_{II}Z_I^2 = -\sum_{J\neq I}^{N} d_{IJ}Z_IZ_J$. Substituting this result in (1.9.6),

$$\mathrm{MSE}(\hat{Y}) = -\sum_{\substack{I,J=1\\J\neq I}}^{N}\sum d_{IJ}Z_IZ_J\left(R_I^2-R_IR_J\right)$$

$$\equiv -\sum_{\substack{I,J=1\\J>I}}^{N}\sum d_{IJ}Z_IZ_J\left(R_I-R_J\right)^2 . \qquad (1.9.2)$$

(b) Any unbiased non-negative quadratic estimator may be written

$$\mathrm{mse}(\hat{Y}) = \sum_{I,J=1}^{N}\sum d_{IJ}(s)Y_IY_J \qquad (1.9.7)$$

$$\equiv -\sum_{\substack{I,J=1\\J>I}}^{N}\sum d_{IJ}(s)Z_IZ_J\left(R_I-R_J\right)^2 + \sum_{I,J=1}^{N}\sum d_{IJ}(s)Z_IZ_JR_I^2 . \qquad (1.9.8)$$

By unbiasedness, non-negativity, and the ratio estimator property, $\sum_{I,J=1}^{N}\sum d_{IJ}(S)Z_IZ_J$ must be zero when all the $R_I = c$. But the $d_{IJ}(S)$ are independent of the Y_I, hence it must always be zero. Hence, by the lemma $\sum_{J=1}^{N} d_{IJ}(S)Z_IZ_J = 0$ for all I, and the second term on the right hand side of (1.9.8) disappears, leaving expression (1.9.4). //

Infinitely many choices for $d_{IJ}(s)$ satisfying (1.9.5) are possible. Rao (1979) makes two suggestions. The first is

$$d_{IJ}(s) = d_{IJ}/\pi_{IJ} \qquad (1.9.9)$$

and the second, valid for unbiased $\hat{Y}$ only, is

$$d_{IJ}(s) = d_{Is}d_{Js} - f_{IJ}(s)/p(s) \qquad (1.9.10)$$

where $f_{IJ}(s)$ is any choice satisfying $\sum_{s\ni I,J} f_{IJ}(s) = 1$, $J > I$. The second has the advantage that the d_{IJ} need not be calculated.

For the important special case where the sample is constrained to consist of two distinct units (I and J) the only possible non-negative unbiased estimator of $\mathrm{MSE}(\hat{Y})$ is

$$\mathrm{mse}(\hat{Y}) = -\frac{d_{IJ}}{\pi_{IJ}}Z_IZ_J\left(R_I-R_J\right)^2 . \qquad (1.9.11)$$

For the Hansen-Hurwitz estimator used with *ppswr* the choice

$$d_{IJ}(s) = -\frac{1}{n}\frac{t_It_J}{n(n-1)P_IP_J} , \qquad (1.9.12)$$

where t_I is the number of times the Ith unit appears in sample, and $P_I = Z_I/Z$, leads to the usual variance estimator,

$$v(y'_{HH}) = \frac{1}{n(n-1)}\sum_{I\in s} t_I\left(\frac{Y_I}{P_I} - y'_{HH}\right)^2 . \qquad (1.9.13)$$

Alternatively (1.9.9) leads to Rao's new non-negative variance estimator

$$v_1(y'_{HH}) = \frac{1}{n}\sum_{\substack{I,J\in s\\J>I}}\sum\frac{P_IP_J}{\pi_{IJ}}\left(\frac{Y_I}{P_I} - \frac{Y_J}{P_J}\right)^2 . \qquad (1.9.14)$$

For the Horvitz-Thompson estimator used with fixed sample size *πpswor* (1.9.9) leads to the Sen-Yates-Grundy variance estimator (1.4.8) which for $n = 2$ is the only possible non-negative unbiased variance estimator. But (1.9.10) with $f_{IJ} = M_2^{-1}$ where $M_2 = \binom{N-2}{n-2}$, gives a new variance estimator for $n > 2$, that is

$$v_2(y'_{HT}) = \sum_{\substack{I,J\in s\\J>I}}\sum\frac{\pi_I\pi_J-p(s)M_2}{p(s)M_2}\left(\frac{Y_I}{\pi_I} - \frac{Y_J}{\pi_J}\right)^2 . \qquad (1.9.15)$$

Further, (1.9.10) with $f_{IJ}(s) = p(s|I, J)$ (the conditional probability of getting s given that I and J were selected in the first two draws) gives another new estimator for $n > 2$, that is

$$v_3(y'_{HT}) = \sum_{\substack{I,J\in s\\J>I}}\sum\frac{\pi_I\pi_Jp(s|I,J)-p(s)}{p(s)}\left(\frac{Y_I}{\pi_I} - \frac{Y_J}{\pi_J}\right)^2 . \qquad (1.9.16)$$

The properties of (1.9.15) and (1.9.16) remain to be investigated, but (1.9.15) in particular appears to offer some gain in computational simplicity.

Rao (1969) also applied the above theory to Murthy's estimator and the Rao-Hartley-Cochran estimator, but without producing any new variance estimator of practical interest. The work of Rao and Vijayan (1977), in producing new variance estimators for the unbiased ratio estimator used with selection probabilities proportional to aggregative size, is considered in Chapter 4.

CHAPTER 2

DESCRIPTIONS OF PROCEDURES FOR SAMPLING WITH UNEQUAL PROBABILITIES WITHOUT REPLACEMENT

2.1 INTRODUCTION

In Chapter 1, 50 *upswor* procedures were listed. In this Chapter these selection procedures will be described in detail. The descriptions will be aimed at indicating the relationships between selection procedures and at showing how the disadvantages of some methods have led to the suggestion of others. The following descriptive abbreviations will be used.

strmps: probability of inclusion strictly proportional to size,

strwor: strictly without replacement,

n fixed: number of units in sample fixed,

syst: systematic,

d by d: draw by draw,

ws: whole sample,

ord: ordered,

unord: unordered.

In addition to these purely descriptive abbreviations certain disadvantages will be indicated as follows:

inexact: fails to satisfy at least one of the three descriptions *strmps*, *strwor* and *n fixed* above,

$n = 2$ only: limited to two sample units per stratum,

b est var: estimator of variance generally biased,

j p enum: calculation of joint probabilities of inclusion in sample involves enumeration of all possible selections, or at least a large number of them,

j p iter: calculation of joint probabilities of inclusion in sample involves iteration on computer,

not gen app: not generally applicable,

non-rotg: non-rotating.

The last two of these disadvantages require some explanation.

(i) Not Generally Applicable

Since the probability of inclusion is proportional to size and no probability of inclusion can be greater than unity, the theoretical limit to the size of individual units is Z/n. If the procedures break down before this theoretical limit on maximum size of unit is reached, it will be described as *not gen app*.

(ii) Non-Rotating

In large scale field surveys it is often desirable to be able to rotate the sample, that is, to drop a portion of the sample and replace it by another at predetermined intervals. The principal reason for wishing to do this is to avoid the kinds of response bias and non-representativeness which can result from being in sample on a number of occasions; phenomena known generally as *sample fatigue*. Rotation will be considered in greater detail in Chapter 3. Meanwhile it should be noted that certain procedures make specific allowance for rotation; that others can be used in rotating samples by selecting initially more units than are required immediately and rotating the excess into the sample as required; and that the remainder (including all those limited to $n = 2$) can only be rotated as whole stratum at a time. Both the latter categories will be described as *non-rotg*.

2.2 DESCRIPTIONS OF SELECTION PROCEDURES

A description of each of the selection procedures listed in Chapter 1 follows.

The format of these descriptions will be

Procedure number and mnemonic title,
Descriptive abbreviations and disadvantages,
Principal references,
Prose description of selection procedure,
Comments.

PROCEDURE 1: Ordered Systematic Procedure

Strmps, *strwor*, *n fixed*, *syst*, *ord*, *b est var*, *j p enum*

Madow (1949), Hartley (1966), Cassel *et al* (1977, p. 17).

Arrange the population units in any convenient order. Cumulate the measures of size

down this order. Divide the total measure of size Z by the required number of units in sample, n, to obtain the *skip interval* Z/n. Choose a *random start*, that is, a random number greater than or equal to zero and less than the skip interval. The first unit selected is that for which the cumulated size measure is the smallest greater than or equal to the random start. The second unit is that for which the cumulated size measure is the samllest greater than or equal to the random start plus the skip interval. In general the $(r+1)$th unit selected is that for which the cumulated size measure is the smallest greater than or equal to the random start plus r times the skip interval.

This is the simplest way of selecting a sample with unequal probabilities without replacement. Because of the ordering process, π_{IJ} will be zero for most pairs I, J. In consequence the Yates-Grundy variance estimator will yield considerable under-estimates of variance. Hartley (1966) sought to overcome this disadvantage by making an assumption about the nature of the population sampled. This assumption is that, for any given unit, the value of the variable being estimated depends on the order in which it appears in the population. The population is therefore divided into quasi-strata, one for each sample unit, and the variance calculated accordingly.

PROCEDURE 2: Random Systematic Procedure

Strπps, strwor, n fixed, unord, b est var, j p enum

Goodman and Kish (1950), Horvitz and Thompson (1952), Hartley and Rao (1962), Rao (1963b), Raj (1964, 1965), Connor (1966), Hanif (1974), Asok and Sukhatme (1976), Cassel *et al* (1977, p. 17).

This procedure is identical with the Ordered Systematic Procedure 1, except that the population units are listed in random order prior to selection.

For this type of selection procedure Hartley and Rao (1962) have given a formula for π_{IJ}, which is asymptotically correct as $N \to \infty$ under certain conditions (see Chapter 1, Section 1.8).

Connor (1966) gave the exact formula for π_{IJ} for any value n and N for this selection procedure.

The main drawbacks of the systematic procedures are the difficulty of calculating the joint probabilities of inclusion for the purpose of estimating the variance, and the fact that one or more of these joint probabilities is sometimes zero. A simple example of a situation in which one of the π_{IJ} is zero is given by $n = 2$; $N = 5$; $Z_I = 1, 2, 4, 5, 6$.

PROCEDURE 3: Grundy's Systematic Procedure

Strπps, strwor, n fixed, syst, unord, b est var, j p enum

Grundy (1954).

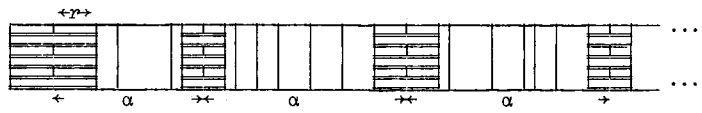
This procedure is a modification of the Random Systematic Procedure 2 which avoids listing all population units in random order. Instead, a single unit is selected with probability proportional to size using a procedure originally devised by Lahiri (1951) (see Procedure 45) and the remaining units are selected systematically using the size of the largest unit in the population Z_{max} (or a somewhat larger round number for convenience) as skip interval. Since it is rather difficult to follow this procedure unless it is spelled out step by step, Grundy's description will be repeated (with the notation slightly altered in order to avoid confusion with symbols used elsewhere in this monograph).

(i) Let α denote either Z_{max}, or, if more convenient a round number slightly larger. (The inequalities $Z_{max} \le \alpha \le (Z-Z_{max})/(n-1)$ are the essential conditions on α .)

(ii) Choose a random number r in the range $0 < r < \alpha$ and a random integer s in the range 1 to N. If $r \le Z_s$ accept unit s as a member of the sample and proceed to (iii); otherwise repeat (ii).

(iii) Choose further integers s_2, s_3 in the range 1 to N, distinct from each other and from s, but otherwise random. Note the sizes of the corresponding sampling units. Each time the cumulative sum $r + Z_{s_2} + Z_{s_3} + \ldots + Z_{s_t}$ first exceeds one of the values $\alpha, 2\alpha, \ldots, (n-1)\alpha$, accept the unit s_t as a member of the sample.

This selection procedure may be represented by the following diagram, which corresponds to $n = 4$.



Because the skip interval is smaller than for the Random Systematic Procedure 2, cases where the joint probabilities of selection of one or more pairs of units are zero will be still less common. However it is still relatively easy to produce such cases. One is given by $n = 2$; $N = 5$; $Z_I = 1, 2, 4, 5, 8$; skip interval = 8 . The joint probability of selection of the smallest pair is then zero. It is, of course, only under such circumstances that the formulation of an unbiased estimator of variance is impossible.

Another advantage of this modified procedure over the Random Systematic Procedure is that if N is large compared with nZ/Z_{max} it is only necessary to order a small portion of the population randomly.

PROCEDURE 4: Yates Grundy Draw By Draw Procedure

Not *strπps*, *strwor*, n *fixed*, *d by d*, *unord*, *inexact*, *non-rotg*.
Yates and Grundy (1953), Raj (1956a), Murthy (1957), Hanurav (1962b), Hajek (1964), Rao and Bayless (1969), Bayless and Rao (1970), Cassel *et al* (1977, pp. 15, 24, 42-3, 153ff).

Select the first unit in the sample with probability proportional to size Z_I ; the second unit, without replacement, again with probability proportional to size. The total probability of the inclusion of the Ith unit to be in sample is

$$\pi_I = P_I\left[1 + \sum_{J=1}^{N} \frac{P_J}{1-P_J} - \frac{P_I}{1-P_I}\right] . \qquad (2.2.1)$$

The joint probability of selecting the Ith and Jth unit is

$$\pi_{IJ} = P_I P_J\left[\frac{1}{1-P_I} + \frac{1}{1-P_J}\right] . \qquad (2.2.2)$$

The selection may proceed to $n = 3$ or more, but the formulae for π_I, π_{IJ} and so on, become rapidly complicated.

The procedure is *inexact*, but the unbiased estimators of Raj (1956a) and Murthy (1957), compensate for this inexactness. They will be considered in detail in Chapter 4.

PROCEDURE 5: Yates-Grundy Rejective Procedure

Not *strπps*, *strwor*, n *fixed*, *rej*, *unord*, *inexact*, *non-rotg*.
Yates and Grundy (1953), Durbin (1953a), Hajek (1964).

Select each of the n sampling units in turn with probability proportional to size with replacement. If any unit is selected more than once in the sample, reject the whole sample selected up to that point and continue selecting a new sample with replacement until n distinct units are selected in the sample.

For $n = 2$, the probability of rejecting the initial sample because unit I was selected twice is P_I^2 , that of selecting units I and J in either order is $2P_IP_J$. The probability of including unit I in the final sample is therefore

$$\pi_I = 2P_I(1-P_I) \Big/ \left(1 - \sum_{\substack{J=1 \\ J\neq I}}^{N} P_J^2\right) . \qquad (2.2.3)$$

The sample for $n > 2$ may be obtained similarly, but formulae become complicated very rapidly.

The order of approximation to exactness is poorer than for the Yates-Grundy Draw-by-Draw Procedure 4.

PROCEDURE 6: Midzuno's Procedure

Strπps, *strwor*, n *fixed*, *d by d*, *unord*, *j p iter*, *not gen app*, *non-rotg*.
Horvitz and Thompson (1952), Yates and Grundy (1953), Rao (1963a).

Select the first unit using a specially calculated set of working probabilities P_I^* , and the remaining units with equal probabilities without replacement. The probabilities used at the first draw are such that the total probability of inclusion of each unit is exactly proportional to size, that is,

$$\pi_I = P_I^* + \frac{n-1}{N-1}\left(1-P_I^*\right) , \quad \sum P_I^* = 1 . \qquad (2.2.4)$$

From (2.2.4),

$$P_I^* = \frac{N-1}{N-n}\pi_I - \frac{n-1}{N-n} . \qquad (2.2.5)$$

Similarly

$$\pi_{IJ} = \frac{n-1}{N-1}\left[\frac{N-n}{N-2}\left(P_I^*+P_J^*\right) + \frac{n-2}{N-2}\right] . \qquad (2.2.6)$$

Horvitz and Thompson mentioned that this selection procedure was suggested by Midzuno, presumably on the analogy of his other selection procedure for selecting samples with probabilities proportional to the aggregate measures of size (PPAS) of the units in the sample (Procedure 46). Procedure 6 will break down unless $Z_I \geq Z(n-1)/n(N-1)$ for all I . This is a very stringent requirement; consequently the procedure is frequently not applicable.

Rao (1963a) has shown that for $n = 2$ the variance of y'_{HT} with this Procedure is always smaller than the variance of the *ppswr* estimator y'_{HH} provided $Z_I > Z/2(N-1)$, which is also the condition for non-negativity of the working probabilities.

PROCEDURE 7: Narain's Procedure

Strπps, *strwor*, n *fixed*, *d by d*, *unord*, $n = 2$ (for all practical purposes), *j p iter*, *non-rotg*.
Narain (1951), Horvitz and Thompson (1952), Yates and Grundy (1953), Brewer and Undy (1962), Rao (1963b), Cassel *et al* (1977, p. 21).

Calculate a set of working probabilities P_I^* . Select the first unit using these

P_I^* , and the second unit without replacement with probabilities proportional to the same P_I^* . These working probabilities are calculated so that each unit's probability of inclusion in sample is proportional to size. They thus conform to the requirement:

$$\pi_I = P_I^* + \sum_{J \neq I}^{N} \frac{P_I^*}{1-P_J^*} P_J^* \ . \qquad (2.2.7)$$

A numerical procedure for obtaining the P_I^* is given in Appendix A.

The joint probability of inclusion of the Ith and the Jth units together in sample is:

$$\pi_{IJ} = P_I^* P_J^* \left(\frac{1}{1-P_I^*} + \frac{1}{1-P_J^*} \right) \ . \qquad (2.2.8)$$

The sample values of the π_{IJ} are required for the Sen-Yates-Grundy variance estimator.

For $n = 2$ this procedure is generally applicable and the Horvitz and Thompson estimator is always more efficient than the corresponding multinomial sampling estimator. For $n > 2$ the equations for the working probabilities are so complicated that no solution has been proposed.

PROCEDURE 8: Brewer's Procedure

Strπps, strwor, n fixed, d by d, unord, non-rotg, j p enum (for $n > 2$)
Brewer (1963, 1975), Rao and Bayless (1969), Rao and Singh (1973), Chromy (1974), Fuller (1971), Sadasivan and Sharma (1974), Cassel *et al* (1977, p. 16).

Select the rth last sample unit, from among those not already selected, with working probabilities proportional to $P_I(1-P_I)/(1-rP_I)$. Brewer (1975) gives a recursive formula for the joint probabilities of inclusion in sample which involves the consideration of the selection probabilities of all subsets of the sample containing $(n-m)$ units from population of $(N-m)$ $(m = 1, 2, \ldots, n-2)$.

This takes a simple form when $n = 2$, that is,

$$\pi_{IJ} = \left(2P_I P_J \left[\frac{1}{1-2P_I} + \frac{1}{1-2P_J} \right] \right) \div \left(1 + \sum_{K=1}^{N} \frac{P_K}{1-2P_K} \right) \ . \qquad (2.2.9)$$

In this case the joint probabilities of inclusion, and hence also the variance estimator, are simple functions of size. Rao (1963a) found that the Horvitz-Thompson estimator was always more efficient than the corresponding Hansen-Hurwitz estimator for multinomial sampling, and that its variance estimator was never negative. Chromy (1974) found that the π_{IJ} for this procedure (still for $n = 2$) asymptotically minimized the expected variance of the Horvitz-Thompson estimator when $\gamma = \frac{1}{2}$.

PROCEDURE 9: Durbin's 'Method I'

Strπps, strwor, n fixed, d by d, unord, not gen app for $n > 2$.
Durbin (1967), Rao and Bayless (1969), Brewer and Hanif (1970), Fuller (1971), Cassel *et al* (1977, p. 16).

For $n = 2$, select the first unit with probability P_I and the second unit without replacement with probability proportional to $P_J \left[\frac{1}{1-2p_1} + \frac{1}{1-2P_J} \right]$, where p_1 is the sample value of the normed size measure P_I of the unit already selected at the first draw.

The joint probability of inclusion of the Ith and Jth units is

$$\pi_{IJ} = \left(2P_I P_J \left[\frac{1}{1-2P_I} + \frac{1}{1-2P_J} \right] \right) \div \left(1 + \sum_{K=1}^{N} \frac{P_K}{1-2P_K} \right) \qquad (2.2.10)$$

which is the same as given for Procedure 8 (2.2.9). Brewer's Procedure 8 and Durbin's 'Method I' are therefore in the same equivalence class for $n = 2$. Hence in this case Procedure 9 shows the same properties of simplicity of variance estimation and the same superior efficiency vis-à-vis multinomial sampling as Procedure 8.

For $n > 2$, the probability of selection of the Jth population unit at the rth draw, conditional on the results of all the previous draws and in particular given that the Ith unit was selected at the $(r-1)$th draw, is proportional to

$$P_{(r-1)J} \left[\frac{1}{1-2P_{(r-1)I}} + \frac{1}{1-2P_{(r-1)J}} \right] \ ,$$

where $P_{(r-1)I}$ is the probability of selection of the Ith unit at the $(r-1)$th draw, conditional on the results of the previous draws.

This extension to $n > 2$ is not generally applicable. Fuller (1971) with his 'Scheme A' extended its range of applicability by introducing certain modifications when some units had size Z_I close to Z/n . Even so, the extension remained not generally applicable.

The procedure would be suitable for rotation, in that the probability of inclusion in sample is constant from draw to draw, but for its lack of general applicability for $n > 2$. An example of where the procedure fails to be generally applicable is given by $n = 3$, $N = 102$; $P_1 = P_2 = 0.3$, $P_3 = P_4 = \ldots = P_{102} = 0.004$.

PROCEDURE 10: Durbin's "Grouped Method"

Strπps, strwor, n fixed, d by d, unord, n = 2 only, non-rotg.

Durbin (1967), Cassel *et al* (1977, p. 16).

Arrange the population units in groups such that each group contains as few units as possible subject to the requirement that the size of the largest unit in each group is less than or equal to half the total size of the group. Select two units from the whole population with replacement. If the units are from different groups, accept both; otherwise accept the first one, replacing the second unit by the second selection using Procedure 9 within the doubly selected group only. For any two units coming from different groups their joint probability of inclusion in sample is $\pi_{IJ} = 2P_I P_J$. For any two units in the same group, their unconditional joint probability of inclusion in sample is

$$\pi_{IJ} = \left(P_I\left(\sum\nolimits' P_I\right)P_J'\left(\frac{1}{1-2P_I'} + \frac{1}{1-2P_J'}\right)\right) \div \left(1 + \sum\nolimits' \frac{P_K'}{1-2P_K'}\right) , \qquad (2.2.11)$$

where $\sum'$ denotes summation over the units in the group and

$$P_I' = P_I \div \sum\nolimits' P_I .$$

This selection procedure is slightly less convenient than some others because it requires grouping, but on the other hand it avoids the need for any special calculation of the π_{IJ} if the two units initially selected are from different groups. It thus achieves a measure of simplification in the estimation of variance, but at the cost of some stability in the Sen-Yates-Grundy variance estimator. By using a randomization device in the variance estimation procedure, still further simplification may be obtained at a slight extra cost in stability. This procedure was, in fact, specially devised for handling variance estimation at two or more stages in a simple fashion.

PROCEDURE 11: Rao-Sampford Rejective Procedure

Strπps, strwor, n fixed, rej, unord, non-rotg.
Rao (1965), Sampford (1967), Rao and Bayless (1969), Bayless and Rao (1970), Asok and Sukhatme (1976), Cassel *et al* (1977, p. 17).

Select the first unit with probability proportional to measure of size. At each subsequent draw select with probability of selection proportional to $P_I/(1-nP_I)$ with replacement. If any unit is selected twice, reject the whole sample selected and start again. The joint probability of selection for any n is

$$\pi_{IJ} = \frac{I_n \lambda_I \lambda_J}{n(n-1)} \sum_{t=2}^{n} \left[\frac{\{t-n(P_I+P_J)\}L_{n-t}^{(\bar{I},\bar{J})}}{n^{t-2}}\right] , \qquad (2.2.12)$$

where

$$K_n = \left[\sum_{I=1}^{N} \frac{tL_{n-t}}{n^t}\right]^{-1} ,$$

$$L_0 = 1 ,$$

$$\lambda_I = Z_I(Z-nZ_I) ,$$

and

$$L_m = \sum_{S(m)} \lambda_{I1}, \lambda_{I2}, \ldots, \lambda_{Im} ,$$

the summation $\sum_{S(m)}$ being over all possible sets of m distinct units drawn from the population. $L_m(\bar{I})$, $L_m(\bar{I}, \bar{J})$ are defined similarly to L_m but relate to the subpopulation formed by omitting unit I and units I and J respectively from the population.

For $n = 2$,

$$\pi_{IJ} = K_2 P_I P_J\left(\frac{1}{1-2P_I} + \frac{1}{1-2P_J}\right)$$

$$= \left(2P_I P_J\left[\frac{1}{1-2P_I} + \frac{1}{1-2P_J}\right]\right) \div \left(1 + \sum_{K=1}^{N} \frac{P_K}{1-2P_K}\right) \qquad (2.2.13)$$

which is identical with (2.2.9) and (2.2.10) so that in this case Procedures 8, 9 and 11 are in the same equivalence class.

This procedure was first suggested by Rao (1965) for $n = 2$ only. Sampford (1967) extended this procedure to cover $n > 2$, but the formula for the π_{IJ}, (2.2.12), is then rather complex. The difficulty in their calculation stems from the large number of decimals which must be stored if they are to be calculated with any acceptable degree of accuracy.

Asok and Sukhatme (1976) compared this procedure with Procedure 2 and proved that the Rao-Sampford Procedure was the more efficient asymptotically for y_{HT}' (1.4.1). They also provided good approximations for the π_{IJ}.

PROCEDURE 12: Durbin-Sampford Procedure

Strπps, strwor, n fixed, d by d, unord, non-rotg.
Sampford (1967), Cassel *et al* (1977, pp. 16-17).

This procedure is difficult to describe for general n. Sampford's description of the procedure for $n = 4$ will therefore be reproduced with appropriate notational changes.

Select the first unit (say the Ith) with probability

$$P_I(1) = C_1\lambda_I \sum_{\substack{K\neq I, J<K \\ I\neq J}}\sum \lambda_J\lambda_K\left(1-P_I-P_J-P_K\right) . \qquad (2.2.14)$$

Select the Jth unit if the Ith unit has already been selected with probability

$$P_J(2|I) = C_2\lambda_J \sum_{K\neq I, J} \lambda_K\left(1-P_I-P_J-P_K\right) . \qquad (2.2.15)$$

Select the Kth unit when the Ith and Jth units are already selected with probability

$$P_K(3|I, J) = C_3\lambda_K\left(1-P_I-P_J-P_K\right) . \qquad (2.2.16)$$

Select the Lth unit when the Ith, Jth and Kth have already been selected with probability

$$P_L(4|I, J, K) = C_4P_L . \qquad (2.2.17)$$

Equating the sum of the $P_I(1)$ at each stage to 1 we have

$$\left.\begin{aligned} P_K(3|I, J) &= C_3\lambda_K C_4^{-1} , \\ P_J(2|I) &= C_2\lambda_J C_3^{-1} , \\ P_I(1) &= \tfrac{1}{2}C_1\lambda L C_2^{-1} , \end{aligned}\right\} \qquad (2.2.18)$$

so that the probability of drawing the Ith, Jth, Kth and Lth units in that order is $\frac{1}{2}C_1\lambda_I\lambda_J\lambda_K P_L$; that of drawing the Ith, Jth and Kth in any order, followed by the Lth , is $3C_1\lambda_I\lambda_J\lambda_K P_L$ and that of drawing the Ith, Jth, Kth and Lth in any order is

$$3C_1\lambda_I\lambda_J\lambda_K P_K\left(\frac{P_I}{\lambda_I}+\frac{P_J}{\lambda_J}+\frac{P_K}{\lambda_K}+\frac{P_L}{\lambda_L}\right) . \qquad (2.2.19)$$

Although this procedure does not set out to use working probabilities, the selection probabilities for the $(n-1)$th and the nth draws are in fact dependent on working probabilities. Consequently for $n = 2$ this procedure is identical in every way with Brewer's Procedure 8. For $n > 2$, the selection probabilities are somewhat more difficult to calculate than $P_I\left(1-nP_I\right)$, so that the Rao-Sampford Procedure 11 is more convenient to use in practice.

PROCEDURE 13: Fellegi's Procedure

Strπps, strwor, n fixed, d by d, unord, j p iter.

Fellegi (1963), Brewer (1967), Rao and Bayless (1969), Bayless and Rao (1970), Cassel *et al* (1977, p. 16).

Select a unit using working probabilities equal to the normed measures of size. At each subsequent draw select one unit without replacement using working probabilities calculated in such a way that the *a priori* probabilities of selection at that draw are also proportional to size. These working probabilities must be calculated by an iterative procedure, which is fairly simple for the second draw but becomes less tractable as the number of the draw increases, especially if any of the Z_I is close to Z/n . Iterative processes for calculating these working probabilities for $n = 2$ are given in Appendix A. The iteration for $n > 2$ can be slow if the population units are very unequal in size, and care should be taken to ensure that the process has converged before using the working probabilities to select the sample.

This procedure has only been demonstrated to be generally applicable for $n = 2$, but appears to have this property for all values of n .

This procedure was devised specifically to facilitate rotation of the sample. The probability of inclusion in sample is maintained proportional to size of unit because the probability of selecting the Ith unit in sample at each draw is P_I .

PROCEDURE 14: Carroll-Hartley Rejective Procedure

Strπps, strwor, n fixed, rej, unord, j p iter.

Rao (1963b), Carroll and Hartley (1964), Hajek (1964), Rao and Bayless (1969), Bayless and Rao (1970), Cassel *et al* (1977, p. 16).

Select the sample of n units with working probabilities P_I^* with replacement. If not all the units selected are distinct, discard the sample and repeat the same procedure again until n distinct units are selected in the sample. The working probabilities P_I^* must be so chosen that the probability of including the Ith unit in sample is nP_I .

For $n = 2$ this procedure is in the same equivalence class as Fellegi's Procedure 13. For larger values of n it is nearly, but not, quite equivalent (that is, the joint probabilities of selection are nearly the same). Like Fellegi's procedure it appears to be generally applicable for $n > 2$. Rao (1963b) proved that for $n = 2$ the variance of y'_{HT} for this procedure is smaller than that for the corresponding multinomial sampling estimator y'_{HH} .

PROCEDURE 15: Carroll-Hartley Draw-by-Draw Procedure

Strπps, strwor, n fixed, d by d, unord, j p iter.

Carroll and Hartley (1964), Cassel *et al* (1977, p. 16).

This, the draw by draw procedure equivalent of Procedure 14, was used by Carroll and Hartley as an aid in determining the working probabilities for Procedure 14.

To use Procedure 15, as opposed to Procedure 14, involves additional calculation. On the other hand it avoids the selection and consequent rejection of unacceptable samples. For $n = 2$ it is identical in every way with Fellegi's Procedure 13.

PROCEDURE 16: Carroll-Hartley Whole Sample Procedure

Strmps, Strwor, rej, n fixed, ws, unord, j p iter.

Carroll and Hartley (1964), Cassel *et al* (1977, p. 16).

This whole sample equivalent of Procedures 14 and 15 was mentioned by Carroll and Hartley solely for purpose of logical completeness. It is less convenient for selection than either of those procedures and appears to offer no compensating advantages.

PROCEDURE 17: Durbin-Hanurav Procedure (or, Hanurav's Scheme B-A)

Strmps, strwor, n fixed, rej, unord, n = 2 only.

Durbin (1953b), Hanurav (1966, 1967), Cassel *et al* (1977, p. 16).

(i) Arrange the population units in ascending order of size, so that the normed measure of size of the largest unit is P_N, and of the next largest is P_{N-1}.

(ii) Conduct a Bernoulli trial (Hanurav's Scheme B) in which the probability of success is

$$\delta = \frac{2(1-P_N)(P_N-P_{N-1})}{1-P_N-P_{N-1}}, \quad 0 \le \delta < 1 .$$

(iii) If the trial is successful at step (ii), the sample consists of the largest unit and one other selected from the remainder with probabilities proportional to the measure of size.

(iv) If the trial is not successful at step (ii),

(a) the measure of size of the largest unit is reduced to that of the next largest,

(b) all the measures of size, thus modified, are normed to sum to unity. (These normed measures and sizes will be denoted by P_I^*. Note that $P_N^* = P_{N-1}^*$.)

A sample of two units is then selected using a specifically devised scheme which depends for its validity on the equality of the two largest units. Hanurav has suggested three separate schemes at this point (Schemes A, A' and A"), and each gives rise to a different set of joint probabilities of selection.

(v) The first of these is the scheme A, described below. Select two units with replacement with working probabilities P_I^*. If the sample consists of two distinct units, accept it. Otherwise select two units with replacement with probabilities proportional to P_I^{*2}. Again if the sample consists of two distinct units accept it. Otherwise reject the sample and select two units with replacement with probabilities proportional to P_I^{*4}, and so on, until at the kth trial two distinct units are selected using probabilities proportional to $P_I^{*2^{k-1}}$. The process terminates with probability one.

Durbin (1953b) derived the formulae appropriate to Hanurav's Scheme A, but without Scheme B it could only be used when $P_N = P_{N-1}$. Durbin did not publish this derivation and Hanurav developed Scheme B-A independently of Durbin's work.

Hanurav indicated in an abstract (1966) that this procedure could be extended to $n > 2$. Although full details were not given it would appear that the first part of the procedure would be identical with Steps (i)-(iv) of the Hanurav-Vijayan Procedure 19 (q.v.). The remaining part appropriate to the situation when the population is arranged in ascending order of size and $P_{N-n+1}^* = P_N^*$ is described by Hanurav in his 1966 abstract.

PROCEDURE 18: Hanurav's Scheme B-A'

Strmps, strwor, n fixed, d by d, unord.

Hanurav (1967, 1969), Cassel *et al* (1977, p. 16).

In this selection procedure the first four steps are the same as for the Durbin-Hanurav Procedure 17. The remainder are as follows:

(v) Select the first unit with a set of working probabilities calculated so as to ensure that the total probabilities of inclusion are exactly proportional to size. The working probabilities α_N for this step were given at first incorrectly in Hanurav (1967) but correctly in Hanurav (1969) as

$$\alpha_I = 2P_I^* - \frac{\alpha_1}{N-1} - \frac{\alpha_2}{N-2} - \ldots - \frac{\alpha_{I-1}}{N-I+1}, \quad 1 \le I \le N-1 . \qquad (2.2.20)$$

(Note that when α_N is defined by (2.2.20) it takes the value zero.)

(vi) Select the second unit with equal probabilities from among those units which are later in the population's size ordering then the first selected unit.

PROCEDURE 19: Hanurav-Vijayan Procedure (Hanurav's Scheme B-A")

Strmps, strwor, n fixed, d by d, unord.

Hanurav (1967, 1969), Vijayan (1968), Cassel *et al* (1977, p. 16).

The selection procedure as originally set out by Hanurav is limited to $n = 2$. Steps (i)-(iv) are as for Procedure 17. The remainder comprise Hanurav's Scheme A" and are as follows:

(v) Select the first unit with a special set of working probabilities. These are calculated so as to ensure that the total probabilities of inclusion are exactly proportional to size. The working probabilities β_I for this step were incorrectly given in Hanurav (1967) but correctly in Hanurav (1969) as

$$\beta_I = P_I^* \left(2 - \frac{\beta_1}{1-P_1^*} - \frac{\beta_2}{1-P_1^*-P_2^*} - \cdots - \frac{\beta_{I-1}}{1-P_1^*-\cdots-P_{I-1}^*} \right) , \quad 1 \le I \le N-1 . \qquad (2.2.21)$$

(Note that when β_N is defined by (2.2.21) it takes the value zero.)

(vi) The second unit is selected with probabilities proportional to P_I^* from among those units later in the population's size ordering then the one selected at step (v).

The selection procedure given by Vijayan for $n \ge 2$ is as follows:

(i) Arrange the population units in ascending order of size.

(ii) Choose one of the numbers $1, 2, \ldots, n$ with probabilities δ_i where

$$\delta_i = n\left(P_{N-n+i+1} - P_{N-n+i}\right) \frac{S+iP_{N-n+1}}{S} , \qquad (2.2.22)$$

$$S = \sum_{I=1}^{N-n} P_I ,$$

and, by convention, $P_{N+1} = n^{-1}$. (Note: when $n = 2$, δ_1 is the δ of Hanurav's Scheme B and $\delta_2 = 1 - \delta_1$.)

(iii) If the number i is chosen, the last $(n-i)$ population units form part of the sample and the remaining i are selected in accordance with Steps (iv)-(vi) following.

(iv) Define new normed measures of size $P_I^*(i)$ for the $(N-n+i)$ population units as yet unselected where

$$P_I^*(i) = \begin{cases} \dfrac{P_I}{S+iP_{N-n+1}} & \text{for } I < N-n+1 , \\ \dfrac{P_{N-n+1}}{S+iP_{N-n+1}} & \text{for } N-n+1 \le I \le N-n+i . \end{cases} \qquad (2.2.23)$$

(v) Select the first of the remaining i units from the first of the $(N-n+1)$ population units as yet unselected with probabilities proportional to $a_j(1)$ where

$$\left.\begin{aligned} a_1(1) &= nP_1^*(i) , \\ a_j(1) &= nP_j^*(i) \prod_{k=1}^{j-1} \{1-(i-1)P_k^{**}(i)\} \quad (j = 2, \ldots, N-n+1) , \end{aligned}\right\} \qquad (2.2.24)$$

and where

$$P_k^{**}(i) = \left(P_k^*(i)\right) \Big/ \left(\sum_{l=k+1}^{N-n+i} P_l^*(i) \right) .$$

(vi) If the unit selected at step (v) is the j_1th, select the second unit from the set consisting of the (j_1+1)th up to the $(N-n+2)$nd with probabilities proportional to $a_j(2, j_1)$ where

$$\left.\begin{aligned} a_{j_1+1}(2, j_1) &= (i-1)P_{j_1+1}^*(i) , \\ a_j(2, j_1) &= (i-1)P_j^*(i) \prod_{k=j_1+1}^{j-1} \{1-(i-2)P_k^{**}(i)\} , \\ & \qquad (j = j_1+2, \ldots, N-n+2) . \end{aligned}\right\} \qquad (2.2.25)$$

(vi) Proceed in a similar manner until the last sample unit is selected. In general if the $(l-1)$st of the i units remaining to be selected at the end of step (iv) is chosen to be the j_{l-1}th population unit, then select the lth unit from among the $(j_{l-1}+1)$st up to the $(N-n+l)$th with probabilities proportional to $a_j(l, j_{l-1})$ where

$$\left.\begin{aligned} a_{j_{l-1}+1}(l, j_{l-1}) &= (n-l+1)P_{j_{l-1}+1}^*(i) , \\ a_j(l, j_{l-1}) &= (n-l+1)P_j^*(i) \prod_{k=j_{l-1}+1}^{j-1} \{1-(i-l)P_k^{**}(i)\} \\ & \qquad (j = j_{l-1}+1, \ldots, N-n+l) . \end{aligned}\right\} \qquad (2.2.26)$$

PROCEDURE 20: Raj's Variance Minimization Procedure

Strπps, strwor, n fixed, ws, unord, n = 2 only, b est var, non-rotg.
Raj (1956b).

Use a linear programming technique to find those value of π_{IJ} which minimize the variance (1.4.5) of the Horvitz-Thompson estimator (1.4.1) on the assumption that the estimand variables Y_I and measures of size Z_I are related by

$$Y_I = \alpha + \beta Z_I \ , \tag{2.2.27}$$

where α and β are unknown parameters.

This procedure is insufficiently defined for $n > 2$. However values of π_{IJ} determined using the same linear programming technique and minimization criterion can be fed into one of Sinha's Procedures 42-43 (q.v.) to obtain a solution.

The value of the procedure is questionable in that, if the Y_I and Z_I satisfy (2.2.27) exactly, the regression estimator will have zero variance while the Horvitz-Thompson estimator in general, will not. Further, linear programming tends to give extreme solutions with one or more of the π_{IJ} taking the value zero. Hence the Sen-Yates-Grundy variance estimator is seldom, if ever, unbiased.

PROCEDURE 21: Hanurav's Simple Junctional Procedure

Strπps, strwor, n not fixed, d by d, unord[1], *inexact, b est var, non-rotg.*
Hanurav (1962a), Cassel *et al* (1977, p. 16).

If k_1 is the positive integer such that

$$\delta_1 = \sum_{I=1}^{k_1} \pi_I \le 1 < \delta_1 + \pi_{k_1+1} \ , \tag{2.2.28}$$

select the Ith unit with the probabilities

$$P_I(1) = \begin{cases} \pi_I & \text{if } 1 \le I \le k_1 \ , \\ 1-\delta_1 & \text{if } I = k_1 + 1 \ , \\ 0 & \text{if } I \ge k_1 + 2 \ . \end{cases} \tag{2.2.29}$$

Then (k_1+1)th unit is called a junctional unit. Since it is selected with probabilities $1 - \delta_1$ at the first draw, it will have probability π_{k_1+1} of being selected at least once (and hence of being included in the sample) if its probability

[1] Although Hanurav's suggestion is that the population should be arranged in size order, the procedure may be applied with any ordering.

of selection at the second draw is

$$P_{k_1+1}(2) = (\pi_{k_1+1} - (1-\delta_1))/\delta_1 \ , \tag{2.2.30}$$

and if at all subsequent draws its probability of selection is zero. The second draw is therefore made with the following selection probabilities:

$$P_I(2) = \begin{cases} 0 & \text{if } 1 \le I \le k_1 \ , \\ (\pi_{k_1+1} - (1-\delta_1))/\delta_1 & \text{if } I = k_1 + 1 \ , \\ \pi_I & \text{if } k_1+2 \le I \le k_1+k_2 \ , \\ 1-\delta_2 & \text{if } I = k_1 \ne k_2+1 \ , \\ 0 & \text{if } I \ge k_1+k_2+2 \ , \end{cases} \tag{2.2.31}$$

where k_2 is the integer such that

$$\delta_2 = P_{k_1+1}(2) + \sum_{k_1+2}^{k_1+k_2} \pi_I \le 1 < \delta_2 + \pi_{k_1+k_2+1} \ . \tag{2.2.32}$$

The number of distinct units selected in the first two draws is thus two, with probability $1 - P_{k_1+1}(1)P_{k_1+1}(2)$, and otherwise only one. The procedure is continued until it terminates.

Note that if neither the Ith nor the Jth unit is a junctional unit, and there is no junctional unit between them in the ordering, then $\pi_{IJ} = 0$. Hence an unbiased Sen-Yates-Grundy variance estimator cannot be obtained using this procedure. The next three procedures were devised in order to circumvent this problem.

PROCEDURE 22: Hanurav's Modified Junctional Procedure

Strπps, strwor, n not fixed, d by d, unord, inexact, non-rotg.
Hanurav (1962a), Cassel *et al* (1977, p. 16).

Select one unit from the entire population with probability proportional to size and the remaining units by Procedure 21 in such a fashion that the probability of selection at least once in the second or subsequent draws is

$$\delta_I = \pi_I \left(1 - \frac{1}{\nu}\right)\left(1 - \frac{\pi_I}{\nu}\right) \ . \tag{2.2.33}$$

The procedure does ensure that every pair of population units has a non-zero probability of inclusion in sample. However the joint probabilities of most of the pairs are only of order $P_I P_J/\nu$ and the Sen-Yates-Grundy variance estimator is unstable particularly for large values of ν .

PROCEDURE 23: Hanurav's Double Junctional Procedure

Strπps, strwor, n not fixed, d by d, unord, inexact, non-rotg.
Hanurav (1962a), Cassel *et al* (1977, p. 16).

Select two independent subsamples using Procedure 21, both with the same expected sample number, with the probability of inclusion of the Ith unit in each subsample given by

$$\pi_I' = 1 - (1-\pi_I)^{\frac{1}{2}} . \tag{2.2.34}$$

A unit is included in sample if it appears in either subsample. The total probability of inclusion of the Ith unit is therefore

$$1 - (1-\pi_I')^2 = \pi_I . \tag{2.2.35}$$

PROCEDURE 24: Hanurav's Sequential Procedure

Strπps, strwor, n not fixed, d by d, unord, inexact, non-rotg.
Hanurav (1962a), Cassel *et al* (1977, p. 16).

Select the Ith unit at the rth draw with probability

$$P_I(r) = \frac{P_I(K_r - K_{r-1})}{1-K_{r-1}P_I} , \tag{2.2.36}$$

when by convention $K_0 = 0$ and

$$K_r = K_{r-1} + \left[\sum_{I=1}^{N} \frac{P_I}{1-K_{r-1}} P_I\right]^{-1} . \tag{2.2.37}$$

This procedure resembles the three preceding ones in that although selection can occur more than once, inclusion in the sample is dependent only on selection at least once. It differs from these procedures in that the probability of inclusion in sample is proportional to size after each draw, so that sampling may be terminated once a desired sample number is obtained. The expected sample number after r draws is K_r. Hanurav supposed that as $r \to \infty$ all units in the population would eventually be included, so that $K_r \to N$. He deduced that after a certain number of draws, the probabilities of selection would turn negative. In fact, however, the limit of K_r as $r \to \infty$ is $P_{\max}^{-1}$. Since this implies that the probability of inclusion of the largest unit tends to unity as $r \to \infty$, this procedure is generally applicable in a rather special and interesting sense.

PROCEDURE 25: Rao-Hartley-Cochran Procedure

Not strπps, strwor, n fixed, d by d, unord, inexact, non-rotg.

Rao, Hartley and Cochran (1962), Stuart (1964), Raj (1966), Hanurav (1967), Rao and Bayless (1969), Bayless and Rao (1970), Singh and Singh (1974), Chotai (1974), Singh and Kishore (1975), Singh and Lal (1978), Cassel *et al* (1977, pp. 15, 153ff).

Divide the population units at random into n groups of N_J units, $J = 1, 2, \ldots, n$, where the N_J are predetermined. Select one unit from each group with probabilities proportional to the normed measures of size within the group.

A special unbiased estimator for use with this selection procedure will be discussed in Chapter 4. Although this estimator is very similar to the Horvitz-Thompson estimator, the two are not identical and their variance estimators are entirely different.

PROCEDURE 26: Stevens' Procedure

Not strπps, strwor, n fixed, d by d, unord, inexact, non-rotg.
Stevens (1958), Kish (1965), Durbin (1967).

Form the population units into groups of not less than n units each, all units within the group being as close in size as possible. Select n groups with replacement with probability proportional to their total measure of size. If the Jth group is selected n_J times, select n_J units from this group with equal probabilities without replacement. Slight variations of this procedure can be found from author to author and even from the same author but the principle involved is the same.

In practice the restriction to groups of at least n units may be relaxed, but then the procedure ceases to be strictly without replacement.

PROCEDURE 27: Poisson Sampling

Strπps, strwor, n not fixed, d by d, unord, inexact.
Hajek (1964), Ogus and Clark (1971), Brewer, Early and Joyce (1972), Brewer, Early and Hanif (1980), Cassel *et al* (1977, p. 17).

Assign a probability of inclusion in sample to each population unit. Conduct a set of N Bernoulli trials, using each of these probabilities in turn, to determine whether or not the corresponding unit is to be included in sample. The sample consists of all the units for which the trials have been successful.

This procedure will be considered in detail in Chapter 4. The variability of the number of units in sample in this procedure is large as compared with other procedures in which the sample number is a random variable. Nevertheless its extreme simplicity may will commend it for large samples.

Poisson sampling is known in forestry as 3-P sampling; see for instance Sethumadhavi and Rajagopalan (1974).

PROCEDURES 28-31: Hajek's Methods I-IV (respectively)

Not πps, *strwor*, *n fixed*, *rej*, *unord*, *inexact*.

Hajek (1964).

These are four approximations to the Carroll-Hartley Rejective Procedure devised by Hajek (1964) and labelled by him I-IV in descending order of accuracy. They differ from the Carroll-Hartley Rejective Procedure only in their working probabilities, which are given below (but not in Hajek's notation).

Procedure 28: Hajek's "Method I"

$$P_I(r) = \left(P_I(1-\lambda P_I)/(1-nP_I)\right) \Big/ \left(\sum_{J=1}^{N} P_J(1-\lambda P_J)/(1-nP_J)\right) \qquad (2.2.38)$$

where

$$\lambda^{-1} = \sum_{J=1}^{N} P_J(1-nP_J) + \left(\sum_{J=1}^{N} P_J^2(1-nP_J)\right) \Big/ \left(\sum_{J=1}^{N} P_J(1-nP_J)\right) . \qquad (2.2.39)$$

(Note that λ is of the order of unity.)

Procedure 29: Hajek's "Method II"

$$P_I(r) = \left(P_I(1-P_I)/(1-nP_I)\right) \Big/ \left(\sum_{J=1}^{N} P_J(1-P_J)/(1-nP_J)\right) . \qquad (2.2.40)$$

Procedure 30: Hajek's "Method III"

$$P_I(r) = \left(P_I/(1-\overline{n-1}P_I)\right) \Big/ \left(\sum_{J=1}^{N} P_J/(1-\overline{n-1}P_J)\right) . \qquad (2.2.41)$$

Procedure 31: Hajek's "Method IV"

$$P_I(r) = \left(P_I/(1-nP_I)\right) \Big/ \left(\sum_{J=1}^{N} P_J/(1-nP_J)\right) . \qquad (2.2.42)$$

Hajek's "Method I" appears to be a very accurate approximation. In the example Hajek gives, the probabilities of inclusion are exact to within one or two parts in a thousand. It may therefore be useful as an entry into the iterative procedure for the Carroll-Hartley Rejective Procedure 14.

PROCEDURE 32: Deming's Systematic Procedure

Strπps, *not strwor*, *n fixed*, *syst*, *ord*, *inexact*.

Deming (1960).

Select systematically m subsamples of n/m units using m different random starts and a skip interval of mZ/n .

Variances may be estimated in an unbiased fashion with $(m-1)$ degrees of freedom by comparing the results of the m systematic samples. The larger m is made,

however, the greater the departure from sampling without replacement. This procedure was originally suggested in the context of equal probability sampling, but is just as applicable to sampling with unequal probabilities.

PROCEDURE 33: Variance Estimator Optimization Procedure

Strπps, *strwor*, *n fixed*, *w.s.*, $n = 2$, *j p iter*.

Brewer and Hanif (1969a).

This procedure fixes the joint probabilities of inclusion in sample in such a fashion as to optimize the stability of the Sen-Yates-Grundy variance estimator under the stochastic model introduced in Chapter 1. Only the π_{IJ} can thus be specified so that this procedure by itself is insufficiently defined for $n > 2$. It can in principle be combined with one of Sinha's Procedures, 42, or 43, to arrive at a defined sample for any value of n . The stability of the Sen-Yates-Grundy variance estimator will be considered in detail in Chapter 3.

PROCEDURE 34: Jessen's Method 1

Strπps, *strwor*, *n fixed*, *w.s.*, *unord*, *non-rotg*.

Jessen (1969).

Choose a decrement D_m which is an aliquot part of every π_I . (If no such number exists the procedure is not applicable. It is simplest if the largest possible value is chosen for D_m .)

Set out a tableau, in which the first row displays the π_I , $I = 1, 2, \ldots, N$.

The first possible sample consists of these n units with the largest value of π_I . Ties are split arbitrarily. Subtract D_m from those n largest values. Enter in the second row of the tableau $(\pi_I - D_m)$ for those units included in the first sample and π_I for those units not in the first sample.

The rth possible sample consists of those n units with the largest values in the rth row of the tableau. Decrease these values by D_m and enter them in the $(r+1)$th row. Repeat the remaining values from the rth row into the $(r+1)$ row. Continue until the process terminates.

There will then be D_m^{-1} possible samples, some of which will in general be duplicated. Select one of these D_m^{-1} samples with equal probabilities.

EXAMPLE (Modified from Jessen, 1969)

Steps and Decrements	Population Unit No.				Interpretation
	1	2	3	4	
Step No. 1 (π_I)	.2	.4	.6	.8	Sample containing units 3 and 4
Decrements	.0	.0	.2	.2	
Step No. 2	.2	.4	.4	.6	Sample containing units 2 and 4
Decrements	.0	.2	.0	.2	
Step No. 3	.2	.2	.4	.4	Sample containing units 3 and 4
Decrements	.0	.0	.2	.2	
Step No. 4	.2	.2	.2	.2	Sample containing units 1 and 2
Decrements	.2	.2	.0	.0	
Step No. 5	.0	.0	.2	.2	Sample containing units 3 and 4
Decrements	.0	.0	.2	.2	
	.0	.0	.0	.0	

Result: Sample contains units 1 and 2 with probability 0.2 ;
2 and 4 with probability 0.2 ;
3 and 4 with probability 0.6 .

N.B. Jessen used the decrement 0.1 and obtained a different sample space. For comments, see Procedure 36, Jessen's Method 3.

PROCEDURE 35: Jessen's Method 2

Strmps, strwor, n fixed, w.s., unord, non-rotg.

Jessen (1969).

This is identical with Procedure 34, except that the decrement varies from step to step. An $(N+1)$th column is needed, with the initial entry unity, to indicate the probability not yet allocated to any sample prior to the rth step. The entries in this column are decremented at every step.

For the mth step the decrement is chosen to be as large as possible subject to two requirements. The first requirement is that, for every unit to be included in the mth sample, the decrement must not exceed the smallest of the entries in the rth row; otherwise one or more entries in the $(m+1)$th row could be negative. The second requirement is that, for every unit not included in the mth sample, the decrement must not exceed the smallest of the differences between the initial π_I and the entry for the Ith unit in the rth row; otherwise the unit corresponding to that entry would be left with an unallocated balance at the end of the process.

The selection of one out of the possible samples is then made with probabilities

given by corresponding decrements.

EXAMPLE (Source: Jessen 1969)

Steps and Decrements	Population Unit No.				$(N+1)$th column (Unallocated Probability)	Intrepretation
	1	2	3	4		
Step No. 1 (π_I)	.2	.4	.6	.8	1.0	Sample containing units 3 and 4
Decrements	.0	.0	.6	.6	.6	
Step No. 2	.2	.4	.0	.2	.4	Sample containing units 1 and 2
Decrements	.2	.2	.0	.0	.2	
Step No. 3	.0	.2	.0	.2	.2	Sample containing units 2 and 4
Decrements	.0	.2	.0	.2	.2	
	.0	.0	.0	.0	.0	

Result: Sample contains units 1 and 2 with probability 0.2 ;
2 and 4 with probability 0.2 ;
3 and 4 with probability 0.6 ;

N.B. This sample space is the same as for the example given under Procedure 34. For comments, see Procedure 36, Jessen's Method 3.

PROCEDURE 36: Jessen's Method 3

Strmps, strwor, n fixed, w.s., unord, non-rotg.

Jessen (1969).

Like Procedure 35, this uses a tableau with a probability decrement that varies from row to row, the $(N+1)$th column indicating, at each step, the probability as yet unallocated to any sample. Here, however, a step may relate to more than one sample.

At the mth step, the units are divided into three sets.

Set 1. Those whose entry in the tableau equals the unallocated probability in the $(N+1)$th column. These units may be included in the rth and all subsequent samples.

Set 2. Those whose entry is intermediate between zero and the corresponding entry in the $(N+1)$th column. These units must be included in some but not all of the mth and subsequent samples.

Set 3. Those whose entry is zero at that step. These units cannot be included in any further samples.

For the mth step, the rule is to include in sample with certainty all units in Set 1, and a random selection of the units in Set 2 large enough to make up the required sample size.

If the probability decrement entered in the $(N+1)$th column at the mth step is

D_m, the entries of those units in Set 1 must each also be decremented by D_m, while those of the units in Set 2 must be decremented by $D_m n_m/N_m$ where N_m is the number of units in Set 2 at the mth step, and n_m is the number of units in Set 2 required to make up the difference between the required sample size and the number of units in Set 1.

The decrement D_m is chosen to be as large as possible subject to two requirements. The first requirement is that $D_m n_m/N_m$ must not exceed the smallest entry for any unit in Set 2 (otherwise the entries could become negative). The second is that D_m must not exceed $N_m/(N_m-n_m)$ times the smallest of the differences between the original π_I and the current tableau entry in the Ith column for any unit in Set 2 (otherwise the unit corresponding to that entry would be left with an unallocated balance at the end of the process).

EXAMPLE (Source: Jessen 1969)

Steps and Decrements	Population Unit No. 1	2	3	4	(N+1)th Column (Unallocated Probability)	Interpretation
Step No. 1 (π_I)	.2	.4	.6	.8	1.0	Random non-replacement samples of 2 units from units 1, 2, 3 and 4
Decrements	.2	.2	.2	.2	.4	
Step No. 2	.0	.2	.4	.6	.6	Samples including unit 4 and a random selection of one from units 2 and 3
Decrements	.0	.2	.2	.4	.4	
Step No. 3	.0	.0	.2	.2	.2	Sample containing units 3 and 4 only
Decrements	.0	.0	.2	.2	.2	
	.0	.0	.0	.0	.0	

After completing the tableau, a random number is chosen between 0 and 1. The type of sample to be selected is indicated by the 'unallocated probability' entry in the (N+1)th column next larger than that random number. In the example given above, a random number in the range 0.2 up to but not including 0.6 would indicate that the sample should include unit 4 and a random selection of one from units 2 and 3.

These first three of Jessen's methods are reasonably convenient for selection purposes and the π_{IJ} are readily calculable. However extreme solutions with one or more of the $\pi_{IJ} = 0$ are common for Procedures 34 and 35. The π_{IJ} for Procedure 36 are always strictly positive. In the example given by Jessen they are identical to those obtained by the Random Systematic Procedure 2. These π_{IJ} are not close to the

optimum set for stability of the Sen-Yates-Grundy variance estimator (see Section 3.7).

PROCEDURE 37: Jessen's Method 4

Strmps, strwor, n fixed, w.s., n = 2 only, unord, non-rotg.
Jessen (1969).

Approximate the π_{IJ} by $\left\{\pi_I\pi_J - \left[n - \sum \pi_I^2\right]/N(N-1)\right\}$. Use trial and error to adjust these approximations so as to ensure that $\sum_{J\neq I}^{N} \pi_{IJ} = n\pi_I$ for all I. Select a sample of two units using these π_{IJ} to define the sample space.

This procedure approximates equality for the $(\pi_I\pi_J - \pi_{IJ})$ for all combinations I, J. Exact equality (which is generally impossible) would simplify the variance of y'_{HT} to

$$V(y'_{HT}) = \left[n - \sum_{I=1}^{N} \pi_I^2\right]\left[\left[\sum_{I=1}^{N} \left(\frac{Y_I}{\pi_I} - \frac{Y}{n}\right)^2\right] / (N-1)\right] . \qquad (2.2.43)$$

There is no necessity to use trial and error, since the variance of the $(\pi_I\pi_J - \pi_{IJ})$ could be minimized analytically. The use of trial and error in Jessen's example results in a set of π_{IJ} which are unfavourable for the stability of the Sen-Yates-Grundy variance estimator. Use of the analytical solution in this case yields a solution close to the theoretical optimum for the stability of this variance estimator. Two alternative formulae for the π_{IJ} which would not require the use of trial and error are given in (3.7.2) and (3.7.3).

PROCEDURE 38: Modified Poisson Sampling

Strmps, strwor, n not fixed, d by d, unord, inexact.
Ogus and Clark (1971), Brewer, Early and Hanif (1980).

Select an ordinary Poisson Sample (Procedure 27). If no units are selected in that sample, reselect repeatedly until a non-empty sample is achieved.

Assuming that the probability of inclusion in the sample of the Ith population unit is to be held constant at π_I, the probability of selecting this unit at each ordinary Poisson draw must be $\pi_I(1-P_0^*)$ where P_0^* is the probability of selecting an empty sample at each attempt. Then $P_0^* = \prod_{I=1}^{N} \{1-\pi_I(1-P_0^*)\}$ and its value may be obtained iteratively using the initial value zero.

Modified Poisson sampling was devised to reduce the variability in sample size which obtains with ordinary Poisson sampling, and in particular to ensure that an

empty sample is never selected. This procedure will be discussed in detail in Chapter 4.

PROCEDURE 39: Collocated Sampling

Strπps, strwor, n not fixed, d by d, unord, inexact.
Brewer, Early and Joyce (1972), Brewer, Early and Hanif (1980).

Collocated sampling is similar to Poisson sampling, but reduces the variation in sample size by requiring the random variables r_I used in the Bernoulli trials to be uniformly *spaced* instead of uniformly *distributed* over the interval $[0, 1)$. A random ordering L $(L_I = 1, 2, \ldots, N)$ is chosen with equal probabilities, and a random variable θ is also selected from a uniform distribution over the interval $[0, 1)$. The value of r_I is then defined as $(L_I+\theta-1)/N$. This procedure will also be discussed in detail in Chapter 4.

PROCEDURE 40: Das-Mohanty Procedure

Strπps, strwor, n fixed, w.s., unord, non-rotg.
Das and Mohanty (1973).

Form a sample space containing $b = b_1 + b_2$ samples, each of n distinct units. The b_1 samples are to be such that each population unit appears precisely r times in all samples combined, and that for each pair of units there is to be at least one sample where both appear together. In the b_2 samples the Ith population unit is to appear $(cZ_I - r)$ times in all samples combined. Select one sample at random from the complete sample space giving equal probability to each sample. Das and Mohanty supply three schemes for the construction of sample spaces with these characteristics.

The advantage of this procedure lies in the simplicity of calculation of the π_{IJ}. The difficulty lies in the construction of the sample spaces such that the resulting π_{IJ} provide stable variance estimators. The procedure can be used only when the Z_I are integers. When the original Z_I are not integers, they can - to any desired level of accuracy - be replaced by new integer-valued measures of size. If these are large, however, the procedure becomes difficult to manage. The same is true if Z_{max}/Z is close to n^{-1}, as then the required value of c is large.

PROCEDURE 41: Mukhopadhyay's Procedure

Strπps, strwor, n fixed, unord, d by d, non-rotg.
Mukhopadhyay (1972), Sinha (1973).

This procedure is one which enables a sample of any size n to be selected given the π_I and the π_{IJ} only. A description of this procedure is omitted here for the

following reasons:

1. it is extremely complicated both to describe and to use;
2. it can be considered as superseded by Sinha's Procedures 42-43.

Readers who wish to consider this procedure in detail are referred to Mukhopadhyay (1972). It will not be considered further in this monograph.

PROCEDURE 42: Sinha's 'Extension" Procedure

Strπps, strwor, n fixed, w.s., unord, non-rotg.
Sinha (1973).

Given any set of non-negative π_I and π_{IJ} which are feasible in the sense that the π_I sum to n, that the π_{IJ} sum over J to $(n-1)\pi_I$, and that no π_{IJ} exceeds $\min(\pi_I, \pi_J)$ or $(\pi_I+\pi_J-1)$, Sinha's Extension Procedure will provide one possible sample space consistent with that feasible set. The procedure is as follows.

Form the sample space for a sample of $N - 2$ $(\geq n)$ units with inclusion probabilities π_I^* and joint inclusion probabilities π_{IJ}^* given by

$$\left.\begin{aligned} \pi_I^* &= \pi_I(N-2)/n \ , \\ \pi_{IJ}^* &= \pi_{IJ}(N-2)(N-3)/n(n-1) \ . \end{aligned}\right\} \qquad (2.2.44)$$

The sample space for this sample of $N - 2$ units is defined by

$$Pr\{\text{sample excludes units } I \text{ and } J\} = 1 + \pi_{IJ}^* - \pi_I^* - \pi_J^* \ .$$

If $n = N - 2$, the procedure terminates at this point. Otherwise, for each possible sample of $N - 2$ in that sample space, calculate the probabilities of selecting each possible subsample of n from the $N - 2$, using *srswor*. Add these probabilities over the sample space for each possible subsample of n units, and select one such subsample using these probabilities.

It will sometimes happen that the set $\{\pi_I^*, \pi_{IJ}^*\}$ is not feasible in the sense stipulated above for the set $\{\pi_I, \pi_{IJ}\}$. In these cases some of the 'possible' samples of $N - 2$ units will have negative probabilities. If, however, the resulting negative probabilities for the *srswor* subsamples of n units are added algebraically, the procedure will still yield a feasible solution.

PROCEDURE 43: Sinha's Reduction Procedure

Strπps, strwor, n fixed, w.s., unord, non-rotg.
Sinha (1973).

Given any set of non-negative π_I and π_{IJ} which is feasible in the sense

defined for Procedure 42, Sinha's Reduction Procedure will provide the identical sample space to that obtainable from Procedure 42. The method is as follows.

Form the sample space for a sample of two units with inclusion probabilities π_I^{**} and joint inclusion probabilities π_{IJ}^{**} given by the implicit formulae

$$\left.\begin{aligned} \pi_I &= \frac{n-2}{N-2} + \frac{N-n}{N-2}\,\pi_I^{**} \,, \\ \pi_{IJ} &= \frac{(n-2)(n-3)}{(N-2)(N-3)} + \frac{(N-n)(n-2)}{(N-2)(N-3)}\left(\pi_I^{**}+\pi_J^{**}\right) + \left[1 - \frac{2(n-2)}{N-2} + \frac{(n-2)(n-3)}{(N-2)(N-3)}\right]\pi_{IJ}^{**} \,. \end{aligned}\right\} \quad (2.2.45)$$

If $n = 2$ the above equations result in the trivial solutions $\pi_I^{**} = \pi_I$, $\pi_{IJ}^{**} = \pi_{IJ}$, and a whole sample of two units can be selected with probabilities π_{IJ} in the usual way. For $n > 2$ the probability of selecting any given sample of n units is the sum (over the sample space of samples of two units) of the probabilities of selecting *srswor* the remaining $(n\text{-}2)$ units from the complementary set of $N - 2$ units.

As with Procedure 42 it will sometimes happen that the $\{\pi_I^{**}, \pi_{IJ}^{**}\}$ do not form a feasible set. If, however, the resulting negative probabilities are added algebraically the procedure will still yield a feasible solution.

PROCEDURE 44: Chaudhuri's Procedure

Strπps, *strwor*, n *fixed*, d *by* d, *unord*, *not gen app*.
Chaudhuri (1976).

Choose any exact *πpswor* procedure generally applicable for $n = 2$. Select the first two units using that procedure, but with working probabilities calculated so as to ensure that the final probabilities of inclusion (after the entire sample is selected) are proportional to size. Select the remaining $n - 2$ units from among those population units not previously selected, using *srswor*.

This procedure is somewhat analogous to Midzuno's (Procedure 6). Whereas Midzuno selects one unit with unequal and the remainder with equal probabilities, Chaudhuri selects two units with unequal and the remainder with equal probabilities. This procedure is applicable when $P_I > (n\text{-}1)/n(N\text{-}1)$ for all I.

PROCEDURE 45: Lahiri's Procedure

Not strπps, *strwor*, n *fixed*, *w.s.*, *unord*, *inexact*, *non-rotg*.
Lahiri (1951), Sankaranarayanan (1969), Rao and Bayless (1969), Bayless and Rao (1970), Vijayan (1975), Cassel *et al* (1977, pp. 120-121, 154ff).

Select a set of n units using *srswor* and find the aggregate size measure of those units. Choose a random number between zero and the sum of the sizes of the n largest units (or any number greater than this). If this random number exceeds the aggregate size of the *srswor* sample of n units, reject the sample as a whole; otherwise accept it. If the sample is rejected, repeat the process until a sample is accepted. Clearly the probability that a sample will be accepted is proportional to the aggregate measure of size of the sample units and in consequence the conventional ratio estimator is unbiased. This procedure will be considered in detail in Chapter 4.

PROCEDURE 46: Ikeda-Midzuno Procedure

Not strπps, *strwor*, n *fixed*, d *by* d, *unord*, *inexact*, *non-rotg*.
Midzuno (1952), Avadhani and Srivastava (1972), Singh (1975b).

Select r units using *srswor*. Select the Ith unit from the remaining $N - r$ units with probability $P_I + \sum_{j=1}^{r} p_j/(N\text{-}r)$. Select the remaining $n - r - 1$ units using *srswor*. The special case of this procedure with $r = 0$ was devised by Ikeda, and the general case by Midzuno. Like Lahiri's Procedure 45, it selects samples with probabilities proportional to their aggregate measures of size and in consequence the conventional ratio estimator is unbiased. Further discussion will be given in Chapter 4.

PROCEDURE 47: Fuller's Scheme B

Strπps, *strwor*, n *fixed*, d *by* d, $n = 2$ *only*, *not gen app*.
Fuller (1971).

Select the first draw with probabilities equal to the normed measure of size P_I, and at the second with probabilities given by

$$P_{J|I} = \tfrac{1}{2}P_J + \tfrac{1}{2}P_I P_J^2\left(\frac{1}{1+D}\right)\left(\left(1 \,/\, \sum_{K=1}^{N} P_K^2\text{-}2P_I^2\right) + \left(1 \,/\, \sum_{K=1}^{N} P_K^2\text{-}2P_J^2\right)\right) \,, \quad (2.2.46)$$

where

$$D = \sum_{L=1}^{N}\left(P_0^2 \,/\, \sum_{K=1}^{N} P_K^2\text{-}2P_L^2\right) \,. \quad (2.2.47)$$

This method has joint probabilities of inclusion π_{IJ} which minimize

$$\sum_{\substack{I,J=1 \\ J>I}}^{N}\left(\frac{\frac{1}{2}\pi_I\pi_J - \pi_{IJ}}{\pi_I\pi_J}\right)^2 \,, \quad (2.2.48)$$

and in consequence are nearly proportional to $\pi_I\pi_J$. The procedure is applicable only when

$$2P_{max}^2 < \sum_{I=1}^{N} P_I^2 , \qquad (2.2.49)$$

where P_{max} is the largest of the P_I.

Fuller (1971) indicated how this procedure could in principle be extended to cover $n > 2$ but did not give details. (N.B. Fuller's Scheme A for $n = 2$ is identical with Procedure 9, q.v.)

PROCEDURE 48: Singh's Procedure

Strmps, strwor, n fixed, syst, ordered, j p enum, not gen app.
Singh (1978).

1. Select a sample of size

$$n' = (N+1)/2 \quad \text{if } N \text{ is odd}$$
$$= N/2 + 1 \quad \text{if } N \text{ is even.}$$

(A) Select a random number I from 1 to N by a predetermined probability $P(I)$. For N odd, choose $P(I) = \left(n'\left(Z_I+Z_{I-1}\right)/Z\right) - 1$. For N even the specification of $P(I)$ is available but cumbersome.

(B) Starting with I select two contiguous units and thereafter $(n'-2)$ units in a circular systematic fashion with skip interval 2.

2. Select a sample of the required size n by simple random sampling from the n' already selected.

From the form of $P(I)$ with N odd it will be seen that the method can only be applied when $Z_I + Z_{I-1} \geq Z/n'$ for all I. Even with an optimum arrangement of the units (largest, smallest, second largest, second smallest, and so on) it is easy to produce counter-examples which violate this condition. The procedure does, however (like the other systematic methods), have good rotational properties.

PROCEDURE 49: Choudhry's Procedure

Strmps, strwor, n fixed, d by d, unord, j p iter, non-rotg.
Choudhry (1979).

This uses the Yates and Grundy Procedure 4 for all draws except the last, and at the last uses a set of working probabilities such that the total probabilities of inclusion in sample are proportional to size. For $n = 2$ the procedure is equivalent to Fellegi's Procedure 13. For $n > 2$ it has the advantage that only one set of working probabilities need be calculated instead of $(n-1)$. The procedure appears to be generally applicable for $n > 2$ but no proof of this is available.

PROCEDURE 50: Chromy's Procedure

Strmps, strwor, n fixed, ord, j p enum, non-rotg.
Chromy (1979)

Each population unit is considered in turn, and given a probability of inclusion in sample conditional on the history of the selection process up to that point.

Let $\text{Int}(I) = \left[\sum_{J=1}^{I} \pi_J\right]$ (that is the integral portion of the expression in square brackets) and $\text{Frac}(I) = \sum_{J=1}^{I} \pi_J - \text{Int}(I)$ (that is the fractional portion of the same expression).

By convention $\text{Int}(0) = \text{Frac}(0) = 0$.

The procedure is such that the number of units selected in sample prior to consideration of the Ith population unit is either $\text{Int}(I-1)$ or $\text{Int}(I-1) + 1$. The following table indicates the conditional probabilities of inclusion in sample of the Ith population unit, given the number of units already selected and the relationship between $\text{Frac}(I)$ and $\text{Frac}(I-1)$.

Case No.	Relationship	Conditional probability of inclusion in sample given	
		$\text{Int}(I-1)$ units previously selected	$\text{Int}(I-1) + 1$ units previously selected
(1)	$\text{Frac}(I) = 0$	1	0
(2)	$\text{Frac}(I) > \text{Frac}(I-1) \geq 0$	$\frac{\text{Frac}(I)-\text{Frac}(I-1)}{1-\text{Frac}(I-1)}$	0
(3)	$\text{Frac}(I-1) > \text{Frac}(I) > 0$	1	$\frac{\text{Frac}(I)}{\text{Frac}(I-1)}$

(Note that the above table is appropriate only when $0 < \pi_1 < 1$ for all I.)

This selection procedure ensures that at every point in the selection process the expected value of the (cumulated) number of sample units already selected is equal to $\sum_{J=1}^{I} \pi_J$.

It can easily be generalized to the case where some units are so large that $nZ_I/Z > 1$. The number of times such large units are to be included in sample is taken to be $\left[nZ_I/Z\right] + 1$ with probability $nZ_I/Z - \left[nZ_I/Z\right]$, and $\left[nZ_I/Z\right]$ with probability $1 - nZ_I/Z + \left[nZ_I/Z\right]$. It is then convenient to express the selection rules in a slightly different kind of table.

Case No.	Relationship	Conditional probability that cumulated number of units selected is to be $\text{Int}(I) + 1$ given	
		Previous cumulated number selected was $\text{Int}(I-1)$	Previous cumulated number selected was $\text{Int}(I-1) + 1$
(1)	$\text{Frac}(I) = 0$	0	0
(2)	$\text{Frac}(I) \geq \text{Frac}(I-1) \geq 0$	$\frac{\text{Frac}(I)-\text{Frac}(I-1)}{1-\text{Frac}(I-1)}$	1
(3)	$\text{Frac}(I-1) > \text{Frac}(I) > 0$	0	$\frac{\text{Frac}(I)}{\text{Frac}(I-1)}$

(This table may be used even if some of the nZ_I/Z take zero or integer values.)

To ensure that an unbiased variance estimator can be obtained for samples with $n > 2$, Chromy suggests the following steps:

"(1) Develop an ordered sampling frame of N [population] units;

(2) Select a unit with probability proportional to its size to receive the label 1;

(3) Continue labelling serially to the end of the sampling frame;

(4) Assign the next serial label to the first unit at the beginning of the list and continue until all [population] units are labelled;

(5) Apply the sequential ... sample selection algorithm starting with the sampling unit labelled 1."

Chromy suggests the use of this selection procedure with meaningfully ordered lists in order to obtain the reductions in variance associated with systematic, stratified or zone sampling. An unordered form of his procedure (starting with a randomized ordering of the population in Step (1) above) could, however, be used, and may yield π_{IJ} values closer to those required for the optimum stability of the Sen-Yates-Grundy variance estimator.

CHAPTER 3

UNEQUAL PROBABILITY PROCEDURES AND THE HORVITZ-THOMPSON ESTIMATOR

3.1. SELECTION PROCEDURES APPROPRIATE FOR USE WITH THE HORVITZ-THOMPSON ESTIMATOR

As mentioned in Chapter 1, the Horvitz-Thompson Estimator has a number of desirable properties when used with an exact sampling procedure. To the three given in Chapter 1 can now be added a fourth, that under model (1.8.5), the expected variance of the Horvitz-Thompson estimator achieves the lower bound of the expected variance for any design-unbiased estimator (Godambe and Joshi, 1965).

The conditions required in Chapter 2 for the description of a procedure as *exact* are that the selection should be strictly without replacement, that the probabilities of inclusion in sample should be strictly proportional to the original measures of size, and that the number of units in sample should be fixed. In Chapter 2, 32 of the 50 selection procedures described had these properties. In this Chapter an attempt will be made to evaluate these procedures under the assumption that they are being used together with the Horvitz-Thompson estimator of total and the Sen-Yates-Grundy estimator of variance.

The criteria for comparison will be limited to samples of size $n = 2$, general applicability, simplicity in selection, simplicity in variance estimation, the efficiency of the Horvitz-Thompson estimator of total, the unbiasedness and stability (that is, efficiency) of the Sen-Yates-Grundy variance estimator, and rotatability.

3.2. LIMITATION TO SAMPLES OF SIZE $n = 2$

If the number of units in the population is large and it is inconvenient to divide into strata, it is imperative that the sample should not be limited to two units. On the other hand it is often convenient to divide large strata into small ones, particularly if they are geographical entities. Moreover, each of the small strata thus formed can usually be made more homogeneous than the original population, making the sample more representative than would be possible without stratification. The case $n = 2$ is, in fact, the limiting case where the maximum advantage in stratification occurs consistent with obtaining an unbiased estimator of variance. (It is possible, by selecting one unit from each of a random subset of the strata and two units from each of the remainder, to push this advantage still further, but it is still necessary to have a suitable method of selecting two units from some of these strata.) Hence the limitation to $n = 2$, while important, is not as critical as it might appear.

The following are formally limited to the case $n = 2$; Procedures 7, 20, 33 and 37. All of these, however, with the exception of Procedure 7, can be extended to the case $n > 2$ by calculating values of π_{IJ} in accordance with the criterion suggested by the chosen procedure, and feeding these π_{IJ} into one of Sinha's Procedures 42-43.

3.3. GENERAL APPLICABILITY

Procedures were described in Chapter 2 as being generally applicable if they could be used given any feasible set of inclusion probabilities $\{\pi_I\}$. To constitute a feasible set the π_I must satisfy the conditions $0 < \pi_I \leq 1$ and $\sum_{I=1}^{N} \pi_I = n$ (an integer). Now the desired π_I are typically derived from non-negative size measures Z_I using the relationship $\pi_I = n\, Z_I/Z$ and thus automatically from a feasible set provide only that $Z_{max} \leq Z/n$. It is an obvious and serious inconvenience when feasible sets $\{\pi_I\}$ are encountered which cannot be catered for by a chosen procedure. Those procedures which are not generally applicable even for $n = 2$, namely, Procedures 6, 41, 44, 47 and 48, will therefore be excluded from further consideration. The following procedures, though defined for $n > 2$, are generally applicable for $n = 2$ only; Procedures 9, 10, and 18. These will be considered further in the context of that special case only. Procedures 13, 17, and 49 appear to be generally applicable for $n > 2$, but no proofs are available.

3.4. SIMPLICITY IN SELECTION

Simplicity is of great importance in the choice of a selection procedure, but it is difficult to be entirely objective in the comparison. It is nevertheless possible to draw the reader's attention to the salient features of the various selection processes and provide tentative assessments of their ease or difficulty. This is the approach which will be followed in this Section.

The Systematic Procedures 1, 2, 3, and 48 are particularly simple. Of these, the Ordered Systematic Procedure 1 requires no randomization of the ordering. In Grundy's Systematic Procedure 3, a portion of the population must be put into random order, and in the remaining Procedures 2 and 48, the whole of it. Chromy's Procedure 50 is somewhat less simple than these.

Jessen's two generally applicable decrement-based Procedures 35 and 36, are also quite simple. The decision rules to be used are straigthforward for Procedure 35 but less so for Procedure 36.

For the remaining procedures the cases $n = 2$ and $n > 2$ will be considered separately.

CASE 1: $n = 2$

Sinha's Reduction Procedure 43 takes a particularly simple form when $n = 2$. Since, however, the π_{IJ} are arbitrary, some method must be used for specifying them, which may itself be simple or complicated. Three simple methods will be presented in Section 3.7.

Brewer's Procedure 8 (which for $n = 2$ is identical with Durbin's Method I, that is Procedure 9), the Rao-Sampford Procedure 11, the Durbin-Sampford Procedure 12, the Hanurav Scheme B-A', 18 and the Hanurav-Vijayan Procedure 19 are all quite easy to apply since the Selection Procedure depends on the calculation of probabilities which are simple functions of measure of size. The Rao-Sampford Procedure 11, and the Durbin-Hanurav Procedure 17, being rejective, involve a slight extra complication over the others mentioned. Durbin's Grouped Method (Procedure 10) is less convenient than Durbin's Method I (Procedure 9) in that it requires grouping, but on the other hand it avoids the need for any special calculation whenever the two units initially selected are from different groups. This procedure must also be classed as easy to use.

Jessen's Method 4 (Procedure 37) involves the use of trial and error, and is consequently somewhat inconvenient.

Raj's Variance Minimization Procedure 20 uses linear programming to determine probabilities of whole samples. This is tedious especially when N is large. For any appreciable number of strata a computer program is necessary.

Fellegi's Procedure 13, the Carroll-Hartley Rejective Procedure 14 and the

Carroll-Hartley Draw-by-Draw Procedure 15 involve virtually identical iterative calculations. Several iterative algorithms have been proposed by Fellegi (1963) and by Carroll and Hartley (1964). One of those devised by Carroll and Hartley is claimed to be fairly rapid (2 decimal places per cycle) unless the largest π_I is near unity. Fellegi (1963) also reported quite rapid convergence for his algorithm, provided that all the π_I were less than 0.85-0.90. Alternatively, using the geometrical properties of Fellegi's Procedure pointed out by Brewer (1967) it is possible to obtain an algorithm which achieves a very satisfactory convergence rate even for quite extreme sets of values of π_I. This algorithm, written in BASIC, is given in Appendix A.

Narain's Procedure 7 also requires iteration to obtain the working probabilities used in selection. Procedures for obtaining iterative solutions have been described by Yates and Grundy (1953) and Brewer and Undy (1962). Appendix A contains a BASIC algorithm based on the geometrical properties described by the latter authors. Again a very satisfactory rate of convergence was found even for quite extreme sets of values of π_I.

The approximate values of Hajek's Method I (Procedure 28) may be useful as entry points to iterations for the Carroll-Hartley Rejective Procedure 14.

The Das-Mohanty Procedure 40 is relatively simple when the measures of size can be written as small integers. More usually, however, the number of possible samples which have to be considered is very large, and the selection procedure correspondingly tedious.

The procedures can therefore be arranged in six groups: the Systematic Procedures for which selection is particularly simple; then Jessen's Procedures 35 and 36, together with Sinha's Reduction Procedure 43; next Brewer's Procedure, Durbin's Method I (Procedure 9), Durbin's Grouped Method (Procedure 10), the Rao-Sampford Procedure 11, the Durbin-Sampford Procedure 12, the Durbin-Hanurav Procedure 17, Hanurav's Scheme B-A' 18 and the Hanurav-Vijayan Procedure 19 which are also fairly easy to use; Jessen's Procedure 37, which involves trial and error; Narain's Procedure 7, Fellegi's Procedure 13, the three Carroll-Hartley Procedures 14-16, and Raj's Variance Minimization Proecdure 20, all of which need iterative algorithms; and finally the Das-Mohanty Procedure 40 which in general involves the construction of quite complicated sample space. This last procedure will not be mentioned further.

The Carroll-Hartley Whole Sample Procedure 16 is of course, less simple to use for selection purposes than the corresponding draw-by-draw and rejective procedures. Since it possesses no advantages on the basis of any of the other criteria used in this study, it will also be dropped from the discussion from this point on. Sinha's Extension Procedure 42 is not appropriate for $n = 2$.

CASE 2: $n > 2$

In this situation the Systematic Procedures 1-3 and 48, Chromy's Procedure 50, and Jessen's Procedures 35-36 remain simple. Sinha's Extension Procedure 42 will usually be simple for populations and samples of small size. His Reduction Procedure 43 (which gives the same solution) is generally somewhat less simple for $n > 2$.

Of the remainder, Brewer's Procedure 8 and the Rao-Sampford Procedure 11 retain the same kind of simplicity as for $n = 2$. The Carroll-Hartley Rejective Procedure 14 is again somewhat more convenient to use than its draw-by-draw equivalent Procedure 15. Iteration is required for working out the selection probabilities, both for these for Fellegi's Procedure 13 and for Choudhry's Procedure 49. All the iterative algorithms mentioned for $n = 2$ are available except those based on the geometric properties of the solution. Choudhry's Procedure 49 requires only one set of working probabilities to be calculated iteratively. There is some doubt as to whether the iterative algorithms for Fellegi's Procedure 13 for $n > 2$ converge (see Section 3.7). The approximate working probabilities of Hajek's Method I (Procedure 28) would probably serve as useful entry points into the iterative algorithm for the Carroll-Hartley Rejective Procedure 14.

The order of simplicity is therefore much the same as for $n = 2$; the Systematic Procedures 1-3 and 48 with Jessen's Procedures 34-36, Chromy's Procedure 50 and Sinha's Procedure 42; next Brewer's Procedure 8 and the Rao-Sampford Procedure 11, and finally the Carroll-Hartley Procedures 14-15, Fellegi's Procedure 13, and Choudhry's Procedure 49 all of which require iteration.

3.5. SIMPLICITY IN VARIANCE ESTIMATION

This criterion is closely related to simplicity in selection. This is because the Sen-Yates-Grundy variance estimator contains the joint inclusion probabilities π_{IJ}, and with some notable exceptions (Procedures 1-3 and 48 for all n, Procedures 8 and 11 for $n > 2$) these follow readily from the same kinds of calculations as are needed to carry out selection.

The Ordered Systematic Procedure is in a special position here. Although the π_{IJ} can be simply enough calculated for any given population ordering, many of them will be zero, and the Sen-Yates-Grundy variance estimator will be so biased as to be meaningless. Consequently it is to all intents and purposes impossible to estimate the variance this way.

Connor (1966) produced an exact formula for the π_{IJ} for the Random Systematic Procedure 2. However the evaluation of this formula for any pair of units involves

adding contributions from all possible combinations of units separating the two in the pair. This can become tedious for large N. Nevertheless Connor's formula does (at present) make the estimation of variance for the Random Systematic Procedure 2 more amenable to computer programming than it is for the other Systematic Procedures or Chromy's Procedure 50.

CASE 1: $n = 2$

For Sinha's Procedures 42 and 43 the π_{IJ} are arbitrary and a method must be adopted for specifying them. The Rao-Sampford Procedure 11 and the Durbin-Hanurav Procedure 17 involve appreciably more work in estimating variance than in selection. For the Rao-Sampford Procedure 11 the formula for the π_{IJ} is still reasonably compact and involves no iteration. That for the Durbin-Hanurav Procedure 17 is a good deal more difficult to use, because the probabilities of selection change from draw to draw. In this regard it is distinctly less simple to use than Brewer's Procedure 8, Durbin's Method I (Procedure 9), Durbin's Grouped Method (Procedure 10), the Rao-Sampford Procedure 11, the Durbin-Sampford Procedure 12, Hanurav's Scheme B-A' 18, and the Hanurav-Vijayan Procedure 19. It is also easy to use Jessen's Procedures 35-36, as only simple calculations are involved.

Durbin (1967) has suggested the use of a randomization device in the estimation of variance for Durbin's Grouped Method (Procedure 10) which makes it slightly more simple to use than the others. Noting that the value of the coefficient $\left(\pi_1\pi_2\pi_{12}^{-1}-1\right)$ is unity for most pairs of units (see Table 3.1), he suggested that it might be dispensed with entirely by using the value $\left((y_1/\pi_1)-(y_2/\pi_2)\right)^2$ as the estimator with probability $\left\{\pi_1\pi_2\pi_{12}^{-1}-1\right\}$ whenever this value is less than unity, and with probability one when it is equal to or greater than unity. If it exceeds unity for any pair of population units, a bias is introduced, but it would appear to be small for most populations and is actually zero for the case $N = 9$, $n = 2$, $\pi_I = 0.04, 0.08, 0.08, 0.18, 0.18, 0.24, 0.30, 0.40, 0.50$, as may be seen from an inspection of the bottom left hand triangle in Table 3.1.

A glance at the other half of Table 3.1 suggests that the bias might be serious if Durbin's suggestion were used for the Rao-Sampford Procedure 11, for in this case 16 out of the 36 values of $\left(\pi_1\pi_2\pi_{12}^{-1}-1\right)$ exceed unity. (This device also increases the instability of the variance estimator slightly as will be noted in Section 3.7.) The gain in simplicity therefore amounts to avoiding the calculation for some but not all values of $\left((y_1/\pi_1)-(y_2/\pi_2)\right)^2$. (The use of the above randomization device in multistage sampling will be considered in Chapter 5.)

To summarize: Durbin's Grouped Method (Procedure 10) particularly with randomization device, is slightly easier to use than Brewer's Procedure 8, Durbin's Method I (Procedure 9), the Rao-Sampford Procedure 11, the Durbin-Sampford Procedure 12, Hanurav's Scheme B-A' 18, and the Hanurav-Vijayan Procedure 19, more or less in that order. The Durbin-Hanurav Procedure 17 is next most convenient. All the other relevant procedures, that is Raj's Variance Minimization Procedure 20, Narain's Procedure 7, the three Carroll-Hartley Procedures, Fellegi's Procedure 13, and even more the Random Systematic Procedure 2 and Grundy's Systematic Procedure 3, involve considerable calculations which indicate the need for a computer.

TABLE 3.1

Values of $\left(\pi_I\pi_J\pi_{IJ}^{-1}-1\right)$ for $N = 9$, $n = 2$;

$\pi_I = 0.04, 0.08, 0.08, 0.18, 0.18, 0.24, 0.30, 0.40, 0.50$

Values for Equivalence Class A

(Procedures 8, 9, 11, 12)

Values for Durbin's Grouped Method (Procedure 10)

π_I \ π_J	0.04	0.08	0.08	0.18	0.18	0.24	0.30	0.40	0.50
0.04	-	1.380	1.380	1.240	1.240	1.149	1.051	0.870	0.665
0.08	0.600	-	1.330	1.196	1.196	1.103	1.014	0.840	0.641
0.08	0.600	0.067	-	1.196	1.196	1.108	1.014	0.840	0.641
0.18	1.000	1.000	1.000	-	1.077	0.998	0.913	0.755	0.573
0.18	1.000	1.000	1.000	0.800	-	0.998	0.913	0.755	0.573
0.24	1.000	1.000	1.000	0.200	0.200	-	0.846	0.698	0.528
0.30	1.000	1.000	1.000	1.000	1.000	1.000	-	0.637	0.477
0.40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-	0.382
0.50	1.000	1.000	1.000	1.000	1.000	1.000	0.250	0.111	-

CASE 2: $n > 2$

Setting aside the procedures considered at the start of the Section, and Sinha's Procedures 42 and 43 for which the π_{IJ} are unspecified, the choice narrows down to Brewer's Procedure 8, the Rao-Sampford Procedure 11, the Durbin-Sampford Procedure 12, Fellegi's Procedure 13 and the Carroll-Hartley Rejective and Draw-by-Draw Procedures 14-15, and Choudhry's Procedure 49. The Rao-Sampford Procedure 11 and the Durbin-Sampford Procedure 12 do not involve iteration and are probably the simplest (though not very simple) for calculating π_{IJ}.

Procedures 8, 13-15 and 49 involve iteration and require a computer. Choudhry's Procedure 49 requires fewer calculations then the remainder of these.

Because of the general complexity of the variance estimation process (especially with $n > 2$) and the relative instability of the Sen-Yates Grundy variance estimator, even when the π_{IJ} are chosen to optimize it, an alternative variance estimator which does not depend on the π_{IJ} will be presented in Section 3.7.

3.6. EFFICIENCY OF THE HORVITZ-THOMPSON ESTIMATOR

Although this might be expected to figure importantly in the comparisons, the efficiency of the Horvitz-Thompson estimator varies so little in practice from one selection procedure to another that it can to all intents and purposes be ignored. Raj (1956b) produced Procedure 20 with a view to minimizing the variance of that estimator, but did so using the assumption that the Y_I and Z_I values were related by the deterministic equation

$$Y_I = \alpha + \beta Z_I \ , \tag{3.6.1}$$

where α and β were constant. The contribution to the variance from the βZ_I terms is then zero and the variance of y'_{HT} is

$$V(y'_{HT}) = \alpha^2 V(n'_{HT}) \ , \tag{3.6.2}$$

where n'_{HT} is the Horvitz-Thompson estimator of N , obtained by writing unity for each Y_I in (1.4.5).

Raj (1956b) minimized $V(n'_{HT})$ by linear programming. This characteristically results in an extreme solution in which one or more of the π_{IJ} takes the value zero. As already noted, zero values of π_{IJ} bias the Sen-Yates-Grundy variance estimator. (If a way through this dilemma is sought by setting the relevant π_{IJ} positive but very small, the Sen-Yates-Grundy variance estimator is unbiased but highly unstable.)

Further, model (3.6.1) is only one of many possible derivations from the ideal $Y_I = \beta Z_I$, and other models give different solutions to the problem of variance minimization. In particular if the model is (1.8.5) the expected variance of y'_{HT} , given by Bayless and Rao (1970), is

$$E^*V(y'_{HT}) = \sigma^2 (Z/n)^{2\gamma} \sum_{I=1}^{N} (1-\pi_I)\pi_I^{2\gamma-1} \ , \tag{3.6.3}$$

which does not depend on the π_{IJ} at all. Thus under (1.8.5) all exact selection procedures yield equally efficient Horvitz-Thompson estimators.

A similar conclusion may be reached by comparing the asymptotic variance formulae (1.8.1) and (1.8.2). The leading term in these variance formulae is of order N^2 . To order N^1 they are identical at

$$V(y'_{HT}) \doteq \sum_{I=1}^{N} \pi_I \left(1 - \frac{n-1}{n}\pi_I\right)\left((Y_I/\pi_I)-(Y/n)\right)^2 \ . \tag{1.8.4}$$

Again the conclusion is that (except for very small populations) the choice of selection procedure has only minimal influence on the efficiency of the Horvitz-Thompson estimator.

There is, however, an exception here in the case of the Ordered Systematic Procedure 1, which is not covered by any such asymptotic variance formulae. The variance of y'_{HT} using this procedure can depend critically on the particular ordering chosen, though not if the Y_I follow the model (1.8.5).

Equations (3.6.3) for the expected variance of y'_{HT} under the model (1.8.5), and (1.8.2) for the asymptotic variance of y'_{HT} to order N^1 , may be compared with the corresponding expressions for the Hansen-Hurwitz estimator using multinomial sampling (*ppswr*). These are

$$E^*V(y'_{HH}) = Z^2 \frac{\sigma^2}{n^{2\gamma}} \sum_{I=1}^{N} \left(1 - \frac{\mu_I}{n}\right)\mu_I^{2\gamma-1} \ , \tag{3.6.4}$$

and

$$V(y'_{HH}) = \sum_{I=1}^{N} \mu_I \left(\frac{Y_I}{\mu_I} - \frac{Y}{n}\right)^2 \ , \tag{3.6.5}$$

exactly.

In these expressions, μ_I is written for nP_I , so that μ_I is the expected number of appearances of the Ith population unit in sample. (For without replacement sampling the meanings of π_I and μ_I are identical.) The contribution of the Ith unit to the expected variance is smaller for the y'_{HT} by the factor $(1-\pi_I)/\left(1 - \frac{\mu_I}{n}\right)$. This is very close to the extra factor $\left(1 - \frac{n-1}{n}\pi_I\right)$ which appears in (1.8.4) but not in (3.6.4). Both are of the order of $(N-n)/(N-1)$, the finite population correction factor for simple random sampling.

Rao and Bayless (1969) and Bayless and Rao (1970) conducted both empirical and

semi-empirical studies of the efficiency of the Horvitz-Thompson estimator. They compared Procedures 8, 13, and 14 for $n = 2$; Procedures 11, 13, and 14 for $n = 3$; and Procedures 11 and 14 for $n = 4$. They found that with the exception of Procedure 13 for $n = 3$ (which rated as slightly less efficient in the empirical comparison) there were no appreciable differences in the performance of these procedures. The exception would be explained if the iterative algorithm for Procedure 13 failed to converge (see Footnotes 2 and 3 in Bayless and Rao (1970)).

3.7. UNBIASEDNESS AND STABILITY OF THE SEN-YATES-GRUNDY VARIANCE ESTIMATOR

Any procedure which allows any of the π_{IJ} to take the value zero - and this includes Procedures, 1, 2, 3, 20, 34, and 35 - can for that reason result in a biased Sen-Yates-Grundy variance estimator. (It has already been noted that the Ordered Systematic Procedure 1 allows so many of the π_{IJ} to take the value zero that the Sen-Yates-Grundy variance estimator is meaningless.) Sinha's Procedures 42 and 43 provide a special case here, in that the π_{IJ} are arbitrary provided only that they constitute a feasible set.

Rao and Bayless (1969) and Bayless and Rao (1970) used their empirical and semi-empirical populations to compare the stability of the Sen-Yates-Grundy variance estimator for some of the remaining procedures. The semi-empirical studies were based on the model (1.8.5) with normally distributed error terms, under which the expected variance of the Sen-Yates-Grundy variance estimator was taken to be

$$\left.\begin{aligned} E^*E\{v_{SYG}(y'_{HT}) - E^*V(y'_{HT})\}^2 &= E^*E\{v_{SYG}(y'_{HT})\}^2 - \{E^*V(y'_{HT})\}^2 , \\ \text{where} \quad E^*E\{v_{SYG}(y'_{HT})\}^2 &= 3\sigma^4(Z/2)^{4\gamma} \sum_{\substack{I,J=1 \\ J>I}}^{N} \frac{\pi_I\pi_J - \pi_{IJ}}{\pi_{IJ}} \left(\pi_I^{2\gamma-2} + \pi_J^{2\gamma-2}\right)^2 , \end{aligned}\right\} \qquad (3.7.1)$$

and $E^*V(y'_{HT})$ is given by (3.6.3), which is a function of the π_{IJ} . Some procedures which differ quite considerably in the actual process of selection end up with the same set $\{\pi_{IJ}\}$ for any given set $\{\pi_I\}$. The following equivalence classes were recognised for $n = 2$ in Chapter 1.

Equivalence Class A, Procedures 8, 9, 11, 12,
Equivalence Class B, Procedures 13, 14, 15, 16, 49.

For the case $n > 2$ the only equivalence class known is:

Equivalence Class B_1, Procedures 14, 15, 16.

In their 1969 paper, Rao and Bayless considered the case $n = 2$, comparing Procedures 8 and 13 (in Equivalence Classes A and B respectively) and Procedure 17. Bayless and Rao (1970) considered the case $n = 3$, comparing Procedures 11, 13 and 14 and the case $n = 4$ comparing Procedures 11 and 14 only (the last being in Equivalence Class B_1).

Their empirical findings were that for $n = 2$ the three procedures had about equally stable variance estimators. Procedure 17 performed rather better for two of the 20 natural populations, but worse in six of the seven artifical populations. For $n = 3$ (and 4) the stabilities of the three (two) variance estimators were virtually identical.

In their semi-empirical studies they found that, for $n = 2$, Procedure 13 (Equivalence Class B) was consistently more stable than Procedure 8 (Equivalence Class A) but that the gains were small. The stabilities of Procedures 13 and 17 were essentially equal, but Procedure 13 was consistently more stable for $\gamma = 1$. Procedure 17 appeared to be consistently more stable than Procedures 8 and 13 for $\gamma = 0.75$. . For $n = 3$, Procedure 13 was found to be consistently less stable than the other two, but there is some doubt as to whether the iterative algorithm for Procedure 13 converged (see Section 3.6).

Brewer and Hanif (1969a) carried out similar semi-empirical studies for the case $N = 4$, $n = 2$, $\pi_I = 0.2, 0.4, 0.6, 0.8$ only. The results, shown in Table 3.2, compare values

$$E^*E\{v_{SYG}(y'_{HT}) - E^*V(y'_{HT})\}^2 / \{E^*V(y'_{HT})\}^2 ,$$

that is to say the relative expected variances of the Sen-Yates-Grundy variance estimator, for all the generally applicable exact procedures with the exception of Sinha's Procedures 42-43 for which the π_{IJ} are arbitrary.

In Table 3.2 all the draw-by-draw and rejective procedures can be seen to approximate the Optimization of Stability Procedure 33 for $\gamma = 1$. The values of the π_{IJ} used to construct Table 3.2 together with some others are given in Table 3.3.

Durbin's Grouped Method (Procedure 10) could not be compared with the other procedures for so small population. Table 3.5 shows the comparison between Equivalence Class A and Procedure 10 both without and with the randomization device mentioned in Section 3.4.

For most of the range $0.50 \leq \gamma \leq 1$, the Rao-Sampford Procedure 12 (together with the other procedures in the Equivalence Class A) has a more stable variance estimator than Durbin's Grouped Method (Procedure 10). The reason for this seems to be that the values of $\left(\pi_I\pi_J\pi_{IJ}^{-1}\, 1\right)$ for Equivalence Class A decrease with the π_I and π_J , particularly the larger of the pair, in much the same way as the values of this

TABLE 3.2

Values of relative expected variances of Sen-Yates-Grundy variance estimators for $N = 4$, $n = 2$, $\pi_I = 0.2, 0.4, 0.6, 0.8$

Procedure or Equivalence Class	Relative Expected Variances of $v_{SYG}(y'_{HT})$		
	$\gamma = 0.50$	$\gamma = 0.75$	$\gamma = 1$
Optimization of Stability (Procedure 33) for			
$\gamma = 0.50$	6.44	5.96	5.63
$\gamma = 0.75$	6.58	5.86	5.33
$\gamma = 1.00$	7.02	5.98	5.21
Random systematic 2*	10.35	10.65	10.96
Grundy's systematic 3	10.91	8.55	6.85
Equivalence Class A	8.17	6.61	5.43
Equivalence Class B	7.50	6.22	5.27
Narain 7	7.60	6.27	5.29
Durbin-Hanurav 17	7.53	6.24	5.28
Hanurav's Scheme B-A' 18	7.32	6.33	5.64
Hanurav-Vijayan 19	7.43	6.19	5.27

* For this example Jessen's Procedure 36 is equivalent to Procedure 2.

TABLE 3.3

Values of joint probability of inclusion in sample of pairs of units for $N = 4$, $n = 2$, $\pi_I = 0.2, 0.4, 0.6, 0.8$

Procedure or Equivalence Class	Joint Probabilities of Inclusion					
	$I = 1$ $J = 2$	$I = 1$ $J = 3$	$I = 1$ $J = 4$	$I = 2$ $J = 3$	$I = 2$ $J = 4$	$I = 3$ $J = 4$
Optimization of Stability (Procedure 33) for						
$\gamma = 0.50$	0.0422	0.0588	0.0990	0.0990	0.2588	0.4422
$\gamma = 0.75$	0.0386	0.0559	0.1055	0.1055	0.2559	0.4386
$\gamma = 1.00$	0.0344	0.0535	0.1121	0.1121	0.2535	0.4344
Random systematic 2	0.0667	0.0667	0.0667	0.0667	0.2667	0.4667
Grundy's systematic 3	0.0333	0.0333	0.1333	0.1333	0.2333	0.4333
Equivalence Class A	0.0277	0.0535	0.1188	0.1188	0.2535	0.4277
Equivalence Class B	0.0311	0.0530	0.1158	0.1158	0.2530	0.4311
Narain 7	0.0306	0.0531	0.1163	0.1163	0.2531	0.4306
Durbin-Hanurav 17	0.0323	0.0505	0.1172	0.1172	0.2505	0.4323
Hanurav's Scheme B-A'	0.0444	0.0444	0.1111	0.1111	0.2444	0.4444
Hanurav-Vijayan 19	0.0333	0.0500	0.1167	0.1167	0.2500	0.4333
Chromy 50*	0.0480	0.0987	0.0533	0.0533	0.2987	0.4480

* With the population units arranged in ascending order of size for step (1).

TABLE 3.4

Values of relative expected variances of the Sen-Yates-Grundy variance estimator for $N = 9$, $n = 2$, $\pi_I = 0.04, 0.08, 0.08, 0.18, 0.18, 0.24, 0.30, 0.40, 0.50$ assuming normality of the ε_I

Procedure or Equivalence Class	Related expected variance of $v_{SYG}(y'_{HT})$		
	$\gamma = 0.50$	$\gamma = 0.75$	$\gamma = 1$
Equivalence Class A	4.74	3.01	2.31
Durbin's Grouped Method (Procedure 10) (without randomization device)	4.63	3.36	3.04
Durbin's Grouped Method (Procedure 10) (with randomization device)	4.74	3.46	3.16

TABLE 3.5

Values of π_{IJ} for $N = 9$, $n = 2$;

$\pi_I = 0.04, 0.08, 0.08, 0.18, 0.18, 0.24, 0.30, 0.40, 0.50$

Values for Equivalence Class A

Values for Durbin's Grouped Method (Procedure 10)

π_I \ π_J	0.04	0.08	0.08	0.18	0.18	0.24	0.30	0.40	0.50
0.04	-	.0013	.0013	.0032	.0032	.0045	.0059	.0086	.0120
0.08	.0020	-	.0027	.0066	.0066	.0041	.0119	.0174	.0244
0.08	.0020	.0060	-	.0066	.0066	.0091	.0119	.0174	.0244
0.18	.0036	.0072	.0072	-	.0156	.0216	.0282	.0210	.0572
0.18	.0036	.0072	.0072	.0180	-	.0216	.0282	.0410	.0572
0.24	.0048	.0096	.0096	.0360	.0360	-	.0390	.0565	.0786
0.30	.0060	.0120	.0120	.0270	.0270	.0360	-	.0733	.1026
0.40	.0080	.0160	.0160	.0360	.0360	.0290	.0600	-	.1447
0.50	.0100	.0200	.0200	.0450	.0450	.0600	.1200	.1800	-

coefficient did in the earlier example, whereas for Durbin's Grouped Method (Procedure 10) they are fixed at unity for most pairs of units. This difference is shown up in Table 3.5, in which for convenience the values for Equivalence Class A have been entered above the main diagonal and those for Durbin's Grouped Method (Procedure 10) below it.

The π_{IJ} used in the construction of Table 3.4 are shown in Table 3.5. The values of the factor $\left(\pi_I \pi_J \pi_{IJ}^{-1} - 1\right)$ were given in Table 3.1.

Sinha's Procedures 42 and 43 are atypical in that the π_{IJ} are not products of

the selection process, but arbitrary inputs. One simple way of choosing a feasible set of π_{IJ} as an input to Sinha's Procedures is to set

$$\left.\begin{aligned} \pi_{IJ} &= A\pi_I\pi_J + B(\pi_I+\pi_J) + C(\pi_I^2+\pi_J^2) , \\ \text{where} \quad A &= (n^2)/\left(n^2 - \sum_{J=1}^{N}\pi_J^2\right) , \\ B &= -\left(n\sum_{J=1}^{N}\pi_J^2\right) / \left\{\left(n^2 - \sum_{J=1}^{N}\pi_J^2\right)(N-2)\right\} , \\ \text{and} \quad C &= (n^2)/\left\{\left(n^2 - \sum_{J=1}^{N}\pi_J^2\right)(N-2)\right\} . \end{aligned}\right\} \qquad (3.7.2)$$

These π_{IJ} will not always be non-negative, but for medium to large values of N approach proportionality to $\pi_I\pi_J$. When this proportionality holds, all the factors $\left(\pi_I\pi_J\pi_{IJ}^{-1}-1\right)$ are equal, and this corresponds to a reasonably, though not optimally, stable Sen-Yates-Grundy variance estimator. Departures from optimality can, however, be serious when N is small, and for the case $N = 4$, $n = 2$, $\pi_I = 0.2, 0.4, 0.6, 0.8$ the value of π_{12} is actually zero.

A more generally satisfactory set of values of π_{IJ} is given by the formula

$$\begin{aligned} \pi_{IJ} &= \frac{n-1}{n}\left\{\pi_I\pi_J + \left(\pi_I^2\pi_J^2\right)/\left(\sum_{K=1}^{N}\pi_K^2\right) + \left(\pi_I^4\pi_J^4\right)/\left(\sum_{K=1}^{N}\pi_K^2\sum_{K=1}^{N}\pi_K^4\right) + \ldots\right\} \\ &= (n-1)\sum_{r=0}^{\infty}\left(\pi_I^{2r}\pi_J^{2r}\right) / \left(\prod_{i=0}^{r}\sum_{K=1}^{N}\pi_K^{2i}\right) . \end{aligned} \qquad (3.7.3)$$

This summation converges rapidly even when one or two values of the π_I are close to unity, each term being less than half the preceding one. The resulting π_{IJ} are necessarily positive, and appear to be close to the optimal values when $\gamma = \frac{1}{2}$. For the case $N = 4$, $n = 2$, $\pi_I = 0.2, 0.4, 0.6, 0.8$ this formula yields $\pi_{12} = 0.0427$, $\pi_{13} = 0.0662$, $\pi_{14} = \pi_{23} = 0.0911$, $\pi_{24} = 0.2662$, $\pi_{34} = 0.4427$ (*cf.* Table 3.3). However when two of the π_I are close to unity (3.7.3) may not result in a feasible set of π_{IJ}.

A third possible choice of π_{IJ} can be made as follows. Multiply each of the given π_I values by $2/n$. The scaled down values of π_I then sum to 2 and

corresponding π_{IJ} can be calculated using the Brewer-Durbin-Sampford formula (2.2.9). If these are then multiplied up by $n(n-1)/2$, the resulting π_{IJ} are always positive, but again in extreme cases may not constitute a feasible set.

It is of interest to compare the stability of the Sen-Yates-Grundy estimator for sampling *πpswor* with that of the usual variance estimator for multinomial sampling. For the latter, the coefficients of $\left\{(y_1/\pi_1)-(y_2/\pi_2)\right\}^2$ are all unity. (In this respect Durbin's Grouped Method (Procedure 10) represents an approach towards multinomial sampling. In view of the explicit use of multinomial sampling as part of this procedure, this result is not surprising.) Using the same assumptions as before, the relative expected variances of the ordinary variance estimator for multinomial sampling are, for this example, 3.98 for $\gamma = \frac{1}{2}$, 2.78 for $\gamma = 3/4$ and 2.57 for $\gamma = 1$. These small values are somewhat illusory, because the variance itself is much larger when sampling is with replacement. To get a more meaningful comparison, the expected variance of the ordinary variance estimator may be divided by $\left\{E^*V(y'_{HT})\right\}^2$. This quotient will be referred to as the *comparison relative expected variance* (CREV) of the ordinary multinomial variance estimator. In this example the CREV takes the values 5.13 for $\gamma = \frac{1}{2}$, 3.93 for $\gamma = 3/4$ and 3.91 for $\gamma = 1$. Referring back to Table 3.5 it will be seen that all these values are higher than the corresponding values for both the Rao-Sampford Procedure 11 and Durbin's Group Method 10.

For the earlier example $\left(N = 4,\ n = 2;\ \pi_I = 0.2, 0.4, 0.6, 0.8\right)$ the relative expected variances of the ordinary variance estimator for multinomial sampling were 3.81 for $\gamma = \frac{1}{2}$, 3.39 for $\gamma = 3/4$ and 3.29 for $\gamma = 1$. The CREV's, however, were 8.56 for $\gamma = \frac{1}{2}$, 8.90 for $\gamma = 3/4$, and 10.06 for $\gamma = 1$. Comparing these values with those in Table 3.2 we find for this example, as for the other, not only that the expected variance is smaller when the sample is drawn without replacement but also that, provided a draw-by-draw or rejective procedure is used, this smaller expected variance is absolutely (though not relatively) more accurately estimated. This is particularly true for the larger values of γ.

Although in this comparison the Sen-Yates-Grundy variance estimator comes out reasonably well, it is still unstable by comparison with some of the variance estimators used with special estimators of total which will be encountered in Chapter 4.

In view of the difficulties encountered in attempting to evaluate the π_{IJ} (particularly for $n > 2$) and of the relative instability of the Sen-Yates-Grundy variance estimator the following approximate variance estimator may be preferred;

$$v_{APP}(y'_{HT}) = \frac{n}{n-1}\left[1 - \left(\sum_{I=1}^{N} \pi_I^{2\gamma}\right)/\left(\sum_{I=1}^{N} \pi_I^{2\gamma-1}\right)\right] \sum_{i=1}^{n} \left((y_i/\pi_i)-(y'_{HT}/n)\right)^2 , \quad (3.7.4)$$

where γ is chosen to be the best available 'guesstimate' of the parameter γ of model (1.8.5). Fortunately, the value of (3.7.4) is not critically dependent on the value of γ chosen.

The rationale behind this estimator is as follows. Under model (1.8.5),

$$E^*E \frac{n}{n-1} \sum_{i=1}^{n} \left((y_i/\pi_i)-(y'_{HT}/n)\right)^2 = \sigma^2(Z/n)^{2\gamma} \sum_{I=1}^{N} \pi_I^{2\gamma-1} . \quad (3.7.5)$$

This is the larger of the two terms in the expected variance of y'_{HT} given in (3.6.3). The ratio of the smaller to the larger term is $\sum_{I=1}^{N} \pi_I^{2\gamma} / \sum_{I=1}^{N} \pi_I^{2\gamma-1}$ and corresponds to the expression n/N found in the finite population correction for equal probability sampling. When $\gamma = \frac{1}{2}$ the ratio is n/N precisely, and although it increases with γ it does not do so rapidly. The manner in which the factor $1 - \sum_{I=1}^{N} \pi_I^{2\gamma} / \sum_{I=1}^{N} \pi_I^{2\gamma-1}$ functions as a finite population correction is further exhibited by remarking that for multinomial sampling the usual unbiased variance estimator may be written

$$v(y'_{HH}) = \frac{n}{n-1} \sum_{i=1}^{n} \left(\frac{y_i}{\mu_i} - \frac{y'_{HH}}{n}\right)^2 , \quad (3.7.6)$$

where $\mu_i = n\pi_i$ is the expected number of appearances in sample of the population unit selected at the ith sample draw and corresponds to the π_i of (3.7.4). When γ is completely unknown, the assumption that $\gamma = \frac{1}{2}$ gives the conservative correction factor $(1-n/N)$. For most populations the value of γ is found to lie between 0.6 and 0.9, and the value 0.75 will usually be a reasonable 'guesstimate'.

3.8. ROTATABILITY

When a number of surveys are to be made at intervals using the same or nearly the same questionnaire, there can be advantages in rotating the sample; that is, in having a regular programme whereby new units are selected to replace old units that have been in the sample for a specified number of surveys.

The advantage of rotation is that the estimate of total can be improved by using information from past periods (Patterson, 1950). A partial overlap between the previous and current samples is required to exploit this improvement. If, however, the

aim is to estimate the changes in total between surveys, it is theoretically best to retain an identical sample. Nevertheless even in this case there are practical advantages in rotation. Objections to keeping the sample unchanged include the following.

(i) Respondents from the first few surveys may refuse to co-operate if asked similar questions on too many successive occasions.

(ii) Respondents who took the trouble to give accurate answers in the first few surveys may become careless. They may for instance continue to give the same answers as before, even though their situation has changed. Interviewers may also become careless in a very similar fashion.

(iii) Respondents who reported candidly to a strange interviewer in the first instance may be reluctant to admit a worsening of their situation, especially to an interviewer who is steadily becoming a more familiar figure.

(iv) Respondents who would remain in a given situation if not questioned about it at regular intervals may be stimulated by this questioning to take steps to change it.

These and other related phenomena are known collectively as *sample fatigue*. All of them tend to diminish the representative character of the sample data. Hence it is usual, in repeating surveys, to arrange that portions of the sample be replaced at regular intervals by new sample units, so that none remains in the sample indefinitely.

Rotation is more often important in multistage designs than in single-stage, but can still be treated quite conveniently in terms of single-stage sampling. Fellegi discusses the two possible alternatives for multistage sampling; a third is mentioned here also.

Alternative I is "the exhaustion of the P.S.U.'s (primary sampling units)". This means that a selected unit is replaced when all its available final stage units have been sampled. (In considering the time when the first rotation need be made, the number of "available units" must be determined by a random mechanism. If all final stage units are regarded as "available", rotation introduces a time-dependent bias, initially in favour of large higher stage units.)

Alternative II involves rotating a P.S.U. every rth survey, the P.S.U.'s selected in the most recent n selections constituting the sample. Selected units remain in the sample for a constant period regardless of their measures of size.

Alternative III is a very crude form of rotation (not mentioned by Fellegi) in which the selected samples are retained in a certain number of strata, and reselected in the remainder.

Fellegi's Procedure 13 is particularly appropriate for Alternative II, since the probability of selection of each population unit is proportional to size at every draw. Alternative II may also be used with any other procedure applicable to $n > 2$ in one of two ways.

(a) If it is recognised beforehand that rotation will be necessary, a larger sample than that immediately necessary can be selected, and the order of these sample units randomized. (Randomization is unnecessary if the procedure of selection is symmetric.) The first n in this random order then constitutes the initial sample and rotation proceeds by dropping the first unit for the $(n+1)$st , the second for the $(n+2)$nd , and so on. When the last sample unit has been used, rotation can still proceed in a limited sense by returning to the first unit dropped from the initial sample.

(b) If the initial sample was already selected and used before the need for rotation was recognised, the larger sample described under (a) must first be tested as to whether it contains all the initial sample units. If not, further larger samples must be selected until one is found which does contain all these units. The initial sample is then taken out and its order randomized. The remaining units of the larger sample are also put in random order. Rotation proceeds by dropping the first unit in random order in the initial sample in favour of the first unit in random order in the remainder, and so on. When the remaining units have been exhausted, rotation can only proceed by returning to the first unit dropped from the initial sample.

A limit to the proportion of the population around which rotation is possible, using Alternative II, is provided by the fact that no sample can be selected larger than n_{max} , where n_{max} is the largest integer less than or equal to Z/Z_{max} . Unless n_{max} is appreciably larger than n , this can be a serious limitation.

If rotation of a self-weighting multistage sample is occuring at the lowest stage of sampling, the minimum rate of rotation for the selected higher stage units is set by the period taken to exhaust the smallest such unit selected. This further limits the extent of rotation possible using Alternative II. With Alternative I, rotation can occur around the entire population.

Alternative III is a very crude form of rotation, and does not even guarantee the certain replacement of any sample unit. If the strata in which reselection is to take place are selected randomly, an estimate of the movement in the estimand can be obtained from the remainder. However since this involves the use of strata to represent other strata, this estimate is not likely to be at all accurate.

To sum up, rotatability depends on which alternatives can be used with which methods. The Systematic Procedures 1-3 and 48 are the only ones for which Alternative I is possible. Alternative II can be used very easily with Fellegi's Procedure and the Carroll-Hartley Procedures 14 and 15, and with somewhat more difficulty with any other

method valid for $n > 2$. Alternative III is the only possibility for any procedure limited to $n = 2$.

3.9. SUMMARY

In Tables 3.6 and 3.7 a summary of the properties of some procedures compared in this Chapter is given for $n = 2$ and $n > 2$ respectively.

The principal conclusions which may be drawn are as follows:

(a) The Systematic Procedures 2-3 score highly on simplicity of selection and ease of rotation, but relatively poorly on most other counts, particularly those relating to variance estimation.

(b) Within Equivalence Class A, the Rao-Sampford Procedure 11 is particularly good for $n > 2$ or for rotation with $n = 2$. For $n = 2$ and no rotation, a draw-by-draw method (Brewer's Procedure 8, Durbin's Method I (Procedure 9) or the Durbin-Sampford (Procedure 12)) will probably be slightly more convenient.

(c) Durbin's Grouped Method I (Procedure 10) has a slight advantage over the Equivalence Class A procedures for simplicity in variance estimation, but at the cost of some stability in the variance estimator.

(d) The Carroll-Hartley Rejective Procedure 14 is superior to the Rao-Sampford Procedure 11 only in that it is unnecessary to resort to randomization when using Alternative II for rotation. The Rao-Sampford Procedure 11 is simpler both for selection and possibly for variance estimation. Otherwise there is no difference of any importance.

(e) Fellegi's Procedure 13 has the further advantage over the Rao-Sampford Procedure 11 that oversampling is not required for rotation. However, the iterative selection calculations are stated by Carroll and Hartley to be less simple than theirs and may not converge for $n > 2$. Choudhry's Procedure 49 has the advantage that fewer calculations are required. The Rao-Sampford Procedure 11 remains the simplest of the four for selection and possibly also for variance estimation.

(f) The Hanurav-Vijayan Procedure 19 is a reasonable alternative to Equivalence Class A procedures for $n = 2$. The same could perhaps be said for Hanurav's Scheme B-A' 18 also, but the Durbin-Hanurav Procedure 17 is decidedly less simple to handle both for selection and for estimation of variance.

TABLE 3.6

Summary of properties of selection procedures for $n = 2$

Procedures	Type	Equivalence class (if any)	Is $v_{SYG}(y'_{HT})$ unbiased?	Stability of $v_{SYG}(y'_{HT})$	Nature of selection process	Requirement for calculation of $\{\pi_{IJ}\}$	Can alternative I or II be used for rotation?
Random syst 2	*syst*	-	not always	variable	randomization plus systematic	*j p enum*	I
Grundy's syst 3	*syst*	-	not always	variable	partial randomization plus systematic	*j p enum*	I
Narain 7	*d by d*	-	yes	near optimum	iterative algorithm	same iterative algorithm	No
Brewer 8	*d by d*	A	yes	near optimum	simple working probabilities	simple closed formula	II[3]
Durbin (I) 9	*d by d*	A	yes	near optimum	simple conditional probabilities	simple closed formula	No[4]
Durbin (Grouped)	*d by d*	-	yes[1]	fair	grouping plus Procedure 9	simple closed formula	No
Rao-Sampford 11	*rej*	A	yes	near optimum	simple working probabilities	simple closed formula	II[3]
Durbin-Sampford 12	*d by d*	A	yes	near optimum	simple working probabilities	simple closed formula	No
Fellegi 13 Choudhry 49 Carroll-Hartley 14	*d by d*	B	yes	near optimum	iterative algorithm	same iterative algorithm	II
Carroll-Hartley 15	*rej*	B B_1	yes	near optimum	iterative algorithm	iterative algorithm	II[2]

72

Table 3.6 (continued)

Procedures	Type	Equivalence class (if any)	Is $v_{SYG}(y'_{HT})$ unbiased?	Stability of $v_{SYG}(y'_{HT})$	Nature of selection process	Requirement for calculation $\{\pi_{IJ}\}$	Can alternative I or II be used for rotation?
Durbin-Hanurav 17	*rej*	-	yes	near optimum	simple working probabilities	closed formula	No
Hanurav B-A' 18	*d by d*	-	yes	near optimum	simple working probabilities	simple closed formula	No
Hanurav-Vijayan 19	*d by d*	-	yes	near optimum	simple working probabilities	simple closed formula	No
Raj's variance minimization 20	*w.s.*	-	no	poor	linear programming	linear programming	No
Jessen (2) 35	*w.s.*	-	not always	variable	decremented	simple calculation	No
Jessen (3) 36	*w.s.*	-	not always	variable	decremented	simple calculation	No
Jessen (4) 37	*w.s.*	-	yes	variable	trial and error	trial and error	No
Sinha (Ext) 42	*w.s.*	-	depends on $\{\pi_{IJ}\}$	depends on $\{\pi_{IJ}\}$	examination of possible samples of $N - 2$ units	no calculations required	II[3]
Sinha (Red) 43	*w.s.*	-	depends on $\{\pi_{IJ}\}$	depends on $\{\pi_{IJ}\}$	cumulation of the π_{IJ}	no calculations required	II[3]
Chromy 50	*d by d*	-	yes (?)	variable (?)	sequential	*j p enum*	II[3]

73

[1] If the randomization device is used to simplify the estimation of variance, $v_{SYG}(y'_{HT})$ will not always be unbiased, and there will be a slightly additional reduction in its stability.

[2] If Alternative II is used, oversampling is needed.

[3] If Alternative II is used, oversampling and random ordering are needed.

[4] If Alternative II is used for rotation with Durbin's Method I it may break down becasue this method is not generally applicable for $n > 2$.

TABLE 3.7

Summary of properties of selection procedures for $n > 2$

Procedures	Type	Equivalence class (if any)	Is $v_{SYG}(y'_{HT})$ unbiased?	Stability of $v_{SYG}(y'_{HT})$	Nature of selection process	Requirement for calculation of $\{\pi_{IJ}\}$	Can Alternative I or II be used for rotation?
Random syst 2	*syst*	-	not always	variable	randomization plus systematic	*j p enum*	I
Grundy's syst 3	*syst*	-	not always	variable	partial randomization plus systematic	*j p enum*	I
Brewer 8	*d by d*	-	yes	near optimum	simple working probabilities	complicated recursive	II[3]
Rao-Sampford 11	*rej*	-	yes	near optimum	simple working probabilities	closed formula	II[3]
Fellegi 13	*d by d*	-	yes	near optimum	iterative algorithm	same iterative algorithm	II
Carroll-Hartley 14	*rej*	B_1	yes	near optimum	iterative algorithm	same iterative algorithm	II[2]
Carroll-Hartley 15	*d by d*	B_1	yes	near optimum	iterative algorithm	same iterative algorithm	II[2]
Jessen (2) 35 Jessen (3) 36	*w.s.*	-	not always	variable	decremented	simple calculation	No
Sinha (Ext) 42	*w.s.*	-	depends on $\{\pi_{IJ}\}$	depends on $\{\pi_{IJ}\}$	examination of possible samples of $N - 2$ units	no calculations required	II[3]
Sinha (Red) 43	*w.s.*	-	depends on $\{\pi_{IJ}\}$	depends on $\{\pi_{IJ}\}$	examination of possible samples of 2 units	no calculations required	II[3]

Table 3.7 (continued)

Procedures	Type	Equivalence class (if any)	Is $v_{SYG}(y'_{HT})$ unbiased?	Stability of $v_{SYG}(y'_{HT})$	Nature of selection process	Requirement for calculation of $\{\pi_{IJ}\}$	Can Alternative I or II be used for rotation?
Choudhry 49	*d by d*	-	yes	near optimum	iterative algorithm	same iterative algorithm	II[3]
Chromy 50	*d by d*	-	yes (?)	variable (?)	sequential	*j p enum*	II[3]

[2] If Alternative II is used, oversampling is needed.

[3] If Alternative II is used, oversampling and random ordering are needed.

(g) The Jessen Procedures 35 and 36 are simple for selection but cannot be rotated easily and score poorly on most other counts.

(h) Sinha's Procedures 42 and 43 look particularly promising for moderate values of n. Since the π_{IJ} are arbitrary they can be chosen to minimize (or using expression (3.7.3) to come close to minimizing) the variance of the Sen-Yates-Grundy variance estimator. For large values of n the procedures become unmanageable.

(j) Because the Systematic Procedures 2 and 3 are so convenient on all counts other than variance estimation, the approximate variance formula (3.7.4) which does not depend on the π_{IJ} may be used to remedy this deficiency.

CHAPTER 4

SELECTION PROCEDURES USING SPECIAL ESTIMATORS

4.1 INTRODUCTION

In Chapter 3 a comparison was made of those selection procedures for which the Horvitz-Thompson estimator possessed the ratio estimator property. It was mentioned, however, in Section 1.7 that certain special estimators had also been devised for use with particular selection procedures, and that in the context of these procedures they also possessed the ratio estimator property. In this Chapter the performance of these special estimators will be compared in the context of their appropriate selection procedures; that is,

(i) Das's estimator with Procedure 4,

(ii) Raj's and Murthy's estimators with Procedure 4,

(iii) the Rao-Hartley-Cochran (RHC) estimator with Procedure 25,

(iv) unbiased and ratio estimators for Poisson sampling with Procedure 27,

(v) unbiased and ratio estimators for Modified Poisson Sampling with Procedure 38,

(vi) unbiased and ratio estimators for Collocated Sampling with Procedure 39, and

(vii) Lahiri's estimator with Procedures 45-46.