CHAPTER 1

PROBABILITY SAMPLING FROM A FINITE UNIVERSE

1.1 INTRODUCTION

A large fraction of the quantitative information that we receive about our economy and our community comes from sample surveys. Statistical agencies of national governments regularly report estimates for items such as unemployment, poverty rates, crop production, retail sales, and median family income. Some statistics may come from censuses, but the majority are based on a sample of the relevant population. Less visible statistics are collected by other entities for business decisions, city planning, and political campaigns. National polls on items beyond politics are regularly reported in newspapers. These reports are so common that few reflect on the fact that almost all people believe that something interesting and (or) useful can be said about a nation of 300 million people on the basis of a sample of a few thousand. In fact, the concept that a probability sample can be so used has only been accepted by the scientific community for about 60 years. In this book we study the statistical basis for obtaining information from samples.

Sampling Statistics. By Wayne A. Fuller

In this chapter we develop a probabilistic framework for the study of samples selected from a finite population. Because the study of estimators often requires the use of large-sample approximations, we define sequences of populations and samples appropriate for such study.

1.2 PROBABILITY SAMPLING

Consider a finite set of elements identified by the integers $U = \{1, 2, \ldots, N\}$. The set of identifiers, sometimes called *labels*, can be thought of as forming a list. The existence of such a list, a list in which every element is associated with one and only one element of the list, is the cornerstone of probability sampling. The list is also called the *sampling frame*. In practice, the frame takes many forms. For example, it may be a list in the traditional sense, such as the list of employees of a firm or the list of patients in a hospital. It is sometimes the set of subareas that exhaust the geographic area of a political unit such as a city or state.

Associated with the jth element of the frame is a vector of characteristics denoted by $\mathbf{y}_j$. In all of our applications, the $\mathbf{y}_j$ are assumed to be real valued. The entire set of N vectors is denoted by $\mathcal{F}$. The set is called a *finite population* or a *finite universe*. A sample is a subset of the elements. Let A denote the set of indices from U that are in the sample. In statistical sampling the interest is in the selection of samples using probability rules such that the probability characteristics of the set of samples defined by the selection rules can be established. Let $\mathcal{A}$ denote the set of possible samples under a particular probability procedure. A person who wishes to obtain information about a population on the basis of a sample must develop a procedure for selecting the sample.

The terms *random samples* and *probability samples* are both used for samples selected by probability rules. Some people associate the term *random sampling* with the procedure in which every sample has the same probability and every element in the population has the same probability of appearing in the sample.

1.2.1 Basic properties of probability samples

In this section we present some basic properties of statistics constructed from probability samples. In the methods of this section, the probabilistic properties depend only on the sampling procedure. The population from which the samples are selected is fixed. Let A be a subset of U and let $\mathcal{A}$ be the collection of subsets of U that contains all possible samples. Let $P[A = a]$ denote the probability that a, $a \in \mathcal{A}$, is selected.

Definition 1.2.1. A *sampling design* is a function $p(\cdot)$ that maps a to $[0, 1]$ such that $p(a) = P[A = a]$ for any $a \in \mathcal{A}$.

A set of samples of primary importance is the set of all possible samples containing a fixed number of distinct units. Denote the fixed size by n. Then the number of such samples is

$$\binom{N}{n} = \frac{N!}{(N-n)!n!}, \tag{1.2.1}$$

where $N! = 1 \times 2 \times \cdots \times N$.

A probability sampling scheme for samples of fixed size n assigns a probability to each possible sample. *Simple random nonreplacement sampling* assigns equal probability to each possible sample. We may occasionally refer to such samples as simple random samples. The *inclusion probability* for element i is the sum of the sample probabilities for all samples that contain element i; that is,

$$\pi_i = P(i \in A) = \sum_{a \in A_{(i)}} p(a),$$

where $A_{(i)}$ is the set of samples that contain element i.

The terms *selection probability*, *probability of selection*, and *observation probability* are also used. In simple random nonreplacement sampling, element i appears in

$$\binom{1}{1}\binom{N-1}{n-1} \tag{1.2.2}$$

samples. If every sample has equal probability, the probability of selecting element i is

$$\pi_i = \left[\binom{N}{n}\right]^{-1}\binom{N-1}{n-1} = \frac{n}{N}. \tag{1.2.3}$$

In discussing probability sampling schemes, we define indicator variables to identify those elements appearing in the sample. Let I_i be the indicator variable for element i. Then

$$\begin{aligned} I_i &= 1 \quad \text{if element } i \text{ is in the sample} \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{1.2.4}$$

Let $\mathbf{d} = (I_1, I_2, \ldots, I_N)$ be the vector of random variables. The probabilistic behavior of functions of the sample depends on the probability distribution of $\mathbf{d}$. The sampling design specifies the probability structure of $\mathbf{d}$, where the inclusion probability for element i is the expectation of I_i,

$$\pi_i = E\{I_i\}. \tag{1.2.5}$$

With this notation, the sum of characteristic y for the elements in the sample is

$$\text{sample sum} \quad = \quad \sum_{i=1}^{N} I_i y_i. \tag{1.2.6}$$

The set A is the set of indices *appearing* in the sample. Thus,

$$A \quad = \quad \{i \in U : I_i = 1\}. \tag{1.2.7}$$

Then the sample sum of (1.2.6) can be written

$$\sum_{i=1}^{N} I_i y_i = \sum_{i \in A} y_i. \tag{1.2.8}$$

The joint inclusion probability, denoted by π_{ik}, for elements i and k is the sum of sample probabilities for all samples that contain both elements i and k. In terms of the indicator variables, the joint inclusion probability for elements i and k is

$$\pi_{ik} \quad = \quad E\{I_i I_k\}. \tag{1.2.9}$$

For simple random nonreplacement sampling, the number of samples that contain elements i and k is

$$\binom{1}{1}\binom{1}{1}\binom{N-2}{n-2} \tag{1.2.10}$$

and

$$\pi_{ik} \quad = \quad [N(N-1)]^{-1} n(n-1). \tag{1.2.11}$$

The number of units in a particular sample is

$$n \quad = \quad \sum_{i=1}^{N} I_i, \tag{1.2.12}$$

and because each I_i is a random variable with expected value π_i, the expected sample size is

$$E\{n\} \quad = \quad \sum_{i=1}^{N} E\{I_i\} \quad = \quad \sum_{i=1}^{N} \pi_i. \tag{1.2.13}$$

Also, the variance of the sample size is

$$V\{n\} = V\left\{\sum_{i=1}^{N} I_i\right\} = \sum_{i=1}^{N}\sum_{k=1}^{N}(\pi_{ik} - \pi_i\pi_k),$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{N}\pi_{ik} - \left(\sum_{i=1}^{N}\pi_i\right)^2, \quad (1.2.14)$$

where $\pi_{ii} = \pi_i$. If $V\{n\} = 0$, we say that the design is a *fixed sample size* or *fixed-size design.* It follows from (1.2.14) that

$$\sum_{\substack{i=1 \\ i\neq k}}^{N}\sum_{k=1}^{N}\pi_{ik} = n^2 - n = n(n-1) \quad (1.2.15)$$

for fixed-size designs. Also, for fixed-size designs,

$$\sum_{k\in U: i\neq k}\pi_{ik} = \sum_{k=1}^{N} E\{I_i I_k\} - \pi_i = (n-1)\pi_i. \quad (1.2.16)$$

Discussions of estimation for finite population sampling begin most easily with estimation of linear functions such as finite population totals. This is because it is possible to construct estimators of totals for a wide range of designs that are unbiased conditionally on the particular finite population. Such estimators are said to be *design unbiased.*

Definition 1.2.2. A statistic $\hat{\theta}$ is *design unbiased* for the finite population parameter $\theta_N = \theta(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)$ if

$$E\{\hat{\theta} \mid \mathcal{F}\} = \theta_N$$

for any vector $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)$, where $E\{\hat{\theta} \mid \mathcal{F}\}$, the design expectation, denotes the average over all samples possible under the design for the finite population $\mathcal{F}$.

Probability sampling became widely accepted in the 1940s. For a number of years thereafter, sampling statisticians who considered estimation problems approached design and estimation problems by treating the N unknown values of the finite population as fixed values. All probability statements were with respect to the distribution created by the sample design probabilities. Thus, in many discussions in the sampling literature the statement that an estimator is "unbiased" means design unbiased for a parameter of the finite universe.

The concept of a linear estimator is also very important in estimation theory. Often, modifiers are required to fully define the construct. If an estimator $\hat{\theta}$ can be written as

$$\hat{\theta} = \sum_{i \in A} w_i y_i, \tag{1.2.17}$$

where the w_i are not functions of the sample y's, we say that the estimator $\hat{\theta}$ is *linear in* y. In the statistical theory of linear models, estimators of the form (1.2.17) are called *linear estimators* provided that the w_i are fixed with respect to the random mechanism generating the y values. Thus, the model specification for the random process and the set of samples under consideration define the statistical linearity property. We will have use for the concept of linearity relative to the design.

Definition 1.2.3. An estimator is *design linear* if it can be written in the form (1.2.17) or, equivalently, as

$$\hat{\theta} = \sum_{i \in U} I_i w_i y_i,$$

where the w_i are fixed with respect to the sampling design.

Observe that for a given finite population, the vector $(w_1 y_1, w_2 y_2, \ldots, w_N y_N)$ is a fixed vector and the elements of the vector are the coefficients of the random variables I_i.

The design mean and design variance of design linear estimators are functions of the selection probabilities. In Definition 1.2.2 we introduced the concept of the expectation over all possible samples for a particular finite population $\mathcal{F}$. We use $V\{\hat{\theta} \mid \mathcal{F}\}$ to denote the analogous design variance.

Theorem 1.2.1. Let $(y_1, y_2, \ldots, y_N)$ be the vector of values for a finite universe of real-valued elements. Let a probability sampling procedure be defined, where π_i denotes the probability that element i is included in the sample and π_{ik} denotes the probability that elements i and k are in the sample. Let

$$\hat{\theta} = \sum_{i \in A} w_i y_i = \sum_{i \in U} I_i w_i y_i$$

be a design linear estimator. Then

$$E\{\hat{\theta} \mid \mathcal{F}\} = \sum_{i=1}^{N} w_i \pi_i y_i \tag{1.2.18}$$

and

$$V\{\hat{\theta} \mid \mathcal{F}\} = \sum_{i=1}^{N}\sum_{k=1}^{N}(\pi_{ik} - \pi_i\pi_k)w_iy_iw_ky_k, \tag{1.2.19}$$

where $\pi_{ik} = \pi_i$ if $i = k$.

If $V(n) = 0$, then $V\{\hat{\theta} \mid \mathcal{F}\}$ can be expressed as

$$V\{\hat{\theta} \mid \mathcal{F}\} = \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\pi_i\pi_k - \pi_{ik})(w_iy_i - w_ky_k)^2. \tag{1.2.20}$$

Proof. Because $E\{I_i\} = \pi_i$ and because w_iy_i, $i = 1, 2, \ldots, N$, are fixed, we have

$$E\{\hat{\theta} \mid \mathcal{F}\} = \sum_{i=1}^{N} E\{I_i \mid \mathcal{F}\}w_iy_i = \sum_{i=1}^{N}\pi_iw_iy_i$$

and (1.2.18) is proven. In a similar manner, and using $E\{I_iI_k\} = \pi_{ik}$, we have

$$\begin{aligned} V\{\hat{\theta} \mid \mathcal{F}\} &= V\left\{\sum_{i=1}^{N} I_iw_iy_j \mid \mathcal{F}\right\} \\ &= E\left\{\left(\sum_{i=1}^{N} I_iw_iy_i\right)^2 \middle| \mathcal{F}\right\} - \left(\sum_{i=1}^{N}\pi_iw_iy_i\right)^2 \\ &= \sum_{i=1}^{N}\sum_{k=1}^{N}(\pi_{ik} - \pi_i\pi_k)w_iy_iw_ky_k \end{aligned}$$

and (1.2.19) is proven. Also see Exercise 1.

To prove (1.2.20) for fixed-size designs, expand the square in (1.2.20) to obtain

$$\begin{aligned} &\frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\pi_i\pi_k - \pi_{ik})(w_i^2y_i^2 - 2w_iy_iw_ky_k + w_k^2y_k^2) \\ &\qquad = \sum_{i=1}^{N}\sum_{k=1}^{N}(\pi_i\pi_k - \pi_{ik})w_i^2y_i^2 \\ &\qquad\quad - \sum_{i=1}^{N}\sum_{k=1}^{N}(\pi_i\pi_k - \pi_{ik})w_iy_iw_ky_k. \end{aligned}$$

The result follows because $\Sigma_{k=1}^{N}(\pi_{ik} - \pi_i\pi_k) = 0$ for fixed-size designs. See (1.2.14). ■

We have stated Theorem 1.2.1 for scalars, but the results extend immediately to vectors. If $\mathbf{y}_i$ is a column vector and

$$\hat{\boldsymbol{\theta}} = \sum_{i \in A} w_i \mathbf{y}_i,$$

the covariance matrix of $\hat{\boldsymbol{\theta}}$ is

$$V\{\hat{\boldsymbol{\theta}} \mid \mathcal{F}\} = \sum_{i=1}^{N}\sum_{k=1}^{N}(\pi_{ik} - \pi_i\pi_k)w_i\mathbf{y}_i w_k\mathbf{y}_k'.$$

Two finite population parameters of particular interest are the finite population total,

$$T_y = \sum_{i \in U} y_i = \sum_{i=1}^{N} y_i, \tag{1.2.21}$$

and the finite population mean,

$$\bar{y}_N = N^{-1}T_y. \tag{1.2.22}$$

If $\pi_i > 0$ for all i, the design linear estimator of the total,

$$\hat{T}_y = \sum_{i \in A} \pi_i^{-1} y_i, \tag{1.2.23}$$

is design unbiased. The estimator (1.2.23) is known as the *Horvitz–Thompson estimator* and is sometimes called the *π estimator*. See Horvitz and Thompson (1952) and Narain (1951). The corresponding design-unbiased estimator of the mean is

$$\bar{y}_{HT} = N^{-1}\hat{T}_y \tag{1.2.24}$$

The properties of the Horvitz–Thompson estimator follow from Theorem 1.2.1.

Corollary 1.2.1.1. Let the conditions of Theorem 1.2.1 hold, let $\pi_i > 0$ for all i, and let the design linear estimator of T_y be $\hat{T}_y$ of (1.2.23). Then

$$E\{\hat{T}_y \mid \mathcal{F}\} = T_y \tag{1.2.25}$$

and

$$V\{\hat{T}_y - T_y \mid \mathcal{F}\} = \sum_{i=1}^{N}\sum_{k=1}^{N}(\pi_{ik} - \pi_i\pi_k)\pi_i^{-1}y_i\pi_k^{-1}y_k. \quad (1.2.26)$$

If $V\{n\} = 0$, then $V\{(\hat{T}_y - T_y) \mid \mathcal{F}\}$ can be expressed as

$$\sum_{i=1}^{N}\sum_{k=1}^{N}(\pi_{ik} - \pi_i\pi_k)(\pi_i^{-1}y_i - n^{-1}T_y)(\pi_k^{-1}y_k - n^{-1}T_y) \quad (1.2.27)$$

or as

$$\frac{1}{2}\sum_{\substack{i=1\\ i\neq k}}^{N}\sum_{k=1}^{N}(\pi_i\pi_k - \pi_{ik})(\pi_i^{-1}y_i - \pi_k^{-1}y_k)^2. \quad (1.2.28)$$

Proof. Results (1.2.25), (1.2.26), and (1.2.28) follow from (1.2.18), (1.2.19), and (1.2.20), respectively, by substituting $w_i = \pi_i^{-1}$.

To show that (1.2.27) is equal to (1.2.26) for fixed-size designs, observe that

$$\sum_{i=1}^{N}\sum_{k=1}^{N}\pi_i\pi_k(\pi_i^{-1}y_i - n^{-1}T_y)(\pi_k^{-1}y_k - n^{-1}T_y)$$

$$= \left(\sum_{i=1}^{N}(y_i - n^{-1}\pi_i T_y)\right)^2 = 0.$$

From (1.2.16), $\Sigma_{k=1}^{N}\pi_{ik} = n\pi_i$. Thus,

$$\sum_{i=1}^{N}(\pi_i^{-1}y_i - n^{-1}T_y)\sum_{k=1}^{N}\pi_{ik}n^{-1}T_y = T_y\sum_{i=1}^{N}(y_i - n^{-1}\pi_i T_y) = 0$$

and (1.2.27) is equal to

$$\sum_{i=1}^{N}\sum_{k=1}^{N}\pi_{ik}\pi_i^{-1}y_i\pi_k^{-1}y_k - T_y^2.$$

■

The Horvitz–Thompson estimator is an unbiased estimator of the total, but it has some undesirable features. The estimator is scale invariant but not

location invariant. That is, for real α, β not zero,

$$\sum_{i\in A} \pi_i^{-1}\beta y_i \;=\; \beta \sum_{i\in A} \pi_i^{-1} y_i,$$

but

$$\sum_{i\in A} \pi_i^{-1}(y_i+\alpha) \;=\; \sum_{i\in A} \pi_i^{-1} y_i + \alpha \sum_{i\in A} \pi_i^{-1}. \tag{1.2.29}$$

The second term of (1.2.29) is $N\alpha$ for many designs, including equal-probability fixed-sample-size designs. However, $\sum_{i\in A} \pi_i^{-1}$ is, in general, a nondegenerate random variable.

The lack of location invariance restricts the number of practical situations in which the Horvitz–Thompson estimator and unequal probability designs are used. One important use of unequal probability sampling is the situation in which the π_i are proportional to a measure of the number of observation units associated with the sampling unit.

Example 1.2.1. Assume that one is interested in the characteristics of households in Des Moines, Iowa. A recent listing of the city blocks and the number of dwelling units in each block is available. On the presumption that the number of households is strongly correlated with the number of dwelling units, we might select a sample of blocks with probability proportional to the number of dwelling units. Assume that all households in the block are observed. In this situation, the fact that the Horvitz–Thompson estimator is not location invariant is relatively unimportant because we are interested in the properties of households, not in the properties of linear functions of blocks. It was in a context such as this that unequal probability sampling was first suggested. See Hansen and Hurwitz (1943). ■■

The fact that the Horvitz–Thompson estimator is not location invariant has another consequence. Associated with each sampling unit is the characteristic, which is always 1. The population total for this characteristic is the number of sampling units in the population. The Horvitz–Thompson estimator of the population size is the coefficient of α in (1.2.29),

$$\hat{N} \;=\; \hat{T}_1 \;=\; \sum_{i\in A} \pi_i^{-1} \tag{1.2.30}$$

with variance

$$V\{\hat{T}_1 \mid \mathcal{F}\} \;=\; \sum_{i=1}^{N}\sum_{k=1}^{N} (\pi_{ik} - \pi_i\pi_k)\pi_i^{-1}\pi_k^{-1}. \tag{1.2.31}$$

Although there are situations in which N is unknown, in many situations N is known. Therefore, the fact that the estimator of the population size is not equal to the true size suggests the possibility of improving the Horvitz–Thompson estimator. We pursue this issue in Section 1.3 and Chapter 2.

Under the conditions that $\pi_i > 0$ for all i and $\pi_{ik} > 0$ for all i and k, it is possible to construct a design-unbiased estimator of the variance of a design linear estimator. Designs with the properties $\pi_i > 0$ for all i and $\pi_{ik} > 0$ for all ik are sometimes said to be *measurable*.

Theorem 1.2.2. Let the conditions of Theorem 1.2.1 hold with $\pi_{ik} > 0$ for all $i,\ k,\ \in U$. Let $\hat{\theta}$ be a design linear estimator of the form (1.2.17). Then

$$\hat{V}\{\hat{\theta} \mid \mathcal{F}\} = \sum_{i,k\in A}\sum \pi_{ik}^{-1}\left(\pi_{ik} - \pi_i\pi_k\right) w_i y_i w_k y_k \qquad (1.2.32)$$

is a design-unbiased estimator of $V\{\hat{\theta} \mid \mathcal{F}\}$. If $V\{n\} = 0$,

$$\tilde{V}\{\hat{\theta} \mid \mathcal{F}\} = \frac{1}{2}\sum_{i,k\in A}\sum \pi_{ik}^{-1}(\pi_i\pi_k - \pi_{ik})(w_i y_i - w_k y_k)^2 \qquad (1.2.33)$$

is a design-unbiased estimator of $V\{\hat{\theta} \mid \mathcal{F}\}$.

Proof. Let $g(y_i,\ y_k)$ be any real-valued function of $(y_i,\ y_k)$. Because $\pi_{ik} > 0$ for all $(i,\ k)$, it follows by direct analogy to (1.2.18) that

$$E\left\{\sum_{i,k\in A} \pi_{ik}^{-1} g(y_i,\ y_k) \mid \mathcal{F}\right\} = \sum_{i=1}^{N}\sum_{k=1}^{N} g(y_i,\ y_k). \qquad (1.2.34)$$

Result (1.2.32) is obtained from (1.2.34) and (1.2.19) by setting

$$g(y_i,\ y_k) = (\pi_{ik} - \pi_i\pi_k) w_i y_i w_k y_k.$$

Result (1.2.33) follows from (1.2.34) and (1.2.20) by setting

$$g(y_i,\ y_k) = (\pi_{ik} - \pi_i\pi_k)\,(w_i y_i - w_k y_k)^2.$$

■

The estimator (1.2.32) for estimator (1.2.19) is due to Horvitz and Thompson (1952), and the estimator (1.2.33) was suggested by Yates and Grundy (1953) and Sen (1953) for estimator (1.2.20).

Theoretically, it is possible to obtain the variance of the estimated variance. The squared differences in (1.2.33) are a sample of all possible differences.

If we consider differences $(w_i y_i - w_k y_k)^2$, for $i \neq k$ there is a population of $N(N-1)$ differences. The probability of selecting any particular difference is π_{ik}. The variance of the estimated difference is a function of the π_{ik} and of the probability that any pair of pairs occurs in the sample. Clearly, this computation can be cumbersome for general designs. See Exercise 12.

Although design unbiased, the estimators of variance in Theorem 1.2.2 have the unpleasant property that they can be negative. If at least two values of $\pi_i^{-1} y_i$ differ in the sample, the variance must be positive and any other value for an estimator is unreasonable.

The Horvitz–Thompson variance estimator also has the undesirable property that it can give a positive estimate for an estimator known to have zero variance. For example, if y_i is proportional to π_i, the variance of $\hat{T}_y$ is zero for fixed-size designs, but estimator (1.2.32) can be nonzero for some designs.

Theorem 1.2.2 makes it clear that there are some designs for which design-unbiased variance estimation is impossible because unbiased variance estimation requires that $\pi_{ik} > 0$ for all (i, k). A sufficient condition for a design to yield nonnegative estimators of variance is $\pi_{ik} < \pi_i \pi_k$. See (1.2.33).

For simple random nonreplacement sampling, $\pi_i = N^{-1} n$ and

$$\pi_{ik} = [N(N-1)]^{-1} n(n-1) \text{ for } i \neq k.$$

Then the estimated total (1.2.23) is

$$\hat{T}_y = N n^{-1} \sum_{i \in A} y_i = N \bar{y}_n, \tag{1.2.35}$$

where

$$\bar{y}_n = n^{-1} \sum_{i \in A} y_i.$$

Similarly, the variance (1.2.26) reduces to

$$\begin{aligned} V\{(\hat{T}_y - T_y) \mid \mathcal{F}\} &= N(N-n) n^{-1} S_{y,N}^2, \\ &= N^2 (1 - f_N) n^{-1} S_{y,N}^2, \end{aligned} \tag{1.2.36}$$

where

$$S_{y,N}^2 = (N-1)^{-1} \sum_{j=1}^{N} (y_j - \bar{y}_N)^2$$

and $f_N = N^{-1} n$. The quantity $S_{y,N}^2$, also written without the N subscript and with different subscripts, is called the *finite population variance*. A few texts define the finite population variance with a divisor of N and change the definition of $V\{\hat{T}_y \mid \mathcal{F}\}$ appropriately. The term $(1 - f_N)$ is called the *finite*

population correction (fpc) or *finite correction term*. It is common practice to ignore the term if the sampling rate is less than 5%. See Cochran (1977, p. 24).

The estimated variance (1.2.33) reduces to

$$\hat{V}\{\hat{T}_y \mid \mathcal{F}\} = Nn^{-1}(N-n)s_{y,n}^2 \tag{1.2.37}$$

for simple random sampling, where

$$s_{y,n}^2 = (n-1)^{-1}\sum_{j\in A}(y_j - \bar{y}_n)^2.$$

The quantity $s_{y,n}^2$ is sometimes called the *sample variance* and may be written with different subscripts. The results for simple random sampling are summarized in Corollary 1.2.2.1.

Corollary 1.2.2.1. Let $U = \{1, 2, \ldots, N\}$ and let $\mathcal{F} = (y_1, y_2, \ldots, y_N)$ be the values of a finite universe. Let a simple random sample of size n be selected from $\mathcal{F}$, let $\bar{y}_n$ be defined by (1.2.35), and let $s_{y,n}^2$ be as defined for (1.2.37). Then

$$\begin{aligned} E\{\bar{y}_n \mid \mathcal{F}\} &= \bar{y}_N, \\ V\{\bar{y}_n \mid \mathcal{F}\} &= N^{-1}(N-n)n^{-1}S_{y,N}^2, \end{aligned} \tag{1.2.38}$$

and

$$E\{\hat{V}(\bar{y}_n \mid \mathcal{F}) \mid \mathcal{F}\} = V\{\bar{y}_n \mid \mathcal{F}\}, \tag{1.2.39}$$

where $\bar{y}_N$ is defined in (1.2.22), $S_{y,N}^2$ is defined in (1.2.36), and

$$\hat{V}\{\bar{y}_n \mid \mathcal{F}\} = (1-f_N)n^{-1}s_{y,n}^2.$$

Proof. For simple random nonreplacement sampling, $\pi_i = N^{-1}n$ and $\pi_{ik} = [N(N-1)]^{-1}n(n-1)$ for $i \neq k$. Thus, by (1.2.25) of Corollary 1.2.1.1,

$$E\{\hat{T}_y \mid \mathcal{F}\} = E\{N\bar{y}_n\} = N\bar{y}_N$$

and we have the first result. The result (1.2.38) is obtained by inserting the probabilities into (1.2.26) to obtain

$$\begin{aligned} V\{\hat{T}_y - T_y \mid \mathcal{F}\} &= \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{N} n^{-1}(N-n)(N-1)^{-1}(y_i - y_k)^2 \\ &= Nn^{-1}(N-n)(N-1)^{-1}\sum_{i=1}^{N}(y_i - \bar{y}_N)^2 \\ &= N^2n^{-1}(1-f_N)S_{y,N}^2. \end{aligned}$$

By the same algebraic argument, the estimator (1.2.33) for the estimated total is

$$\hat{V}\{\hat{T}_y \mid \mathcal{F}\} = Nn^{-1}(N-n)s_{y,n}^2$$

with expectation $Nn^{-1}(N-n)S_{y,N}^2$ by Theorem 1.2.2. ■

In some situations, such as the investigation of alternative designs, it is useful to consider the finite population to be generated by a stochastic mechanism. For example, the $\{y_i\}$, $i = 1, 2, \ldots, N$, might be independent identically distributed (iid) random variables with a distribution function $F(y)$. We then say that the finite population is a sample from the *superpopulation* $F(y)$.

A simple and useful specification is that of Theorem 1.2.3. The combination of a sample design and an estimator is called a *strategy*.

Theorem 1.2.3. Let $\{y_1, y_2, \ldots, y_N\}$ be a set of $iid(\mu, \sigma^2)$ random variables. Let the sample design have probabilities π_i, $\pi_i > 0$, and π_{ik} such that $\Sigma_{i=1}^N \pi_i = n$. Assume that the vector $\mathbf{d}$ of selection indicators is independent of $\{y_1, y_2, \ldots, y_N\}$. Let the estimated total be

$$\hat{T}_y = \sum_{i \in A} \pi_i^{-1} y_i. \tag{1.2.40}$$

Then $\pi_i = N^{-1}n$ and $\pi_{ik} = [N(N-1)]^{-1}n(n-1)$ for $i \neq k$ minimize the variance of $\hat{T}_y - T_y$.

Proof. Under the assumptions, $E\{\hat{T}_y - T_y \mid \mathcal{F}\} = 0$ and d is independent of the y_i. Therefore, the unconditional variance of $\hat{T}_y - T_y$ is the expectation of the conditional variance. Using $E\{y_i^2\} = \sigma^2 + \mu^2$ and $E\{y_i y_k\} = \mu^2$ for $i \neq k$,

$$\begin{aligned} V\{\hat{T}_y - T_y\} &= E\left\{\sum_{i=1}^N \sum_{k=1}^N C\{I_i, I_k\} \pi_i^{-1} y_i \pi_k^{-1} y_k\right\} \\ &= \mu^2 \sum_{i=1}^N \sum_{k=1}^N (\pi_{ik} - \pi_i \pi_k)\pi_i^{-1}\pi_k^{-1} \\ &\quad + \sigma^2 \sum_{i=1}^N (\pi_i - \pi_i^2)\pi_i^{-2}, \end{aligned} \tag{1.2.41}$$

where $C\{x, z\}$ is the covariance of x and z.

Given that $\Sigma_{i=1}^N \pi_i = n$, the second term of (1.2.41) is minimized if all π_i are equal because π_i^{-1} is convex in π_i. The first term of (1.2.41) is nonnegative

because it equals $\mu^2 V\{\Sigma_{i=1}^{N} I_i \pi_i^{-1}\}$. Therefore, the minimum possible value for the first term is zero, which is attained if the π_i are equal to $N^{-1}n$ and $\pi_{ik} = [N(N-1)]^{-1}n(n-1)$ for all $i \neq k$. ■

The formulation of Theorem 1.2.3 deserves further discussion. The result pertains to a property of an estimator of the finite population total. However, the property is an average over all possible finite populations. The result does not say that simple random sampling is the best procedure for the Horvitz–Thompson estimator for a particular finite population. We cannot find a design-unbiased procedure that is minimum variance for all fixed unknown finite populations because the design variance is a function of the N unknown values. See Godambe (1955), Godambe and Joshi (1965), and Basu (1971). On the other hand, if our information about the finite population is such that we are willing to act as if the finite population is a set of *iid* random variables, simple random sampling is the best sampling strategy for the Horvitz–Thompson estimator, where "best" is with respect to the superpopulation specification. If the finite population is assumed to be a sample from a normal distribution, the sample mean is optimal for the finite population mean.

Theorem 1.2.4. Let $\{y_1, y_2, \ldots, y_N\}$ be a set of normal independent random variables with mean μ and variance σ^2, denoted by $NI(\mu, \sigma^2)$ random variables. In the class of sample selection procedures for samples of size n that are independent of $\{y_1, y_2, \ldots, y_N\}$, the procedure of selecting a simple random nonreplacement sample of size n from U and using the estimator $\bar{y}_n$ to estimate the finite population mean, $\bar{y}_N$, is an optimal strategy in that there is no strategy with smaller mean square error.

Proof. By Theorem 1.3.1, the n elements in the sample are $NI(\mu, \sigma^2)$ random variables. Therefore, the sample mean is the minimum mean square error estimator of μ. See, for example, Stuart and Ord (1991, p. 617). Furthermore, the minimum mean square error predictor of $\bar{y}_{N-n}$ is $\bar{y}_n$. Thus,

$$\bar{y}_n = N^{-1}[n\bar{y}_n + (N-n)\bar{y}_n]$$

is the best predictor of $\bar{y}_N$. ■

Because the elements of the original population are identically distributed, any nonreplacement sampling scheme that is independent of $\{y_1, y_2, \ldots, y_N\}$ would lead to the same estimation scheme and the same mean square error. However, probability sampling and the Horvitz–Thompson estimator are robust in the sense that the procedure is unbiased for any set $\{y_1, y_2, \ldots, y_N\}$.

If the normality assumption is relaxed, the mean is optimal in the class of linear unbiased estimators (predictors).

Corollary 1.2.4.1. Let $\{y_1, y_2, \ldots, y_N\}$ be a set of *iid* random variables with mean μ and variance σ^2. Then the procedure of selecting a simple random nonreplacement sample of size n from N and using the estimator $\bar{y}_n$ is a minimum mean square error procedure for $\bar{y}_N$ in the design-estimator class composed of designs that are independent of $\{y_1, y_2, \ldots, y_N\}$ combined with linear estimators.

Proof. Under the *iid* assumption, the sample mean is the minimum mean square error linear predictor of $\bar{y}_{N-n}$ and the result follows. See Goldberger (1962) and Graybill (1976, Section 12.2) for discussions of best linear unbiased prediction.

■

The consideration of unequal probabilities of selection opens a wide range of options and theoretical difficulties. The very fact that one is able to associate unequal π_i with the elements means that we know something that differentiates element i from element j. It is no longer reasonable to treat the elements as exchangeable, that is, as a set for which the joint distribution does not depend on the indexing. However, it may be possible to transform the observations to achieve exchangeability. For example, it might be possible to define π_i such that it is reasonable to treat the $y_i\pi_i^{-1}$ as exchangeable. The nature of auxiliary information and the manner in which it should enter selection and estimation is the subject of survey sampling.

1.2.2 Poisson sampling

A sample design with simple theoretical properties is that in which samples are created by conducting N independent Bernoulli trials, one for each element in the population. If the result of the trial is a success, the element is included in the sample. Otherwise, the element is not part of the sample. The procedure is called *Poisson sampling* or *Bernoulli sampling* or *sieve sampling*.

Theorem 1.2.5. Let $(y_1, y_2, \ldots, y_N)$ be a finite universe of real-valued elements, and let $(\pi_1, \pi_2, \ldots, \pi_N)$ be a corresponding set of probabilities with $\pi_i > 0$ for all $i \in U$. For Poisson samples,

$$V\{\hat{T}_y - T_y \mid \mathcal{F}\} = \sum_{i=1}^{N} \pi_i^{-1}(1-\pi_i)y_i^2, \tag{1.2.42}$$

where T_y is the total defined in (1.2.21) and $\hat{T}_y$ is the Horvitz–Thompson estimator (1.2.23).

The expected sample size is

$$E\{n\} = \sum_{i=1}^{N} \pi_i \tag{1.2.43}$$

and

$$V\{n\} = \sum_{i=1}^{N} \pi_i(1-\pi_i). \tag{1.2.44}$$

A design-unbiased estimator for the variance of $\hat{T}_y$ is

$$\hat{V}\{\hat{T}_y \mid \mathcal{F}\} = \sum_{i\in A}(1-\pi_i)\pi_i^{-2}y_i^2. \tag{1.2.45}$$

If $\pi_i \equiv \pi$,

$$E\left\{n^{-1}\sum_{i\in A} y_i \mid (\mathcal{F},\ n),\ n>0\right\} = \bar{y}_N \tag{1.2.46}$$

and

$$V\left\{n^{-1}\sum_{i\in A} y_i \mid (\mathcal{F},\ n),\ n>0\right\} = N^{-1}(N-n)n^{-1}S_{y,N}^2. \tag{1.2.47}$$

Proof. Results (1.2.42), (1.2.43), and (1.2.44) follow from the fact that the I_i of $\mathbf{d} = (I_1, I_2, \ldots, I_N)$ are independent Bernoulli random variables. If $\pi_i \equiv \pi$, the sample size is a binomial random variable because it is the sum of N *iid* Bernoulli random variables. The set of samples with size $n = n_0$ is the set of simple random nonreplacement samples of size n_0, because every sample of size n_0 has the same probability of selection. Results (1.2.46) and (1.2.47) then follow. ∎

Theorem 1.2.5 gives another example of difficulties associated with unconsidered use of the Horvitz–Thompson estimator. If $\pi_i \equiv \pi$, the Horvitz–Thompson estimator of the total of y for a Poisson sample is

$$\hat{T}_y = \pi^{-1}\sum_{i\in A} y_i \tag{1.2.48}$$

with variance

$$V\{\hat{T}_y \mid \mathcal{F}\} = \pi^{-2}\sum_{i=1}^{N}\pi(1-\pi)y_i^2. \tag{1.2.49}$$

By (1.2.46), another estimator of the total of a Poisson sample with $\pi_i \equiv \pi$ is

$$\begin{aligned}\tilde{T}_y &= N\bar{y}_n \quad \text{if } n > 0 \\ &= 0 \quad \text{if } n = 0,\end{aligned} \tag{1.2.50}$$

where $\bar{y}_n = n^{-1}\Sigma_{i\in A}y_i$. The estimator $N\bar{y}_n$ is conditionally unbiased for T_y for each positive n, and if $n = 0$, $\tilde{T}_y = \hat{T}_y = 0$. The mean square error of $\tilde{T}_y$ is

$$N^2 E\left\{(n^{-1} - N^{-1})S_{y,N}^2 \mid (\mathcal{F}, n),\ n > 0\right\} P\{n > 0\} + T_y^2 P\{n = 0\}. \tag{1.2.51}$$

Now $E\{n\} = \mu_n = N\pi$ and the variance of the Horvitz–Thompson estimator can be written

$$N^2(\mu_n^{-1} - N^{-1})\left[N^{-1}(N-1)S_{y,N}^2 + \bar{y}_N^2\right].$$

While $E\{n^{-1} \mid n > 0\} > \mu_n^{-1}$, it is difficult to think of a situation in which one would choose the Horvitz–Thompson estimator over estimator (1.2.50). Note also that given $\pi_i \equiv \pi$,

$$\hat{V}\{\tilde{T}_y \mid (\mathcal{F},\ n),\ n > 1\} = N^2(n^{-1} - N^{-1})s_{y,n}^2, \tag{1.2.52}$$

where $s_{y,n}^2$ is as defined in (1.2.37), is a conditionally unbiased estimator of the conditional variance of $\tilde{T}_y$, conditional on $n > 1$.

1.2.3 Stratified sampling

Assume that the elements of a finite population are divided into H groups, indexed by $h = 1, 2, \ldots, H$, called *strata*. Assume that the hth stratum contains N_h elements and it is desired to estimate the finite population mean

$$\bar{y}_N = N^{-1}\sum_{h=1}^{H}\sum_{i=1}^{N_h} y_{hi} = \sum_{h=1}^{H} N^{-1}N_h\bar{y}_{Nh}, \tag{1.2.53}$$

where $\bar{y}_{Nh} = N_h^{-1}\Sigma_{i=1}^{N_h}y_{hi}$ and y_{hi} is the ith element in the hth stratum.

Assume that we are willing to treat the elements in each stratum as if they were a random sample from a population with mean μ_h and variance σ_h^2. That is, the unknown values in stratum h are considered to be a realization of N_h

iid random variables. Thus, if one were selecting a sample to estimate the mean of an individual stratum, it is reasonable, by Theorems 1.2.3 and 1.2.4, to select a simple random sample from that stratum. Then, the sample mean of the stratum sample is an unbiased estimator of the population stratum mean. That is,

$$E\{\bar{y}_h \mid \mathcal{F}\} = \bar{y}_{Nh},$$

where

$$\bar{y}_h = n_h^{-1} \sum_{i \in A_h} y_{hi}$$

and A_h is the set of indices for the sample in stratum h. It follows that

$$\bar{y}_{st} = \sum_{h=1}^{H} N^{-1} N_h \bar{y}_h \tag{1.2.54}$$

is unbiased for the population mean, where $\bar{y}_h$ is the mean of a simple nonreplacement sample of size n_h selected from stratum h and the H stratum samples are mutually independent.

The procedure of selecting independent samples from H mutually exclusive and exhaustive subdivisions of the population is called *stratified sampling*, a very common technique in survey sampling. The samples within a stratum need not be simple random samples, but we concentrate on that case.

Because the $\bar{y}_h$ are independent,

$$V\{\bar{y}_{st} - \bar{y}_N \mid \mathcal{F}\} = \sum_{h=1}^{H} (N^{-1} N_h)^2 N_h^{-1} (N_h - n_h) n_h^{-1} S_h^2, \tag{1.2.55}$$

where $\bar{y}_{st}$ is as defined in (1.2.54) and

$$S_h^2 = (N_h - 1)^{-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{y}_{Nh})^2.$$

The estimated variance of the stratum mean $\bar{y}_h$ is

$$\hat{V}\{(\bar{y}_h - \bar{y}_{Nh}) \mid \mathcal{F}\} = N_h^{-1} (N_h - n_h) n_h^{-1} s_h^2,$$

where

$$s_h^2 = (n_h - 1)^{-1} \sum_{i \in A_h} (y_{hi} - \bar{y}_h)^2.$$

It follows that an unbiased estimator of the variance of $\bar{y}_{st}$ is

$$\hat{V}\{(\bar{y}_{st} - \bar{y}_N) \mid \mathcal{F}\} = \sum_{h=1}^{H} (N^{-1}N_h)^2 N_h^{-1}(N_h - n_h)n_h^{-1}s_h^2. \quad (1.2.56)$$

Under the model in which the y_{hi} are realizations of $iid(\mu_h, \sigma_h^2)$ random variables, the unconditional variance is the expected value of (1.2.55),

$$V\{\bar{y}_{st} - \bar{y}_N\} = E\{V[\bar{y}_{st} - \bar{y}_N \mid \mathcal{F}]\} = \sum_{h=1}^{H} N^{-2}N_h(N_h - n_h)n_h^{-1}\sigma_h^2. \quad (1.2.57)$$

Assume that the objective of the design and estimation operation is to estimate the population mean, $\bar{y}_N$, of the characteristic y. Assume that a total amount C is available for sample observation and that it costs c_h to observe an element in stratum h. Under this scenario, one would choose the n_h to minimize the variance (1.2.55), or the variance (1.2.57), subject to the condition that

$$\sum_{h=1}^{H} n_h c_h \leq C.$$

The minimization requires knowledge of the S_h^2 or of σ_h^2.

In practice, one seldom knows the σ_h^2 at the time that one is constructing a sampling design. Thus, it is reasonable for the designer to construct a model for the population, to postulate parameters for the model, and to use the model and parameters as the basis for determining a design. The model is called the *design model*, and the parameters of the model are called *design parameters* or *anticipated parameters*.

The expected value of the design variance of the planned estimator calculated under the designer's model using the postulated parameters is called the *anticipated variance*. Let $\hat{\theta}$ be an estimator of a finite population parameter θ_N. Then the anticipated variance of $\hat{\theta} - \theta_N$ is

$$AV\{\hat{\theta} - \theta_N\} = E\{E[(\hat{\theta} - \theta_N)^2 \mid \mathcal{F}]\} - [E\{E(\hat{\theta} - \theta_N \mid \mathcal{F})\}]^2.$$

For stratified sampling $E\{\bar{y}_{st} - \bar{y}_N \mid \mathcal{F}\} = 0$ and the anticipated variance of $\bar{y}_{st}$ is minimized by minimizing

$$AV\{\bar{y}_{st} - \bar{y}_N\} = \sum_{h=1}^{H} N^{-2}N_h^2(n_h^{-1} - N_h^{-1})\ddot{\sigma}_h^2, \quad (1.2.58)$$

subject to the cost restriction, where $\ddot{\sigma}_h^2$, $h = 1, 2, \ldots, H$, are the anticipated stratum variances. If one uses the method of Lagrange multipliers, one

obtains

$$n_h = \lambda^{-1/2} c_h^{-1/2} N^{-1} N_h \ddot{\sigma}_h, \tag{1.2.59}$$

where λ is the Lagrange multiplier and

$$\lambda^{1/2} = C^{-1} \sum_{h=1}^{H} N^{-1} N_h c_h^{1/2} \ddot{\sigma}_h.$$

In general, the n_h of (1.2.59) are not integers. Also, it is possible for n_h to exceed N_h. The formal feasible solution could be obtained by using integer programming. In practice, the n_h are rounded to integers with all n_h greater than or equal to 2 and less than or equal to N_h. The allocation with n_h proportional to $N_h \ddot{\sigma}_h$ is optimal for constant costs and is sometimes called *Neyman allocation*, after Neyman (1934).

Our discussion is summarized in the following theorem.

Theorem 1.2.6. Let $\mathcal{F}$ be a stratified finite population in which the elements in stratum h are realizations of $iid(\mu_h, \sigma_h^2)$ random variables. Let $\ddot{\sigma}_h^2$, $h = 1, 2, \ldots, H$, be the anticipated variances, let C be the total amount available for sample observation, and assume that it costs c_h to observe an element in stratum h. Then a sampling and estimation strategy for $\bar{y}_N$ that minimizes the anticipated variance in the class of linear unbiased estimators and probability designs is: Select independent simple random nonreplacement samples in each stratum, selecting n_h in stratum h, where n_h is defined in (1.2.59), subject to the integer and population size constraints, and use the estimator defined in (1.2.54).

If it is desired to obtain a particular variance for a minimum cost, one minimizes cost subject to the variance constraint

$$V_S = \sum_{h=1}^{H} N^{-2} N_h^2 (n_h^{-1} - N_h^{-1}) \ddot{\sigma}_h^2,$$

where V_S is the variance specified. In this case,

$$n_h = \left(V_S + \sum_{h=1}^{H} N^{-2} N_h \ddot{\sigma}_h^2 \right)^{-1} \left(\sum_{h=1}^{H} N^{-1} N_h c_h^{1/2} \ddot{\sigma}_h \right) N^{-1} N_h c_h^{-1/2} \ddot{\sigma}_h.$$

In both cases, n_h is proportional to $N^{-1} N_h c_h^{-1/2} \ddot{\sigma}_h$.

1.2.4 Systematic sampling

Systematic sampling is used widely because of the simplicity of the selection procedure. Assume that it is desired to select a sample of size n from a population of size N with probabilities π_i, $i = 1, 2, \ldots, N$, where $0 < \pi_i < 1$. To introduce the procedure, consider the population of 11 elements displayed in Table 1.1. Assume that it is desired to select a sample of four elements with probabilities of selection that are proportional to the measures of size. The sum of the sizes is 39. Thus, π_i is the size of the ith element divided by 39 and multiplied by 4. The third column contains the cumulated sizes, and the fourth column contains the cumulated sizes, normalized so that the sum is 4.

To select a systematic random sample of four elements, we select a random number in the interval (0, 1). For our example, assume that the random number is 0.4714. Then the elements in the cumulated normalized sum associated with the numbers 0.4714, 1.4714, 2.4714, and 3.4714 constitute the sample. Let the cumulated size for element t be C_t, where the elements are numbered $1, 2, \ldots, N$. Then an element is associated with the number ζ if

$$C_{t-1} < \zeta \leq C_t.$$

The probability that one of the numbers $(RN, RN+1, RN+2, RN+3)$, where RN is a random number in (0, 1), falls in the interval $(C_{t-1}, C_t]$ is π_t.

Table 1.1 Selection of a Systematic Sample

Element Number	Measure of Size	Cumulated Size	Normalized Cumulated Size	Random Number and Increments
1	6	6	0.6154	0.4714
2	5	11	1.1282	
3	6	17	1.7436	1.4714
4	4	21	2.1538	
5	5	26	2.6667	2.4714
6	4	30	3.0769	
7	2	32	3.2821	
8	3	35	3.5897	3.4714
9	2	37	3.7949	
10	1	38	3.8974	
11	1	39	4.0000	

The selection is particularly simple if $N = nk$, where k is an integer and the elements are to be selected with equal probability. Then the selection consists of selecting a random integer between 1 and k inclusive, say r. The sample is composed of elements $r,\ r+k,\ r+2k,\ \ldots,\ r+(n-1)k$. For this situation, there are k possible samples and we have

$$\begin{aligned} \pi_i &= k^{-1} && \text{for all } i \\ \pi_{ij} &= k^{-1} && \text{if } j = i + km \text{ or } j = i - km \\ &= 0 && \text{otherwise,} \end{aligned} \tag{1.2.60}$$

where m is an integer. Because $\pi_{ij} = 0$ for some pairs, it is not possible to construct a design-unbiased estimator of the variance of a systematic sample.

If the elements are arranged in random order and if the elements are selected with equal probability, systematic sampling produces a simple random nonreplacement sample. Sometimes, for populations in natural order, the variance is estimated as if the sample were a random nonreplacement sample. Such variance calculation is appropriate if the natural order is equivalent to random order. More often, adjacent pairs are assigned to pseudo strata and the variance estimated as if the sample were a two-per-stratum stratified sample. See Section 5.3.

Systematic samples are sometimes defined with random sample sizes. For example, we might draw a sample from a population of size N by selecting a random integer between 1 and k inclusive, say r. Let the sample be elements $r,\ r+k,\ \ldots$, where the last element is $r+(q-1)k$ and $N-k < r+(q-1)k \leq N$. Let

$$N = kq + L,$$

where q and L are integers and $0 \leq L < k$; then L of the samples are of size $q+1$, and $k-L$ of the samples are of size q. Because every element has a probability k^{-1} of being selected, the Horvitz–Thompson estimator is unbiased for the population total. However, the estimator $\bar{y}_n$, where $\bar{y}_n$ is the sample mean, is slightly biased for the population mean.

To consider systematic sampling for the mean of a population arranged in natural order, assume that the superpopulation satisfies the stationary first-order autoregressive model

$$\begin{aligned} y_t &= \rho y_{t-1} + e_t, \\ e_t &\sim NI(0,\ \sigma^2), \end{aligned} \tag{1.2.61}$$

where $0 < \rho < 1$ and the symbol $\sim$ means "is distributed as." Under this model

$$V\{y_t\} = (1-\rho^2)^{-1}\sigma^2,$$

$$C\{y_t,\ y_{t+j}\} \quad = \quad (1-\rho^2)^{-1}\rho^{|j|}\sigma^2,$$

and the correlation between unit t and unit $t+j$, denoted by $\rho(j)$, is $\rho^{|j|}$. A systematic sample of size n selected from a population of size $N = nk$ generated by model (1.2.61) nearly minimizes the variance of the sample mean as an estimator of the finite population mean. Under the extended model with correlations that satisfy

$$\rho(i) - 2\rho(i+1) + \rho(i+2) \geq 0 \qquad \text{for } i = 0,\ 1,\ 2,\ \ldots,$$

Papageorgiou and Karakostas (1998) show that the optimal design for the population mean using the sample mean as the estimator is the systematic sample with the index of the first unit equal to the integer approximation of $(2n)^{-1}(N-n)$. Blight (1973) pointed out that the optimal linear estimator of the population mean under the model (1.2.61) is a weighted combination that gives more weight to the first and last observations than to the middle observations. Such selection and estimation procedures, although the best under the model, are not design unbiased.

Systematic sampling is efficient relative to simple random nonreplacement sampling for populations with a linear trend. Assume that the population satisfies the model

$$y_t \quad = \quad \beta_0 + \beta_1 t + e_t, \tag{1.2.62}$$

where e_t are $iid(0,\ \sigma^2)$ random variables. Then, for a population of size $N = kn$, the variance, under the model, of the random-start systematic sample mean as an estimator of the population mean is

$$\begin{aligned} V\{\bar{y}_{sys} - \bar{y}_N\} \quad &= \quad (12)^{-1}(k+1)(k^2-k)\beta_1^2 \\ &\quad + n^{-1}k^{-1}(k-1)\sigma^2 \\ &\doteq \quad (12)^{-1}k^3\beta_1^2 + n^{-1}\sigma^2 \end{aligned} \tag{1.2.63}$$

for large k. The variance of the sample mean for a simple random nonreplacement sample is approximately

$$V\{\bar{y}_{srs} - \bar{y}_N\} \quad \doteq \quad (12)^{-1}n^2k^3\beta_1^2 + n^{-1}\sigma^2\ . \tag{1.2.64}$$

If the ordered population is divided into n strata of size k and one element is selected in each stratum, the variance of the stratified mean as an estimator of the population mean is

$$\begin{aligned} V\{\bar{y}_{st} - \bar{y}_N\} \quad &= \quad n^{-1}k^{-1}(k-1)(12)^{-1}(k+1)(k^2-k)\beta_1^2 \\ &\quad + n^{-1}k^{-1}(k-1)\sigma^2 \\ &\doteq \quad n^{-1}(12)^{-1}k^3\beta_1^2 + n^{-1}\sigma^2. \end{aligned} \tag{1.2.65}$$

Thus, because the stratified sample mean averages over the local linear trends, it is more efficient than the systematic sample. It is not possible to construct a design-unbiased estimator of the variance for either the one-per-stratum or the systematic designs.

Systematic sampling can also be inefficient relative to simple random nonreplacement sampling. Assume that the y values satisfy

$$y_t = \sin 2\pi k^{-1} t.$$

Then the values in a systematic sample of interval k are identical. Hence, for this population, the variance of the mean is greater than the variance of a simple random nonreplacement sample. Furthermore, because the within-sample variation observed is zero, the estimated variance is zero when the variance is estimated as if the sample were a simple random sample.

See Bellhouse (1988) for a review of systematic sampling. Variance estimation for systematic samples is considered in Section 5.3.

1.2.5 Replacement sampling

Consider a sampling scheme in which repeated selections of a single element are made from a population of elements. Let the selection probabilities for each selection, or draw, for the N elements be p_{ri}, $i = 1, 2, \ldots, N$, where

$$\sum_{i=1}^{N} p_{ri} = 1.$$

Then a replacement sample of size n is that obtained by selecting an element from the N elements with probability p_{ri} at each of n draws. Such a procedure may produce a sample in which element i appears more than once. An estimator of the population total is

$$\hat{T}_{yR} = n^{-1} \sum_{i \in A} p_{ri}^{-1} y_i t_i = \sum_{d=1}^{n} n^{-1} p_d^{-1} y_d, \tag{1.2.66}$$

where t_i is the number of times that element i is selected in the sample, and (p_d, y_d) is the value of (p_{ri}, y_i) for the element selected on the dth draw. Although simple replacement sampling is seldom used in practice, its properties are useful in theoretical discussions.

We may omit the descriptor *nonreplacement* when discussing nonreplacement samples, but we always use the descriptor *replacement* when discussing replacement samples. A replacement sample can be considered to be a random sample selected from an infinite population with the value $p_{ri}^{-1} y_i =: z_i$

occurring with frequency p_{ri}, where the symbol =: means "is defined to equal." Thus, the variance of the infinite population of z's is

$$\sigma_z^2 = E\{(z_i - \mu_z)^2\} = \sum_{i=1}^{N} p_{ri}(z_i - \mu_z)^2$$
$$= \sum_{i=1}^{N} p_{ri}(p_{ri}^{-1}y_i - T_y)^2, \qquad (1.2.67)$$

where

$$\mu_z = \sum_{i=1}^{N} p_{ri}z_i = \sum_{i=1}^{N} y_i = T_y.$$

The estimator $\hat{T}_{yR}$ is the mean of n *iid* random variables with mean μ_z and variance σ_z^2. Thus,

$$V\{\hat{T}_{yR}\} = V\left\{n^{-1}\sum_{d=1}^{n} z_d\right\} = n^{-1}\sigma_z^2, \qquad (1.2.68)$$

where z_d is the value of z obtained on the dth draw. Furthermore,

$$\hat{V}\{\hat{T}_{yr}\} = n^{-1}(n-1)^{-1}\sum_{d=1}^{n}(z_d - \bar{z}_n)^2, \qquad (1.2.69)$$

where

$$\bar{z}_n = n^{-1}\sum_{d=1}^{n} z_d$$

is unbiased for $V\{\hat{T}_{yR}\}$. The simplicity of the estimator (1.2.69) has led to its use as an approximation in nonreplacement unequal probability sampling when all of the np_{ri} are small.

The fact that some elements can be repeated in the estimator (1.2.66) is an unappealing property. That the estimator is not efficient is seen most easily when all p_{ri} are equal to N^{-1}. Then

$$\hat{T}_{yR} = Nn^{-1}\sum_{i\in A} y_i t_i. \qquad (1.2.70)$$

Because the draws are independent, the sample of unique elements is a simple random nonreplacement sample. Thus, an unbiased estimator of the mean is

$$\bar{y}_u = n_u^{-1}\sum_{i\in A} y_i, \qquad (1.2.71)$$

where n_u is the number of unique elements in the sample. The conditional variance of the mean associated with (1.2.66) conditional on $(t_1, t_2, \ldots, t_{n_u})$ is

$$V\{N^{-1}\hat{T}_{yR} \mid (t_1, t_2, \ldots, t_{n_u})\} = n^{-2}\sum_{i\in A} t_i^2\sigma_y^2, \quad (1.2.72)$$

while

$$V\{\bar{y}_u \mid (t_1, t_2, \ldots, t_{n_u})\} = n_u^{-1}\sigma_y^2. \quad (1.2.73)$$

Because $\Sigma t_i^2 \geq n_u$, with equality for $n_u = n$, the mean of unique units is conditionally superior to the mean associated with (1.2.66) for every $1 < n_u < n$.

1.2.6 Rejective sampling

Rejective sampling is a procedure in which a sample is selected by a particular rule but is accepted only if it meets certain criteria. The selection operation is repeated until an acceptable sample is obtained. The procedure is sometimes called *restrictive sampling*. In most situations, the rejection of certain samples changes the inclusion probabilities. Hájek (1964, 1981) studied two kinds of rejective sampling. In the first, a replacement sample is selected and the sample is kept only if it contains no duplicates. In the second, a Poisson sample is selected and is kept only if it contains exactly the number of elements desired.

To illustrate the effect of the restriction on probabilities, consider the selection of a Poisson sample from a population of size 4 with selection probabilities (0.2, 0.4, 0.6, 0.8) for i =1, 2, 3, 4. Let the sample be rejected unless exactly two elements are selected. The probabilities of the six possible samples of size 2 are (0.0064, 0.0144, 0.0384, 0.0384, 0.1024, 0.2304) for the samples [(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)], respectively. It follows that the rejective procedure gives inclusion probabilities (0.1375, 0.3420, 0.6580, 0.8625) for $i = 1, 2, 3, 4$. See Section 1.4 for references on the use of rejective sampling with unequal probabilities.

To illustrate how the inclusion probabilities are changed by other rejection rules, consider the selection of a sample of size 3 from a sample of 6, where the elements are numbered from 1 to 6. Assume that the procedure is to select a simple random sample of size 3 but to reject the sample if it contains three adjacent elements. Thus, samples (1, 2, 3), (2, 3, 4), (3, 4, 5), and (4, 5, 6) are rejected. If the sample is rejected, a new simple random sample is selected until an acceptable sample is obtained. There are 20 possible simple random samples and 16 acceptable samples. Therefore, the probabilities of inclusion

in an acceptable sample are (9/16, 8/16, 7/16, 7/16, 8/16, 9/16) for elements (1, 2, 3, 4, 5, 6), respectively.

As a second example, let x be the ordered identification for a population of size 8. Assume that we select samples using simple random sampling but reject any sample with a mean of x less than 2.5 or greater than 6.5. Thus, the four samples (1, 2, 3), (1, 2, 4), (5, 7, 8), and (6, 7, 8) are rejected. The resulting probabilities of inclusion are (19/52, 19/52, 20/52, 20/52, 20/52, 20/52, 19/52, 19/52) for elements (1, 2, 3, 4, 5, 6, 7, 8), respectively. These two examples illustrate the general principles that rejecting adjacent items increases the relative probability of boundary elements, and rejecting samples with large $\mid \bar{x}_n - \bar{x}_N \mid$ decreases the relative probability of extreme observations. In these simple examples, one can construct the Horvitz–Thompson estimator using the correct inclusion probabilities.

Many practitioners employ modest types of rejective sampling when the unit identification carries information. For example, let an ordered population be divided into m strata of size k, with two elements selected in each stratum. Practitioners would be tempted to reject a sample composed of the two largest elements in each stratum. The probability of such a sample is $[0.5k(k-1)]^{-m}$ for m strata of size k. If only this sample and the similar sample of the two smallest elements are rejected, the inclusion probabilities will be little affected for large k and m. On the other hand, if a large fraction of possible samples are rejected, the inclusion probabilities can be changed by important amounts.

1.2.7 Cluster samples

In much of the discussion to this point we have considered a conceptual list of units, where the units can be given an identification and the identifications can be used in sample selection. In Example 1.2.1 we introduced the possibility that the units on the frame are not the units of final interest. In that example, households are of interest and are the units observed, but the units sampled are blocks, where there will be several households in a block. Samples of this type are called *cluster samples*. It is also possible for the units of analysis to differ from the sampling units and from the observation units. Assume that data are collected for all persons in a household using a single respondent for the household and that the analyst is interested in the fraction of people who had flu shots. Then the analysis unit is a person, the observation unit is the household, and the sampling unit is the block.

In estimation formulas such as (1.2.23) and (1.2.33), the variable y_i is the total for the ith sampling unit. In Example 1.2.1, y_i is the total for a block. It is very easy for analysts to treat analysis units or observation units incorrectly as sampling units. One must always remember the nature of the units on the sampling frame.

From a statistical point of view, no new concepts are involved in the construction of estimators for cluster samples. If we let M_i be the number of elements in the ith cluster and let y_{ij} be the value for the jth element in the ith cluster, then

$$y_i = \sum_{j=1}^{M_i} y_{ij}$$

and the estimator of a total is

$$\hat{T}_y = \sum_{i \in A} \pi_i^{-1} y_i, \tag{1.2.74}$$

where π_i is the probability of selection for the ith cluster and y_i is the total of the characteristic for all persons in the ith cluster. Similarly, variance estimators such as (1.2.32) and (1.2.33) are directly applicable.

1.2.8 Two-stage sampling

In many situations it is efficient first to select a sample of clusters and then select a subsample of the units in each cluster. In this case, the cluster is called a *primary sampling unit* (PSU), and the sample of primary sampling units is called the *first-stage sample*. The units selected in the subsample are called *secondary sampling units* (SSUs), and the sample of secondary sampling units is called the *second-stage sample*.

We adopt the convention described by Särndal, Swensson, and Wretman (1992, p. 134). If the sample is selected in two steps (stages), if units selected at the second step are selected independently in each first-step unit, and if the rules for selection within a first-step unit depend only on that unit and not on other first step units in the sample, the sample is called a *two-stage sample*.

The Horvitz–Thompson estimator of the total for a two-stage sample is

$$\hat{T}_{2s} = \sum_{i \in A_1} \sum_{j \in B_i} \pi_{(ij)}^{-1} y_{ij}, \tag{1.2.75}$$

where $\pi_{(ij)} = \pi_i \pi_{(ij)|i}$ is the probability that second-stage unit ij is selected in the sample, π_i is the probability that first-stage unit i is selected, $\pi_{(ij)|i}$ is the probability that second-stage unit ij is selected given that first-stage unit i is selected, A_1 is the set of indices for first-stage units in the sample, and B_i is the set of second-stage units in first-stage unit i that are in the sample.

The estimator (1.2.75) is unbiased for the total by the properties of the Horvitz–Thompson estimator. The joint probabilities are

$$\begin{aligned} \pi_{(ij)(km)} &= \pi_i \pi_{(ij)(im)|i} && \text{if } i = k \text{ and } j \neq m \\ &= \pi_{ik} \pi_{(ij)|i} \pi_{(km)|k} && \text{if } i \neq k \text{ and } j \neq m, \end{aligned} \tag{1.2.76}$$

where π_{ik} is the probability that first-stage units i and k are selected, and $\pi_{(ij)(im)|i}$ is the probability that elements ij and im are selected given that PSU i is selected. Given these probabilities, the variances and estimated variances of the Horvitz–Thompson estimator are defined. We present some more convenient expressions for the variance and estimated variance.

Consider a sample of n_1 PSUs selected from a finite population which is, itself, a sample of N PSUs selected from an infinite population of PSUs. Let the ith PSU be selected with probability π_i and let the ith PSU contain M_i secondary sampling units. Let a nonreplacement probability sample of m_i units be selected from the M_i. Then an alternative expression for the estimator of (1.2.75) is

$$\hat{T}_{2s} = \sum_{i \in A_1} \pi_i^{-1} \hat{y}_i, \tag{1.2.77}$$

where

$$\hat{y}_i = \sum_{j \in B_i} \pi_{(ij)|i}^{-1} y_{ij}$$

and B_i is as defined in (1.2.75). The design variance of $\hat{T}_{2s}$ is

$$\begin{aligned} V\{\hat{T}_{2s} \mid \mathcal{F}\} &= V\{E[\hat{T}_{2s} \mid (A_1, \mathcal{F})] \mid \mathcal{F}\} + E\{V[\hat{T}_{2s} \mid (A_1, \mathcal{F})] \mid \mathcal{F}\} \\ &= V_1\{\hat{T}_{1s} \mid \mathcal{F}\} + E\{V[\hat{T}_{2s} \mid (A_1, \mathcal{F})] \mid \mathcal{F}\}, \end{aligned} \tag{1.2.78}$$

where $\hat{T}_{1s}$ is the estimated total with all $m_i = M_i$, $V_1\{\hat{T}_{1s} \mid \mathcal{F}\}$ is the variance of the estimated total with $m_i = M_i$ for all i, and $V[\hat{T}_{2s}|(A_1, \mathcal{F})]$ is the conditional design variance, conditional on the first-stage units selected. Generally, a design consistent estimator of $V[\hat{T}_{2s} \mid (A_1, \mathcal{F})]$ is available and can be used to estimate $E\{V[\hat{T}_{2s} \mid (A_1, \mathcal{F})]\}$. Estimation of $V_1\{\hat{T}_{1s} \mid \mathcal{F}\}$ is more difficult. Consider a quadratic function of the y_i, such as the Horvitz–Thompson estimator, for $V_1\{\hat{T}_{1s} \mid \mathcal{F}\}$,

$$\tilde{V}_1\{\hat{T}_{1s} \mid \mathcal{F}\} = \sum_{i \in A_1} \sum_{j \in A_1} \alpha_{ij} y_i y_j,$$

where α_{ij} are fixed coefficients. If y_i is replaced with $\hat{y}_i$, we have

$$\begin{aligned} E\left\{ \sum_{i \in A_1} \sum_{j \in A_1} \alpha_{ij} \hat{y}_i \hat{y}_j \mid (A_1, \mathcal{F}) \right\} &= \sum_{i \in A_1} \sum_{j \in A_1} \alpha_{ij} y_i y_j \\ &\quad + \sum_{i \in A_1} \alpha_{ii} V\{\hat{y}_i \mid (A_1, \mathcal{F})\}, \end{aligned} \tag{1.2.79}$$

because

$$E\left\{\hat{y}_i^2 \mid (A_1, \mathcal{F})\right\} = y_i^2 + V\left\{\hat{y}_i - y_i \mid (A_1, \mathcal{F})\right\}, \tag{1.2.80}$$

and given that samples are selected independently within each PSU,

$$E\left\{\hat{y}_i \hat{y}_j \mid (A_1, \mathcal{F})\right\} = y_i y_j \qquad \text{for } i \neq j. \tag{1.2.81}$$

Thus, given a quadratic estimator of variance for the first stage, an estimator of $V\{\hat{T}_{2s} \mid \mathcal{F}\}$ is

$$\hat{V}\{\hat{T}_{2s} \mid \mathcal{F}\} = \hat{V}_1\{\hat{T}_{1s} \mid \mathcal{F}\} + \sum_{i \in A_1} (\pi_i^{-2} - \alpha_{ii}) \hat{V}\{\hat{y}_i \mid (A_1, \mathcal{F})\}, \tag{1.2.82}$$

where $\hat{V}_1\{\hat{T}_{1s} \mid \mathcal{F}\}$ is the estimated design variance for the first-stage sample computed with $\hat{y}_i$ replacing y_i.

The coefficient for y_i^2 in the design variance of a design linear estimator is $\pi_i^{-1}(1 - \pi_i)$. It follows that the α_{ii} in a design-unbiased quadratic estimator of the variance of a design linear estimator is $\pi_i^{-2}(1 - \pi_i)$. Therefore, the bias in $\hat{V}_1\{\hat{T}_{2s} \mid \mathcal{F}\}$ as an estimator of $V\{\hat{T}_{2s} \mid \mathcal{F}\}$ is the sum of the $\pi_i V\{y_i \mid (A_1, \mathcal{F})\}$, and the bias is small if the sampling rates are small.

For a simple random sample of PSUs, the variance of the estimator of the total for a complete first stage is

$$V_1\{\hat{T}_{1s} \mid \mathcal{F}\} = N^2(1 - f_1) n_1^{-1} S_{1y}^2, \tag{1.2.83}$$

where $f_1 = N^{-1} n_1$,

$$S_{1y}^2 = (N-1)^{-1} \sum_{i \in U} (y_i - \bar{y}_N)^2,$$

$$y_i = \sum_{j=1}^{M_i} y_{ij},$$

and $\bar{y}_N = N^{-1} \Sigma_{i \in U} y_i$. Given that the samples within the first-stage units are simple random nonreplacement samples,

$$V\{\hat{T}_{2s} \mid (A_1, \mathcal{F})\} = \sum_{i \in A_1} \pi_i^{-2} M_i^2 \left(1 - M_i^{-1} m_i\right) m_i^{-1} S_{2yi}^2, \tag{1.2.84}$$

where

$$S_{2yi}^2 = (M_i - 1)^{-1} \sum_{j=1}^{M_i} (y_{ij} - \bar{y}_{Mi})^2$$

and

$$\bar{y}_{Mi} = M_i^{-1} \sum_{j=1}^{M_i} y_{ij}.$$

For simple random sampling at the second stage,

$$\hat{V}\{\hat{y}_i \mid (A_1, \mathcal{F})\} = M_i^2 \left(1 - M_i^{-1} m_i\right) m_i^{-1} s_{2yi}^2, \quad (1.2.85)$$

where

$$s_{2yi}^2 = (m_i - 1)^{-1} \sum_{j \in B_i} (y_{ij} - \bar{y}_i)^2$$

and

$$\bar{y}_{i\cdot} = m_i^{-1} \sum_{j \in B_i} y_{ij}$$

is design consistent for $E\{V[\hat{T}_{2s} \mid (A_1, \mathcal{F})] \mid \mathcal{F}\}$. The expected value of the estimator of the variance (1.2.83) constructed by replacing y_i with $\hat{y}_i$ is

$$\begin{aligned} E\{\hat{V}_{1,srs}(\hat{T}_{1s} \mid \mathcal{F}) \mid \mathcal{F}\} &= E[C_f (n_1 - 1)^{-1} \sum_{i \in A_1} (\hat{y}_i - \bar{y}_{n1})^2 \mid \mathcal{F}] \\ &= C_f S_{1y}^2 + C_f n_1^{-1} \sum_{i \in A_1} M_i (M_i - m_i) m_i^{-1} S_{2yi}^2, \end{aligned} \quad (1.2.86)$$

where $C_f = N^2 (1 - f_1) n_1^{-1}$, $f_1 = N^{-1} n_1$, and

$$\bar{y}_{n1} = n_1^{-1} \sum_{i \in A_1} \hat{y}_i.$$

Therefore, an unbiased estimator of the variance of $\hat{T}_{2s}$ is, for simple random nonreplacement sampling at both stages,

$$\begin{aligned} \hat{V}\{\hat{T}_{2s} \mid \mathcal{F}\} &= \hat{V}_{1,srs}\{\hat{T}_{2s} \mid \mathcal{F}\} \\ &\quad + f_1 N^2 n_1^{-2} \sum_{i \in A_1} M_i^2 \left(1 - M_i^{-1} m_i\right) m_i^{-1} s_{2yi}^2, \end{aligned} \quad (1.2.87)$$

where $\hat{V}_{1,srs}\{\hat{T}_{1s} \mid \mathcal{F}\}$ is defined in (1.2.86) and s_{2yi}^2 is defined in (1.2.85). The first-stage estimated variance in (1.2.86) is a quadratic in y_i with $\alpha_{ii} = (1 - f_1) N^2 n_1^{-2}$. Furthermore, $\pi_i^{-2} = N^2 n_1^{-2}$ and $\hat{V}_{1,srs}\{\hat{T}_{2s} \mid \mathcal{F}\}$ is the

dominant term in $\hat{V}\{\hat{T}_{2s} \mid \mathcal{F}\}$ when the finite population correction is close to 1.

To construct estimator (1.2.87), we require that $m_i \geq 2$ for all i and require the assumption of independent simple random nonreplacement samples within first-stage units. Estimator (1.2.82) only requires that the second-stage design be such that a reasonable estimator of the second-stage variance is available for every PSU.

If the finite population correction can be ignored, the estimator $\hat{V}_1\{\hat{T}_{2s} \mid \mathcal{F}\}$ is consistent for the variance under any selection scheme for secondary units, such that $\hat{T}_{2s}$ is an unbiased estimator and the selection within a PSU is independent of the selection in other PSUs. This follows from (1.2.80) and (1.2.81). Thus, for example, one could stratify each of the PSUs and select stratified samples of secondary units within each PSU.

Example 1.2.2. We use data from the U.S. National Resources Inventory (NRI) in a number of examples. The NRI is conducted by the U.S. Natural Resources Conservation Service in cooperation with the Iowa State University Center for Survey Statistics and Methodology. The survey is a panel survey of land use conducted in 1982, 1987, 1992, 1997, and yearly since 2000. Data are collected on soil characteristics, land use, land cover, wind erosion, water erosion, and conservation practices. The sample is a stratified area sample of the United States, where the primary sampling units are areas of land called *segments*. Data are collected for the entire segment on such items as urban lands, roads, and water. Detailed data on soil properties and land use are collected at a random sample of points within the segment. The sample for 1997 contained about 300,000 segments with about 800,000 points. The yearly samples are typically about 70,000 segments. See Nusser and Goebel (1997) for a more complete description of the survey.

We use a very small subsample of the Missouri NRI sample for the year 1997 to illustrate calculations for a two-stage sample. The true first-stage sampling rates are on the order of 2%, but for the purposes of illustration, we use the much higher rates of Table 1.2. In Missouri, segments are defined by the Public Land Survey System. Therefore, most segments are close to 160 acres in size, but there is some variation in size due to variation in sections defined by the Public Land Survey System and due to truncation associated with county boundaries. The segment size in acres is given in the fourth column of the table. The points are classified using a system called *broaduse*, where example broaduses are urban land, cultivated cropland, pastureland, and forestland. Some of the broaduses are further subdivided into categories called *coveruses*, where corn, cotton, and soybeans are some of the coveruses within the cropland broaduse.

Table 1.2 Missouri NRI Data

Stratum	PSU	Weight	Segment Size	Total No. Pts	No. Pts. Forest	s^2_{2yi}	$\hat{y}_i$
1	1	3.00	195	3	2	0.1111	130
	2	3.00	165	3	3	0	165
	3	3.00	162	3	2	0.1111	54
	4	3.00	168	3	0	0	0
	5	3.00	168	3	2	0.1111	112
	6	3.00	100	2	1	0.2500	50
	7	3.00	180	3	0	0	0
2	1	5.00	162	3	1	0.1111	54
	2	5.00	174	3	1	0.1111	58
	3	5.00	168	3	2	0.1111	112
	4	5.00	174	3	0	0	0

In this example, we estimate the acres of forestland and define

$$\begin{aligned} y_{ij} &= 1 \quad \text{if point } j \text{ in PSU } i \text{ is forest} \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

The total number of points in the segment is given in the fifth column and the number that are forest is given in the sixth column. In a typical data set there would be a row for each point, and the sum for the segment would be calculated as part of the estimation program. Treating each point as if it represents 1 acre, we have

$$\hat{y}_{hi} = M_{hi} m_{hi}^{-1} \sum_{j=1}^{m_{hi}} y_{hij},$$

where M_{hi} represents the acres (SSUs) in segment i of stratum h and m_{hi} the number of sample points (SSUs) in the segment. Thus, the estimated total acres of forest for PSU 1 in stratum 1 is 130, and the estimated variance for that estimated segment total is

$$\hat{V}\{\bar{y}_{1,1} \mid (A_1, \mathcal{F})\} = 195(195-3)3^{-1}(0.1111) = 1386.67,$$

where $s^2_{2y,1,1} = 0.1111$ is as defined in (1.2.85).

The estimated acres of forest for this small region is

$$\hat{T}_{2s} = \sum_{h=1}^{2} N_h n_{1h}^{-1} \sum_{i \in A_{1h}} \hat{y}_{hi} = 3(514) + 5(222) = 2652,$$

where n_{1h} is the number of sample segments (PSUs) in stratum h.

Equation (1.2.87) extends immediately to stratified sampling and we have

$$\begin{aligned} \hat{V}\{\hat{T}_{2s} \mid \mathcal{F}\} &= \sum_{h=1}^{2} N_h^2 (1 - f_{1h}) n_{1h}^{-1} \hat{s}_{1h}^2 \\ &\quad + \sum_{h=1}^{2} N_h^2 n_{1h}^{-2} f_{1h} \sum_{i=1}^{n_{1h}} M_{hi}(M_{hi} - m_{hi}) m_{hi}^{-1} s_{2y,hi}^2 \\ &= 340.940 + 29.190 = 370.130, \end{aligned}$$

where $f_{1h} = N_h^{-1} n_{1h}$,

$$\hat{s}_{1h}^2 = (n_{1h} - 1)^{-1} \sum_{i \in A_{1h}} (\hat{y}_{hi} - \bar{y}_{h,n1})^2,$$

and $\bar{y}_{h,n1}$ is the stratum analog of $\bar{y}_{n1}$ of (1.2.86). The values of the first-stage estimated variances are $(\hat{s}_{1,1}^2, \hat{s}_{1,2}^2) = (4130.3, 2093.3)$. There is a sizable correlation between points within a segment for a broaduse such as forest, and the between-PSU portion dominates the variance. ■■

1.3 LIMIT PROPERTIES

1.3.1 Sequences of estimators

We define sequences that will permit us to establish large-sample properties of sample designs and estimators. Our sequences will be sequences of finite populations and associated probability samples. A set of indices is used to identify the elements of each finite population in the sequence. To reduce the number of symbols required, we usually assume that the Nth finite population contains N elements. Thus, the set of indices for the Nth finite population is

$$U_N = \{1, 2, \ldots, N\}, \tag{1.3.1}$$

where $N = 1, 2, \ldots$. Associated with the ith element of the Nth population is a column vector of characteristics, denoted by $\mathbf{y}_{iN}$. Let

$$\mathcal{F}_N = (\mathbf{y}_{1N}, \mathbf{y}_{2N}, \ldots, \mathbf{y}_{NN})$$

be the set of vectors for the Nth finite population. The set $\mathcal{F}_N$ is often called simply the Nth *finite population* or the Nth *finite universe*.

Two types of sequences $\{\mathcal{F}_N\}$ may be specified. In one, the set $\mathcal{F}_N$ is a set of fixed vectors from a fixed sequence. In the other, the vectors

$\mathbf{y}_{iN}$, $i = 1, 2, \ldots, N$, are random variables. For example, the $\{\mathbf{y}_{iN}\}$, $i = 1, 2, \ldots, N$, might be the first N elements of the sequence $\{y_i\}$ of *iid* random variables with distribution function $F(y)$ such that

$$E\{y_i\} = \mu \tag{1.3.2}$$

and

$$E\{(y_i - \mu)^2\} = \sigma^2. \tag{1.3.3}$$

If necessary to avoid confusion, we will add subscripts so that, for example, μ_y denotes the mean of y and σ_y^2 or σ_{yy} denotes the variance of y.

As defined previously, the finite population mean and variance for scalar y are

$$\bar{y}_N = N^{-1} \sum_{i=1}^{N} y_{iN} \tag{1.3.4}$$

and

$$S_{y,N}^2 = (N-1)^{-1} \sum_{i=1}^{N} (y_{iN} - \bar{y}_N)^2. \tag{1.3.5}$$

The corresponding quantities for vectors are

$$\bar{\mathbf{y}}_N = N^{-1} \sum_{i=1}^{N} \mathbf{y}_{iN} \tag{1.3.6}$$

and

$$\mathbf{S}_{yy,N} = (N-1)^{-1} \sum_{i=1}^{N} (\mathbf{y}_{iN} - \bar{\mathbf{y}}_N)(\mathbf{y}_{iN} - \bar{\mathbf{y}}_N)'. \tag{1.3.7}$$

Recall that a sample is defined by a subset of the population indices and let A_N denote the set of indices appearing in the sample selected from the Nth finite population. The number of distinct indices appearing in the sample is called the *sample size* and is denoted by n_N. We assume that samples are selected according to the probability rule $p_N(A)$ introduced in Section 1.2.

Example 1.3.1. As an example of a sequence of populations, consider the sequence of sets of $N = 10j$ elements, where $j = 1, 2, \ldots$. Let *iid* Bernoulli random variables be associated with the indexes $1, 2, \ldots, N$. From each set of $10j$ values realized, a simple random nonreplacement sample

of size $n_N = j$ is selected. In this case it is possible to give the exact form of the relevant distributions. Assume that the Bernoulli variable is such that

$$\begin{aligned} x_i &= 1 \quad \text{with probability } p \\ &= 0 \quad \text{with probability } (1-p). \end{aligned}$$

Then the distribution of

$$X_N = \sum_{i=1}^{N} x_i$$

is that of a binomial random variable with parameters $(N,\ p)$ and

$$P\{X_N = a\} = \binom{N}{a} p^a (1-p)^{N-a}.$$

Because the elements are independent, the unconditional distribution of the sample sum, X_n, is binomial with parameters $(n,\ p)$,

$$P\{X_n = a\} = \binom{n}{a} p^a (1-p)^{n-a}.$$

Now a particular finite population, $\mathcal{F}_N$, has X_N elements equal to 1. The conditional distribution of X_n given $\mathcal{F}_N$ is the hypergeometric distribution and

$$P\{X_n = a \mid \mathcal{F}_N\} = \binom{X_N}{a} \binom{N - X_N}{j-a} \left[\binom{N}{j}\right]^{-1}$$

■■

A fully specified sequence will contain a description of the structure of the finite populations and of the sampling probability rules. For example, it might be assumed that the finite population is composed of N *iid* random variables with properties (1.3.2) and (1.3.3), and that the samples are simple nonreplacement samples of size n_N selected from the N population elements. In that situation, a simple random sample of size n_N selected from the finite universe is a set of *iid* random variables with common distribution function $F_y(y)$. A proof, due to F. Jay Breidt, is given in Theorem 1.3.1.

Theorem 1.3.1. Suppose that $y_1,\ y_2,\ \ldots,\ y_N$ are *iid* with distribution function $F(y)$ and corresponding characteristic function $\varphi(t) = E\{e^{ity}\}$. Let $\mathbf{d} = (I_1,\ I_2,\ \ldots,\ I_N)'$ be a random vector with each component supported on $\{0,\ 1\}$. Assume that $\mathbf{d}$ is independent of $(y_1,\ y_2,\ \ldots,\ y_N)'$. Let

$U = \{1, 2, \ldots, N\}$ and define $A = \{k \in U : I_k = 1\}$. If A is nonempty, the random variables $(y_k, k \in A) \mid \mathbf{d}$ are *iid* with characteristic function $\varphi(t)$.

Proof. Let $(t_1, t_2, \ldots, t_N)'$ be an element of N-dimensional Euclidean space. Then, given $\mathbf{d}$, the joint characteristic function of $(y_k, k \in A)$ is

$$\begin{aligned} E\left\{exp\left(\mathrm{i}\sum_{k\in A} t_k y_k\right) \middle| \mathbf{d}\right\} &= E\left\{exp\left(\mathrm{i}\sum_{k\in U} t_k I_k y_k\right) \middle| \mathbf{d}\right\} \\ &= E\left\{\prod_{k\in U} exp(\mathrm{i}t_k I_k y_k) \mid \mathbf{d}\right\} \\ &= \prod_{k\in U} E\{exp(it_k I_k y_k) \mid \mathbf{d}\} \\ &= \prod_{k\in U} \varphi(t_k I_k) \\ &= \prod_{k\in A} \varphi(t_k), \end{aligned} \tag{1.3.8}$$

since $\varphi(0) = 1$. The result follows because (1.3.8) is the characteristic function of $n = \Sigma_{k\in U} I_k$ *iid* random variables with distribution function $F(y)$. ■

The crucial assumption of the theorem is that the probability rule defining membership in the sample, the probability function for $\mathbf{d}$, is independent of $(y_1, y_2, \ldots, y_N)$. It then follows that given $\mathbf{d}$ with component support on $\{0, 1\}$, the sets $\{y_k, k \in A\}$ and $\{y_k, k \notin A\}$ are sets of n and $N - n$ *iid* random variables with distribution function $F_y(y)$. Furthermore, the two sets are independent. The conditional distribution of the two sets is the same for all $\mathbf{d}$ with the same sample size, where the sample size is

$$n = \sum_{k=1}^{N} I_k.$$

Thus, for fixed-sample-size nonreplacement designs and *iid* random variables, the unconditional distribution over all samples is the same as the conditional distribution for a particular sample set of indices.

Example 1.3.2. As a second example of a sequence of populations and samples, let $\mathcal{F}_N = (y_1, y_2, \ldots, y_N)$ be the first N elements in a sequence of independent random variables selected from a normal distribution with mean

μ_y and variance σ_y^2. Let n_N be the largest integer less than or equal to fN, where f is a fixed number in (0, 1). Assume that a simple random sample of size n_N is selected from $\mathcal{F}_N$, let A_N be the set of indices of the sample selected, and let A_N^c be the set of indexes of the $N - n_N$ elements not in A_N. The n_N sample elements are $NI(\mu_y, \sigma_y^2)$ random variables, independent of the $N - n_N$ nonsample $NI(\mu_y, \sigma_y^2)$ random variables. It follows that

$$\bar{y}_n \sim N(\mu_y, n_N^{-1}\sigma_y^2),$$

$$\bar{y}_{N-n} \sim N\{\mu_y, (N - n_N)^{-1}\sigma_y^2\},$$

$$\bar{y}_n - \bar{y}_N \sim N\{0, n_N^{-1}(1 - f_N)\sigma_y^2\},$$

and

$$\left[(1 - f_N)n_N^{-1}s_{y,n}^2\right]^{-1/2}(\bar{y}_n - \bar{y}_N) \sim t_{n-1},$$

where $f_N = N^{-1}n_N$,

$$s_{y,n}^2 = (n_N - 1)^{-1}\sum_{i \in A_N}(y_i - \bar{y}_n)^2,$$

$$\bar{y}_n = n_N^{-1}\sum_{i \in A_N} y_i,$$

$$\bar{y}_{N-n} = (N - n_N)^{-1}\sum_{i \in A_N^c} y_i,$$

and t_{n-1} is Student's t-distribution with $n_N - 1$ degrees of freedom. ■■

Given a model for the stochastic mechanism generating the finite population, we can consider expectations conditional on properties of the random variables. Most often we are interested in a set of samples with some of the same characteristics as those observed in the current sample.

Example 1.3.3. Let $\mathcal{F}_N = [(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)]$ be the first N elements in a sequence of independent random variables from a bivariate normal distribution,

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim NI\left[\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}\right].$$

Let a sequence of simple random samples be selected as described in Example 1.3.2. Let $\mathbf{x}_n = (x_1, x_2, \ldots, x_n)$, let $\mathbf{z}_i = (x_i, y_i)$, and let

$$\begin{pmatrix} s_{x,n}^2 & s_{xy,n} \\ s_{xy,n} & s_{y,n}^2 \end{pmatrix} = (n - 1)^{-1}\sum_{i \in A_N}(\mathbf{z}_i - \bar{\mathbf{z}}_n)'(\mathbf{z}_i - \bar{\mathbf{z}}_n),$$

where $\bar{z}_n$ is the simple sample mean. The least squares regression coefficient for the regression of y on x is

$$\hat{\beta}_n = s_{x,n}^{-2} s_{xy,n}.$$

By Theorem 1.3.1, the sample is a realization of *iid* normal vectors. It follows that under the model, we have the conditional mean and variance,

$$E\{\hat{\beta}_n \mid \mathbf{x}_n\} = \beta$$

and

$$V\{\hat{\beta}_n \mid \mathbf{x}_n\} = \left[\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\right]^{-1} \sigma_e^2,$$

where $\beta = \sigma_x^{-2}\sigma_{xy}$ and $\sigma_e^2 = \sigma_y^2 - \beta\sigma_{xy}$. ■■

In describing rates of convergence for real-valued sequences and for sequences of random variables, a notation for order is useful. We use the conventions given by Fuller (1996, Chapter 5). See Appendix 1A.

Example 1.3.4. In previous examples we have considered sequences of finite populations generated by a random mechanism. To study sampling properties for a sequence of finite populations generated from a fixed sequence, let $\{y_i\}$ be a sequence of real numbers and assume that

$$\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}(y_i,\ y_i^2) = (\theta_1,\ \theta_2),$$

where $(\theta_1,\ \theta_2)$ are finite and $\theta_2 - \theta_1^2 > 0$. Define a sequence of finite populations $\{\mathcal{F}_N\}$, where the Nth finite population is composed of the first N values of the sequence $\{y_i\}$. Let a simple random sample of size $n_N = [fN]$ be selected from the Nth finite population, where $0 < f < 1$ and $[fN]$ is the largest integer less than or equal to fN. By the results of Section 1.2,

$$V\{\bar{y}_n - \bar{y}_N \mid \mathcal{F}_N\} = (1 - f_N)n_N^{-1}S_{y,N}^2,$$

where $f_N = N^{-1}n_N$. By assumption,

$$\lim_{N\to\infty} S_{y,N}^2 = \theta_2 - \theta_1^2$$

is a finite positive number. It follows that we can write

$$V\{\bar{y}_n - \bar{y}_N \mid \mathcal{F}_N\} = O(n_N^{-1})$$

and

$$\bar{y}_n - \bar{y}_N \mid \mathcal{F}_N \;=\; O_p(n_N^{-1/2}),$$

where $\{(\bar{y}_n - \bar{y}_N) \mid \mathcal{F}_N\}$ denotes the sequence of $\bar{y}_n - \bar{y}_N$ calculated from the sequence of samples selected from the sequence $\{\mathcal{F}_N\}$. Because the sampling fraction is fixed, n_N and N are of the same order. For a sequence of finite populations that is generated from a sequence of fixed numbers such as this example, the notational reference to $\mathcal{F}_N$ is not required because the random variation comes only from the design. In complex situations the notation serves to identify properties derived from the sampling design. Even in situations where not required, we often employ the notation. ■■

Once the sequence of populations, sample designs, and estimators is specified, the properties of the sequence of estimators can be obtained. The unconditional properties of the estimator, the properties conditional on the particular finite population, and the properties conditional on some attributes of the particular sample are all of interest. Because of the central importance of the sampling design, it is common in the survey sampling literature to use the term *design consistent* for a procedure that is consistent conditional on the particular sequence of finite populations. The sequence of populations can be composed of fixed numbers, as in Example 1.3.4, or can be a sequence of random variables, as in Example 1.3.2. For a sequence of random variables, the property is assumed to hold almost surely (a.s.); that is, the property holds for all sequences except for a set of measure zero.

Definition 1.3.1. Given a sequence of finite populations $\{\mathcal{F}_N\}$ and an associated sequence of sample designs, the estimator $\hat{\theta}$ is *design consistent* for the finite population parameter θ_N if for every $\epsilon > 0$,

$$\lim_{N\to\infty} P\{\mid \hat{\theta} - \theta_N \mid > \epsilon \mid \mathcal{F}_N\} \;=\; 0 \quad \text{a.s.}, \tag{1.3.9}$$

where the notation indicates that for the sequence of finite populations, the probability is that determined by the sample design.

Observe that $\bar{y}_n$ of Example 1.3.4 is design consistent for $\bar{y}_N$ because $V\{\bar{y}_n - \bar{y}_N \mid \mathcal{F}_N\}$ is $O(n_N^{-1})$.

Example 1.3.5. For the sequence of populations and samples of Example 1.3.2,

$$V\{\bar{y}_n - \bar{y}_N \mid \mathcal{F}_N\} = (1 - f_N)n_N^{-1}S_{y,N}^2,$$

where $f_N = N^{-1}n_N$ and

$$S_{y,N}^2 = (N-1)^{-1}\sum_{i=1}^{N}(y_i - \bar{y}_N)^2.$$

The sequence of populations is created from a sequence of $NI(\mu, \sigma_y^2)$ random variables. Therefore,

$$\lim_{N\to\infty} S_{y,N}^2 = \sigma_y^2 \quad \text{a.s.}$$

It follows that

$$V\{\bar{y}_n - \bar{y}_N \mid \mathcal{F}_N\} = O_p(n_N^{-1}) \quad \text{a.s.},$$

$$(\bar{y}_n - \bar{y}_N) \mid \mathcal{F}_N = O_p(n_N^{-1/2}) \quad \text{a.s.}$$

and hence $\bar{y}_n$ is design consistent for $\bar{y}_N$. ■■

1.3.2 Central limit theorems

Central limit theorems are critical to our ability to make probability statements on the basis of sample statistics. Our first results are for a stratified finite population, where the strata are composed of realizations of *iid* random variables. Under mild conditions, the properly standardized stratified mean converges to a normal random variable. In the theorem statement, $N(0, \sigma^2)$ denotes the normal distribution with mean zero and variance σ^2, and the symbol $\xrightarrow{\mathcal{L}}$ is used to denote convergence in distribution (convergence in law).

Theorem 1.3.2. Let $\{\mathcal{F}_N\}$, where $\mathcal{F}_N = \{y_{hiN}\}$, $h = 1, 2, \ldots, H_N$; $i = 1, 2, \ldots, N_{hN}$, be a sequence of finite populations composed of H_N strata, where the y_{hiN} in stratum h are a sample of $iid(\mu_h, \sigma_h^2)$ random variables with bounded $2+\delta$, $\delta > 0$, moments. Let the sample for the Nth population be a simple random stratified sample with $n_{hN} \geq 1$ for all h, where $\{n_{hN}\}$ is a fixed sequence. Let

$$\begin{aligned}
A_N &= \{hi \in U_N : I_{hiN} = 1\}, \\
n_N &= \sum_{h=1}^{H_N} n_{hN} = \sum_{h=1}^{H_N}\sum_{i=1}^{N_{hN}} I_{hiN}, \\
\bar{y}_N &= N^{-1}\sum_{h=1}^{H_N}\sum_{i=1}^{N_{hN}} y_{hiN} = N^{-1}\sum_{hi\in U_N} y_{hiN}, \\
\hat{\theta}_n &= \sum_{h=1}^{H_N} N^{-1}N_{hN}\bar{y}_{hn},
\end{aligned}$$

and

$$\bar{y}_{hn} = n_{hN}^{-1} \sum_{i=1}^{n_{hN}} y_{hiN},$$

where U_N is the set of indices hi for population N and I_{hiN} is the indicator for sample membership. Assume that

$$\lim_{N\to\infty} \sup_{1\le h\le H_N} \frac{n_{hN}^{-2}\left(N_{hN}-n_{hN}\right)^2+1}{\sum_{g=1}^{H_N} N_{gN}\left(N_{gN}-n_{gN}\right) n_{gN}^{-1}\sigma_g^2} = 0. \tag{1.3.10}$$

Then

$$[V\{\hat{\theta}_n - \bar{y}_N\}]^{-1/2}(\hat{\theta}_n - \bar{y}_N) \xrightarrow{\mathcal{L}} N(0,1), \tag{1.3.11}$$

where

$$V\{\hat{\theta}_n - \bar{y}_N\} = \sum_{h=1}^{H_N} N^{-2} N_{hN}(N_{hN}-n_{hN}) n_{hN}^{-1}\sigma_h^2.$$

Furthermore, if the y_{hiN} have bounded fourth moments, if $n_{hN} \ge 2$ for all h, and

$$\sum_{h=1}^{H_N} \lambda_{hN}^2 n_{hN}^{-1} = o\left(\left(\sum_{h=1}^{H_N} \lambda_{hN}\right)^2\right), \tag{1.3.12}$$

then

$$[\hat{V}\{\hat{\theta}_n - \bar{y}_N\}]^{-1/2}(\hat{\theta}_n - \bar{y}_N) \xrightarrow{\mathcal{L}} N(0,1), \tag{1.3.13}$$

where $\lambda_{hN} = N^{-2} N_{hN}\left(N_{hN}-n_{hN}\right) n_{hN}^{-1}$,

$$\hat{V}\{\hat{\theta}_n - \bar{y}_N\} = \sum_{h=1}^{H_N} N^{-2} N_{hN}(N_{hN}-n_{hN}) n_{hN}^{-1} s_h^2,$$

and

$$s_h^2 = (n_{hN}-1)^{-1} \sum_{j=1}^{N_{hN}} (y_{hjN} - \bar{y}_{hn})^2.$$

Proof. For each N, the design is a fixed-size design and by Theorem 1.3.1 the sample in each stratum is a set of *iid* random variables. Therefore, the

stratified estimator is a weighted average of independent random variables and we write

$$\begin{aligned}\hat{\theta}_n - \bar{y}_N &= N^{-1} \sum_{hi \in U_N} \left(N_{hN} n_{hN}^{-1} I_{hiN} - 1\right) y_{hiN} \\ &=: N^{-1} \sum_{hi \in U_N} c_{hN} y_{hiN},\end{aligned}$$

where

$$\begin{aligned}c_{hN} &= N^{-1}\left(N_{hN} - n_{hN}\right) n_{nN}^{-1} \qquad \text{if } hi \in A_N \\ &= -N^{-1} \qquad \text{if } hi \notin A_N.\end{aligned}$$

Because the random variables are identically distributed and the n_{hN} are fixed, we can treat the c_{hN} as fixed.

The Lindeberg criterion is

$$\begin{aligned}&\lim_{N\to\infty} V_N^{-1} \sum_{hi \in U_N} c_{hN}^2 \int_{R_{hN}} (y - \mu_h)^2 dF_{hy}(y) \\ &\leq \lim_{N\to\infty} V_N^{-1} \sum_{hi \in U_N} c_{hN}^2 \int_{R_{0N}} (y - \mu_h)^2 dF_{hy}(y) \\ &\leq \lim_{N\to\infty} V_N^{-1} \sum_{hi \in U_N} c_{hN}^2 \epsilon^\delta B_N^\delta \int_{R_{0N}} |\, y - \mu_h \,|^{2+\delta}\, dF_{hy}(y) \\ &\leq \lim_{N\to\infty} V_N^{-1} \sum_{hi \in U_N} c_{hN}^2 \epsilon^\delta B_N^\delta E\{|\, y_{hi} - \mu_h \,|^{2+\delta}\}, \qquad (1.3.14)\end{aligned}$$

where $V_N = V\{\hat{\theta}_n - \bar{y}_N\}$,

$$R_{hN} = \{y : |\, y - \mu_h \,| \geq \epsilon V_N^{1/2} \;|\, c_{hN} \,|^{-1}\},$$

$$R_{0N} = \{y : |\, y - \mu_h \,| \geq \epsilon B_N^{-1}\},$$

and

$$B_N = V_N^{-1/2} \sup_{1 \leq h \leq H_N} |\, c_{hN} \,|.$$

By assumption (1.3.10) B_N is converging to zero and (1.3.14) converges to zero because the $2 + \delta$ moments are bounded. Thus, the first result is proven.

If the y_{hiN} have fourth moments bounded by, say, M_4, the variance of the estimated variance is

$$\begin{aligned} V\left\{\sum_{h=1}^{H_N}\lambda_{hn}s_h^2\right\} &= \sum_{h=1}^{H_N}\lambda_{hN}^2(n_{hN}-1)^{-2}V\left\{\sum_{i\in A_h}y_{hiN}^2 - n_{hN}\bar{y}_{hn}^2\right\} \\ &\leq \sum_{h=1}^{H_N}\lambda_{hN}^2(n_{hN}-1)^{-2}(n_{hN}M_4+M_5) \\ &= o([V\{\hat{\theta}_n-\bar{y}_N\}]^2) \end{aligned}$$

by assumption (1.3.12), where M_5 is the bound on $V\{n_{hN}\bar{y}_h^2\}$ $-2C\{n_{hN}\bar{y}_h^2, \sum y_{hiN}^2\}$. See Exercise 49. Therefore,

$$[V\{\hat{\theta}_n-\bar{y}_N\}]^{-1}\hat{V}\{\hat{\theta}_n-\bar{y}_N\} \xrightarrow{P} 1$$

and result (1.3.13) is proven. ■

By Theorem 1.3.2, the stratified mean is approximately normally distributed for a large number of small strata or for a small number of large strata.

It is important to note that in Theorem 1.3.2, as in Theorem 1.3.1, results are obtained by averaging over all possible finite populations under the assumption that the design vector $\mathbf{d}$ is independent of $(y_1, y_2, \ldots, y_N)$. The independence assumption is reasonable for stratified samples because selection in each stratum is simple random sampling. Simple random sampling is a special case of stratified sampling and hence the sample mean of a simple random sample is normally distributed in the limit.

Corollary 1.3.2.1. Let $\{\mathcal{F}_N\}$, where $\mathcal{F}_N = (y_{1N}, y_{2N}, \ldots, y_{NN})$, be a sequence of finite populations in which the $y_{iN}, i = 1, 2, \ldots, N$, are realizations of independent (μ, σ^2) random variables with bounded $2+\delta, \delta > 0$, moments. Let A_N be a simple random nonreplacement sample of size n_N selected from the Nth population. Assume that

$$\lim_{N\to\infty} n_N = \infty$$

and

$$\lim_{N\to\infty} N - n_N = \infty.$$

Let $\bar{y}_n$, $\bar{y}_N$, $S_{y,N}^2$, and $s_{y,n}^2$ be as defined in (1.2.35), (1.2.22), (1.2.36), and (1.2.37), respectively. Then

$$\left[N^{-1}(N-n_N)n_N^{-1}S_{y,N}^2\right]^{-1/2}(\bar{y}_n-\bar{y}_N) \xrightarrow{\mathcal{L}} N(0,1) \tag{1.3.15}$$

and

$$\left[N^{-1}(N-n_N)n_N^{-1}s_{y,n}^2\right]^{-1/2}(\bar{y}_n-\bar{y}_N)\xrightarrow{\mathcal{L}}N(0,1). \tag{1.3.16}$$

Proof. Because both n_N and $N-n_N$ increase without bound as N increases, (1.3.10) is satisfied for $H=1$, and result (1.3.15) follows. By the assumption that $E\{|\ y\ |^{2+\delta}\}$ is bounded,

$$\lim_{N\to\infty} s_{y,n}^2=\sigma^2 \quad \text{a.s.}$$

See Hall and Heyde (1980, p. 36). Result (1.3.16) then follows. ■

Result (1.3.13) permits one to use the estimated variance to construct confidence intervals that are appropriate in large samples. That is,

$$\lim_{N\to\infty} P\{\bar{y}_n-t_\alpha[\hat{V}\{\bar{y}_n\}]^{0.5}\le\bar{y}_N\le\bar{y}_n+t_\alpha[\hat{V}\{\bar{y}_n\}]^{0.5}\}=\alpha,$$

where the probability that a standard normal random variable exceeds t_α is α and

$$\hat{V}\{\bar{y}_n\}=(1-f_N)n_N^{-1}s_{y,n}^2.$$

In Theorem 1.3.2, the basis for the limiting result is a sequence of all possible samples from all possible finite populations. One can also obtain limiting normality for Poisson sampling from a fixed sequence of finite populations. The result is due to Hájek (1960).

Consider the Poisson sampling design introduced in Section 1.2.2. For such a design, define the vector random variable

$$\mathbf{x}_i = \mathbf{g}_i I_i, \qquad i=1,\ 2,\ \ldots,\ N, \tag{1.3.17}$$

where I_i is the indicator variable with $I_i=1$ if element i is selected and $I_i=0$ otherwise,

$$\mathbf{g}_i=(1,\ \mathbf{y}_i',\ \alpha_N\pi_i^{-1},\ \alpha_N\pi_i^{-1}\mathbf{y}_i')', \qquad i=1,\ 2,\ \ldots,\ N, \tag{1.3.18}$$

π_i is the probability that element i is included in the sample, $\mathbf{y}_i$ is a column vector associated with element i, $\alpha_N=N^{-1}n_{BN}$, and $n_{BN}=E\{n_N\mid N\}$, where $n_N=\Sigma_{i\in A}I_i$. The ratio α_N is required only for normalization purposes in limit operations, and is required only if $N^{-1}n_N$ or $1-N^{-1}n_N$ goes to zero as N increases. For a fixed $\mathbf{g}_i$, the mean of $\mathbf{x}_i$ is $\mathbf{g}_i\pi_i$ and

$$V\{\mathbf{x}_i\mid\mathbf{g}_i\}=\pi_i(1-\pi_i)\,\mathbf{g}_i\,\mathbf{g}_i'.$$

The Horvitz–Thompson estimator, $N^{-1}\hat{\mathbf{T}}_y$, of the population mean of $\mathbf{y}$ is the vector associated with $\alpha_N \pi_i^{-1}\mathbf{y}_i$ in the estimated mean vector,

$$\hat{\boldsymbol{\mu}}_x = n_{BN}^{-1} \sum_{i=1}^{N} \mathbf{x}_i. \tag{1.3.19}$$

Theorem 1.3.3. Let $\mathbf{y}_1, \mathbf{y}_2, \ldots$, be a sequence of real vectors and let $\pi_1, \pi_2, \ldots$, be a sequence of probabilities, with $0 < \pi_i < 1$. Let a Poisson sample be selected from $\mathcal{F}_N = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)$, and let $\mathbf{g}_i$ be defined by (1.3.18). Assume that

$$\lim_{N\to\infty} n_{BN}^{-1} \sum_{i=1}^{N} \mathbf{g}_i \pi_i = \boldsymbol{\mu}_x, \tag{1.3.20}$$

$$\lim_{N\to\infty} n_{BN}^{-1} \sum_{i=1}^{N} \pi_i(1-\pi_i)\mathbf{g}_i \mathbf{g}_i' = \boldsymbol{\Sigma}_{xx}, \tag{1.3.21}$$

the submatrix of $\boldsymbol{\Sigma}_{xx}$ associated with $(1, \mathbf{y}_i')$ is positive definite, and the submatrix of $\boldsymbol{\Sigma}_{xx}$ associated with $(\alpha_N \pi_i^{-1}, \alpha_N \pi_i^{-1}\mathbf{y}_i')$ is positive definite. Also assume that

$$\lim_{N\to\infty} \sup_{1\le k\le N} \left(\sum_{i=1}^{N} (\boldsymbol{\gamma}' \mathbf{g}_i)^2 \pi_i(1-\pi_i) \right)^{-1} (\boldsymbol{\gamma}'\mathbf{g}_k)^2 = 0 \tag{1.3.22}$$

for every fixed row vector $\boldsymbol{\gamma}'$ such that $\boldsymbol{\gamma}' \boldsymbol{\Sigma}_{xx} \boldsymbol{\gamma} > 0$. Let $\mathbf{x}_i$, $i = 1, 2, \ldots$, be the independent random variables defined by (1.3.17). Then

$$n_{BN}^{1/2}(\hat{\boldsymbol{\mu}}_x - \boldsymbol{\mu}_{xN}) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Sigma}_{xx}), \tag{1.3.23}$$

where

$$\boldsymbol{\mu}_{xN} = n_{BN}^{-1} \sum_{i=1}^{N} \mathbf{g}_i \pi_i$$

and $\hat{\boldsymbol{\mu}}_x$ is defined in (1.3.19).

If, in addition,

$$\lim_{N\to\infty} n_{BN}^{-1} \sum_{i=1}^{N} \pi_i \mid \mathbf{g}_i \mid^4 = M_g \tag{1.3.24}$$

for some finite M_g, then

$$[\hat{V}\{\hat{\mathbf{T}}_y \mid \mathcal{F}_N\}]^{-1/2}(\hat{\mathbf{T}}_y - \mathbf{T}_y) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}), \tag{1.3.25}$$

where $\hat{\mathbf{T}}_y$ is the Horvitz–Thompson estimator, $\mid \mathbf{g}_i \mid = (\mathbf{g}_i'\mathbf{g}_i)^{1/2}$, and

$$\hat{V}\{\hat{\mathbf{T}}_y \mid \mathcal{F}_N\} \;=\; \sum_{i \in A_N} (1 - \pi_i)\pi_i^{-2}\mathbf{y}_i\mathbf{y}_i'.$$

Proof. Let

$$Z_i \;=\; \boldsymbol{\gamma}'\,\mathbf{g}_i(I_i - \pi_i), \tag{1.3.26}$$

where $\boldsymbol{\gamma}$ is an arbitrary real-valued column vector, $\boldsymbol{\gamma} \neq \mathbf{0}$. Then $\{Z_i\}$ is a sequence of independent random variables with zero means and $V\{Z_i\} = \pi_i(1 - \pi_i)\,(\boldsymbol{\gamma}'\,\mathbf{g}_i)^2 =: v_{ii}$. Letting

$$V_N \;=\; \sum_{i=1}^{N} \pi_i\,(1 - \pi_i)\,(\boldsymbol{\gamma}'\mathbf{g}_i)^2 \;=\; \sum_{i=1}^{N} v_{ii}, \tag{1.3.27}$$

the arguments of the proof of Theorem 1.3.2 can be used to show that

$$V_N^{-1/2} \sum_{i=1}^{N} Z_i \xrightarrow{\mathcal{L}} N(0,\ 1). \tag{1.3.28}$$

Note that all moments exist for the random variables I_i. Multivariate normality follows because $\boldsymbol{\gamma}$ is arbitrary. Now

$$\lim_{N \to \infty} n_{BN}^{-1} \sum_{i=1}^{N} (\boldsymbol{\gamma}'\mathbf{g}_i)^2 \pi_i(1 - \pi_i) \;=\; \boldsymbol{\gamma}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}$$

and we have result (1.3.23).

By assumption (1.3.24), the variance of $\hat{V}\{N^{-1}\boldsymbol{\gamma}_y'\hat{\mathbf{T}}_y \mid \mathcal{F}_N\}$ is

$$\begin{aligned} V\left\{N^{-2} \sum_{i \in A_N} (1 - \pi_i)\pi_i^{-2}(\boldsymbol{\gamma}_y'\mathbf{y}_i)^2\right\} \;&=\; N^{-4} \sum_{i=1}^{N} \pi_i(1 - \pi_i)^3\pi_i^{-4}(\boldsymbol{\gamma}_y'\mathbf{y}_i)^4 \\ &=\; O(N^{-3}) \end{aligned}$$

for any fixed vector $\boldsymbol{\gamma}_y$. Because $V\{\hat{\mathbf{T}}_y\}$ is positive definite,

$$[\hat{V}\{\hat{\mathbf{T}}_y \mid \mathcal{F}_N\}]^{-1/2}[V\{\hat{\mathbf{T}}_y \mid \mathcal{F}_N\}]^{1/2} = \mathbf{I} + O_p(N^{-1/2}) \quad (1.3.29)$$

and result (1.3.25) follows. ■

By Theorem 1.3.3, the limiting distribution of the pivotal statistic for the mean of a Poisson sample is $N(0,\ 1)$ for any sequence of finite populations satisfying conditions (1.3.20), (1.3.21), and (1.3.22). Condition (1.3.22) can be replaced with conditions on the moments of y and on the probabilities.

Corollary 1.3.3.1. Let the sequence of populations and vectors satisfy (1.3.20) and (1.3.21) of Theorem 1.3.3. Replace assumption (1.3.22) with

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} |\ \mathbf{y}_i\ |^{2+\delta} < \infty \quad (1.3.30)$$

for some $\delta > 0$, and assume that

$$K_L < \pi_i < K_U \quad (1.3.31)$$

for all i where K_L and K_U are fixed positive numbers and n_{BN} was defined for (1.3.18). Then the limiting normality of (1.3.23) holds.

Proof. The result follows from the fact that (1.3.30) and (1.3.31) are sufficient for the Lindeberg condition. ■

Hájek (1960) showed that the result for Poisson sampling can be extended to simple random nonreplacement sampling .

Theorem 1.3.4. Let the assumptions of Theorem 1.3.3 hold for a sequence of scalars, $\{y_j\}$, with the exception of the assumption of Poisson sampling. Instead, assume that samples of fixed size $n = \pi N$ are selected by simple random nonreplacement sampling. Then

$$\hat{V}_n^{-1/2}(\bar{y}_n - \bar{y}_N) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(0,\ 1), \quad (1.3.32)$$

where $\hat{V}_n = N^{-1}n^{-1}(N-n)s_{y,n}^2$, and $s_{y,n}^2$ is defined for (1.2.37).

Proof. The probability that any set of r elements, $1 \le r \le n_o$, is included in a Poisson sample of size n_o is

$$\frac{\begin{pmatrix} N-r \\ n_o-r \end{pmatrix} \pi^{n_o}(1-\pi)^{N-n_o}}{\begin{pmatrix} N \\ n_o \end{pmatrix} \pi^{n_o}(1-\pi)^{N-n_o}},$$

which is also the probability for the set of r selected as a simple random sample. Hence, the conditional distribution of the Poisson sample given that n_o elements are selected is that of a simple random nonreplacement sample of size n_o.

Let n_B be the expected sample size of a Poisson sample selected with probability π, where n_B is an integer, and let a realized sample of size n_o be given. We create a simple random sample of size n_B starting with the sample of size n_o. If $n_o > n_B$, a simple random sample of $n_o - n_B$ elements is removed from the original sample. If $n_B > n_o$, a simple random sample of $n_B - n_o$ elements is selected from the $N - n_o$ nonsample elements and added to the original n_o elements. If $n_o > n_B$, the n_B elements form a simple random sample from the n_o, and if $n_o < n_B$, n_o is a simple random sample from n_B.

Consider, for $n_o > n_B$, the difference

$$\bar{y}_o - \bar{y}_B = \frac{\sum_{i\in A_o} y_i}{n_o} - \frac{\sum_{i\in A_o} y_i - \sum_{i\in A_k} y_i}{n_B},$$

where $\bar{y}_o$ is the mean of the original Poisson sample, $\bar{y}_B = \bar{y}_{SRS} = \bar{y}_n$ is the mean of the created simple random sample, A_o is the set of indices in the original Poisson sample, and A_k represents the indices of the $k = n_o - n_B$ elements removed from the original Poisson sample. Because the n_B elements are selected from the n_o elements, $E\{\bar{y}_B \mid (n_o, A_o)\} = \bar{y}_o$ and

$$V\{\bar{y}_o - \bar{y}_B \mid (n_o, A_o)\} = (n_B^{-1} - n_o^{-1})s_{yo}^2,$$

where

$$\begin{aligned} s_{yo}^2 &= (n_o - 1)^{-1} \sum_{i\in A_o} (y_i - \bar{y}_o)^2, \\ \bar{y}_B &= n_B^{-1} \sum_{i\in A_B} y_i, \\ \bar{y}_o &= n_o^{-1} \sum_{i\in A_o} y_i, \end{aligned}$$

and A_B is the set of indices for the n_B elements. Furthermore,

$$\begin{aligned} V\{\bar{y}_o - \bar{y}_B \mid n_o\} &= E\{E[(\bar{y}_o - \bar{y}_B)^2 \mid (n_o, A_o)] \mid n_o\} \\ &= (n_B^{-1} - n_o^{-1})S_{y,N}^2, \end{aligned}$$

where $S_{y,N}^2$ is the finite population variance.

If $0 < n_o \leq n_B$, $E\{\bar{y}_o \mid (n_o, A_B)\} = \bar{y}_B$,

$$V\{\bar{y}_o - \bar{y}_B \mid (n_o, A_B)\} = (n_o^{-1} - n_B^{-1})s_{y,B}^2$$

and

$$V\{\bar{y}_o - \bar{y}_B \mid n_o\} = (n_o^{-1} - n_B^{-1})S_{y,N}^2,$$

where

$$s_{y,B}^2 = (n_B - 1)^{-1} \sum_{i \in A_B} (y_i - \bar{y}_B)^2.$$

Then, for $n_o > 0$,

$$V\{\bar{y}_o - \bar{y}_B \mid n_o\} = \mid n_o^{-1} - n_B^{-1} \mid S_{y,N}^2.$$

We note that n_o satisfies

$$E\{(n_B^{-1} n_o - 1)^{2r}\} = O(n_B^{-r})$$

for positive integer r because $N^{-1} n_o$ is the mean of N Poisson random variables. Now $n_o^{-1} n_B$ is bounded by n_B for $n_o > 0$ and by $K_1^{-1} n_B$ for $n_o > K_1$. It follows from Theorems 5.4.4 and 5.4.3 of Fuller (1996) that

$$E\{(n_o^{-1} - n_B^{-1})^2 \mid n_o > 0\} = O(n_B^{-3}).$$

See Exercise 1.34.

To evaluate $E\{(\bar{y}_o - \bar{y}_B)^2\}$, we define $\bar{y}_o - \bar{y}_B = \bar{y}_B$ when $n_o = 0$, and write

$$E\{(\bar{y}_o - \bar{y}_B)^2\} = E\{(\bar{y}_o - \bar{y}_B)^2 \mid n_o > 0\} P\{n_o > 0\} + \bar{y}_B^2 P\{n_o = 0\}.$$

Because $P\{n_o = 0\}$ goes to zero exponentially as $n_B \to \infty$,

$$E\left\{(\bar{y}_B - \bar{y}_o)^2\right\} = O(n_B^{-3/2})$$

and the limiting distribution of $n_B^{1/2}(\bar{y}_B - \bar{y}_N)$ is the same as that of $n_B^{1/2}(\bar{y}_o - \bar{y}_N)$. By Theorem 1.3.3, the limiting distribution of $n_B^{1/2}(\bar{y}_o - \bar{y}_N)$ is normal. By assumption (1.3.24), $s_{y,n}^2 - S_{y,N}^2$ converges to zero in probability, and result (1.3.32) is proven. ■

Theorem 1.3.4 is for simple random samples, but the result extends immediately to a sequence of stratified samples with a fixed number of strata.

Corollary 1.3.4.1. Let $\{\mathcal{F}_N\}$ be a sequence of populations, where the Nth population is composed of H strata with $\mathcal{F}_{h,N} = \{y_{h1}, y_{h2}, \ldots, y_{h,N_{hN}}\}$, $h = 1, 2, \ldots, H$. Assume that $\{y_{hi}\}$, $h = 1, 2, \ldots, H$, are sequences of real numbers satisfying

$$\lim_{N \to \infty} N_{hN}^{-1} \sum_{i \in U_{hN}} \left[y_{hi}, (y_{hi} - \bar{y}_{hN})^2, y_{hi}^4\right] = (M_{1h}, S_h^2, M_{4h}),$$

where the M_{4h} are finite and the S_h^2 are positive. Then

$$[\hat{V}\{(\bar{y}_{st} - \bar{y}_N) \mid \mathcal{F}_N\}]^{-1/2}(\bar{y}_{st} - \bar{y}_N) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(0,\ 1),$$

where $\bar{y}_{st}$ is the stratified mean and $\hat{V}\{(\bar{y}_{st} - \bar{y}_N) \mid \mathcal{F}_N\}$ is the usual stratified estimator defined in (1.2.56).

Proof. Omitted. ■

In proving Theorems 1.3.3 and 1.3.4, we assumed the elements of the finite population to be fixed and obtained results based on the sequence of fixed populations. In Theorem 1.3.2, the sequence of finite populations was created as samples from an infinite population, and the results were for averages over all possible samples from all possible finite populations. It is also useful to have conditional properties for a sequence of finite populations created as samples from an infinite population. Using the strong law of large numbers, it is shown in Theorem 1.3.5 that the central limit theorem holds conditionally, except for a set of probability zero. The sequence $\{\pi_i\}$ in the theorem can be fixed or random.

Theorem 1.3.5. Consider a sequence of populations, $\{\mathcal{F}_N\}$, where the Nth population is the set $(y_1,\ y_2,\ \ldots,\ y_N)$ and $\{y_i\}$ is a sequence of independent $(\mu,\ \sigma_i^2)$ random variables with bounded $4 + \delta$, $\delta > 0$, moments. Let a Poisson sample be selected from the Nth finite population with probabilities π_i, where the $Nn_{BN}^{-1}\pi_i$ are bounded as described in (1.3.31). If the π_i are random, $(\pi_i,\ y_i)$ is independent of $(\pi_j,\ y_j)$ for $i \neq j$. Assume that

$$\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\sigma_i^2 = V_{11} \tag{1.3.33}$$

and

$$\lim_{N\to\infty} n_{BN}N^{-2}\sum_{i=1}^{N}\pi_i^{-1}(1,\ y_i)'(1,\ y_i) = \boldsymbol{\Sigma}_{22} \text{ a.s.}, \tag{1.3.34}$$

where $E\{n_N \mid \mathcal{F}_N\} = n_{BN}$. Assume that V_{11} is positive and that $\boldsymbol{\Sigma}_{22}$ is positive definite. Then

$$[\hat{V}\{\hat{T}_y \mid \mathcal{F}_N\}]^{-1/2}(\hat{T}_y - T_{yN}) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(0,\ 1) \text{ a.s.}, \tag{1.3.35}$$

where T_{yN} is the population total for the Nth population,

$$\hat{T}_y = \sum_{i\in A_N}\pi_i^{-1}y_i,$$

and

$$\hat{V}\{\hat{T}_y \mid \mathcal{F}_N\} = \sum_{i \in A_N} (1 - \pi_i)\pi_i^{-2} y_i^2.$$

Proof. By the $2 + \delta$ moments of the superpopulation,

$$\lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} (y_i,\ y_i^2) = (\mu,\ \mu^2 + V_{11}) \quad \text{a.s.}$$

and

$$\lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \mid y_i \mid^{2+0.5\delta} < \infty \quad \text{a.s.}$$

Therefore, conditions (1.3.20), (1.3.21), and (1.3.22) are satisfied almost surely and

$$[V\{\hat{T}_y \mid \mathcal{F}_N\}]^{-1/2} (\hat{T}_y - T_{yN}) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(0,\ 1) \quad \text{a.s.}, \tag{1.3.36}$$

where

$$V\{\hat{T}_y \mid \mathcal{F}_N\} = \sum_{i=1}^{N} (1 - \pi_i)\pi_i^{-1} y_i^2,$$

by Corollary 1.3.3.1.

Now $n_{BN} N^{-1} (1 - \pi_i)\pi_i^{-1}$ is bounded by, say, K_c, and

$$V\left\{ \sum_{i \in A} \left[n_{BN} N^{-1} (1 - \pi_i)\pi_i^{-1} \right] \pi_i^{-1} y_i^2 \mid \mathcal{F}_N \right\} \leq \sum_{i=1}^{N} K_c^2 (1 - \pi_i)\pi_i^{-1} y_i^4.$$

Therefore, by the $4 + \delta$ moments of y_i,

$$\lim_{N \to \infty} n_{BN} N^{-1} \sum_{i=1}^{N} K_c^2 (1 - \pi_i)\pi_i^{-1} y_i^4$$

is a well-defined finite number, almost surely. It follows that

$$n_{BN}N^{-2}[\hat{V}\{\hat{T}_y \mid \mathcal{F}_N\} - V\{\hat{T}_y \mid \mathcal{F}_N\}] \mid \mathcal{F}_N = O_p(n_{BN}^{-1/2}) \text{ a.s.} \tag{1.3.37}$$

and result (1.3.35) follows. ■

Theorem 1.3.5 is for Poisson sampling but the result holds for a sequence of stratified samples with a fixed number of strata, by Corollary 1.3.4.1.

Corollary 1.3.5.1. Let $\{\mathcal{F}_N\}$ be a sequence of populations, where the Nth population is composed of H strata with $\mathcal{F}_{h,N} = \{y_{h1}, y_{h2}, \ldots, y_{h,N_{hN}}\}$, $h = 1, 2, \ldots, H$. Assume that the $\{y_{hi}\}$, $h = 1, 2, \ldots, H$, are sequences of independent (μ_h, σ_h^2) random variables with bounded $4 + \delta$, $\delta > 0$ moments. Let a sequence of stratified samples be selected, where $n_{h,N} \geq n_{h,N-1}$, $\lim_{N\to\infty} n_{h,N} = \infty$,

$$\lim_{N\to\infty} N_{h,N}^{-1} n_{h,N} = f_{h,\infty},$$

and

$$\lim_{N\to\infty} N^{-1} N_{h,N} = W_h$$

for $h = 1, 2, \ldots, H$. Assume that $0 \leq f_{h,\infty} < 1$ and $W_h > 0$ for $h = 1, 2, \ldots, H$. Then

$$[\hat{V}\{\hat{T}_y \mid \mathcal{F}_N\}]^{-1/2}(\hat{T}_y - T_{yN}) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(0, 1) \text{ a.s.},$$

where

$$\hat{V}\{\hat{T}_y \mid \mathcal{F}_N\} = \sum_{h=1}^{H} N_h^2(n_h^{-1} - N_h^{-1})s_h^2.$$

Proof. The conditions of Corollary 1.3.3.1 hold almost surely and the result follows by Corollary 1.3.4.1. ■

To extend the results of Theorem 1.3.5 to estimation of parameters of the superpopulation, we require the following theorem, adapted from Schenker and Welsh (1988).

Theorem 1.3.6. Let $\{\mathcal{F}_N\}$ be a sequence of finite populations, let θ_N be a function of the elements of $\mathcal{F}_N$, and let a sequence of samples be selected from $\{\mathcal{F}_N\}$ by a design such that

$$\theta_N - \theta_N^\circ \xrightarrow{\mathcal{L}} N(0, V_{11}), \tag{1.3.38}$$

and

$$(\hat{\theta} - \theta_N) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(0,\ V_{22}) \quad \text{a.s.}, \tag{1.3.39}$$

for a fixed sequence $\{\theta_N^\circ\}$ and an estimator, $\hat{\theta}$, where $V_{11} + V_{22} > 0$. Then

$$(V_{11} + V_{22})^{-1/2}(\hat{\theta} - \theta_N^\circ) \xrightarrow{\mathcal{L}} N(0,\ 1). \tag{1.3.40}$$

Proof. Let $\Phi_{Vj}(\cdot)$ denote the normal cumulative distribution function with mean zero and variance V_{jj}, $j = 1,\ 2$, and let $\Phi_{V3}(\cdot) = \Phi_{V1}(\cdot) * \Phi_{V2}(\cdot)$ denote the normal cumulative distribution function with mean zero and variance $V_{11} + V_{22} = V_{33}$, where $\Phi_{V1}(\cdot) * \Phi_{V2}(\cdot)$ is the convolution of $\Phi_{V1}(\cdot)$ and $\Phi_{V2}(\cdot)$. Consider

$$\mid P\{(\hat{\theta} - \theta_N^\circ) \le t\} - \Phi_{V3}(t) \mid \ = \ \mid E[P\{(\hat{\theta} - \theta_N^\circ) \le t\} \mid \mathcal{F}_N] - \Phi_{V3}(t) \mid .$$

Letting $s = t - (\theta_N - \theta_N^\circ)$ yields

$$\begin{aligned}
&\mid P\{(\hat{\theta} - \theta_N^\circ) \le t\} - \Phi_{V3}(t) \mid \\
&\qquad \le \ \mid E[P\{(\hat{\theta} - \theta_N) \le s \mid \mathcal{F}_N\}] - E\{\Phi_{V2}(s) \mid \mathcal{F}_N\} \mid \\
&\qquad\quad + \mid E\{\Phi_{V2}(s) \mid \mathcal{F}_N\} - \Phi_{V3}(t) \mid \\
&\qquad \le \ E[\ \mid P\{(\hat{\theta} - \theta_N) \le s \mid \mathcal{F}_N\} - \Phi_{V2}(s) \mid ,\ \mid \mathcal{F}_N] \\
&\qquad\quad + \mid E\{\Phi_{V2}(s) \mid \mathcal{F}_N\} - \Phi_{V3}(t) \mid .
\end{aligned} \tag{1.3.41}$$

Because $\mid P\{(\hat{\theta} - \theta_N) \le s \mid \mathcal{F}_N\} - \Phi_{V2}(s) \mid$ is bounded for all $s \in \mathcal{R}$,

$$\begin{aligned}
&\lim_{N\to\infty} \mid P\{(\hat{\theta} - \theta_N) \le s \mid \mathcal{F}_N\} - \Phi_{V2}(s) \mid \mathcal{F}_N \mid \\
&\qquad \le E\{\lim_{N\to\infty} (\sup_{s\in\mathcal{R}} \mid P\{(\hat{\theta} - \theta_N) \le s \mid \mathcal{F}_N\} - \Phi_{V2}(s) \mid ,\ \mid \mathcal{F}_N)\}
\end{aligned} \tag{1.3.42}$$

by the dominated convergence theorem. By Lemma 3.2 of R. R. Rao (1962), assumption (1.3.39) implies that

$$\lim_{N\to\infty} \{\sup_{s\in\mathcal{R}} \mid P\{(\hat{\theta} - \theta_N) \le s \mid \mathcal{F}_N\} - \Phi_{V2}(s) \mid ,\ \mid \mathcal{F}_N\} \ = \ 0 \quad \text{a.s.} \tag{1.3.43}$$

and the expectation in (1.3.42) is zero.

Now

$$\lim_{N\to\infty} E\{\Phi_{V2}(s) \mid \mathcal{F}_N\} = E\left\{\lim_{N\to\infty} \Phi_{V2}[t-(\theta_N-\theta_N^\circ)] \ , \mid \mathcal{F}_N\right\}$$

$$= \Phi_{V2}(t) * \Phi_{V1}(t) = \Phi_{V3}(t) \qquad (1.3.44)$$

by (1.3.38) and the dominated convergence theorem. It follows from (1.3.43) and (1.3.44) that (1.3.41) converges to zero as $N \to \infty$. ■

The V_{11} of (1.3.38) or the V_{22} of (1.3.39) can be zero, but the sum $V_{11}+V_{22}$ is never zero. For example, let $\bar{y}_n$ be the mean of a simple random sample from a finite population that is a set of $iid(\mu, \sigma^2)$ random variables, let $\hat{\theta}_n - \theta_N = N^{1/2}(\bar{y}_n - \bar{y}_N)$, and let $\theta_N - \theta_N^\circ = N^{1/2}(\bar{y}_N - \mu)$. Then if $N - n \to 0$ as N increases, $V_{22} = 0$. Conversely, if $N^{-1}n \to 0$, the limiting variance of $n^{1/2}(\bar{y}_n - \bar{y}_N) = \sigma^2$ and the limiting variance of $n^{1/2}(\bar{y}_N - \mu) = 0$. If $\lim N^{-1}n = f$, for $0 < f < 1$, both V_{11} and V_{22} are positive. These theoretical results have a commonsense interpretation. If the sample is a very small fraction of the finite population, the fact that there is the intermediate step of generating a large finite population is of little importance. Conversely, if we have a very large sampling rate, say a census, we still have variability in the estimator of the superpopulation parameter. See Deming and Stephan (1941) on the use of a census in this context.

Using Theorems 1.3.6 and 1.3.5, one can prove that the limiting distribution of the standardized $\bar{y}_{HT} - \mu$ is normal.

Corollary 1.3.6.1. Let the assumptions of Theorem 1.3.5 hold and assume that

$$\lim_{N\to\infty} N^{-1}n_{BN} = f_\infty,$$

where $0 \le f_\infty \le 1$. Then

$$\left[\hat{V}\{\bar{y}_{HT} - \mu\}\right]^{-1/2} (\bar{y}_{HT} - \mu) \xrightarrow{\mathcal{L}} N(0, 1), \qquad (1.3.45)$$

where $\bar{y}_{HT}$ is as defined in (1.2.24),

$$\hat{V}\{\bar{y}_{HT} - \mu\} = \hat{V}\{\bar{y}_{HT} \mid \mathcal{F}_N\} + N^{-1}\hat{S}_y^2,$$

$$\hat{S}_y^2 = N^{-1}\sum_{i\in A_N} \pi_i^{-1}(y_i - \bar{y}_{HT})^2,$$

$\hat{V}\{\bar{y}_{HT} \mid \mathcal{F}_N\} = N^{-2}\hat{V}\{\hat{T}_y \mid \mathcal{F}_N\}$, and $\hat{V}\{\hat{T}_y \mid \mathcal{F}_N\}$ is as defined in (1.3.25).

Proof. By the moment properties of the superpopulation,

$$\theta_N = N^{1/2}(\bar{y}_N - \mu) \xrightarrow{\mathcal{L}} N(0,\ V_{11})$$

or, equivalently, for $f_\infty > 0$,

$$n_{BN}^{1/2}(\bar{y}_N - \mu) \xrightarrow{\mathcal{L}} N(0,\ f_\infty V_{11}).$$

By (1.3.36) of the proof of Theorem 1.3.5,

$$n_{BN}^{1/2}(\bar{y}_{HT} - \bar{y}_N) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(0,\ V_{22}) \quad \text{a.s.},$$

where

$$\lim_{N\to\infty} n_{BN} N^{-2} \sum_{i=1}^{N} (1 - \pi_i)\pi_i^{-1} y_i^2 \quad = \quad V_{22} \quad \text{a.s.}$$

Therefore, by Theorem 1.3.6,

$$n_B^{1/2}(\bar{y}_{HT} - \mu) \xrightarrow{\mathcal{L}} N(0,\ V_{22} + f_\infty V_{11}). \tag{1.3.46}$$

The estimated variance satisfies

$$\hat{V}\{\bar{y}_{HT} \mid \mathcal{F}_N\} - V\{\bar{y}_{HT} \mid \mathcal{F}_N\} \quad = \quad O_p(n_{BN}^{-1.5})$$

by (1.3.37). Also,

$$V\left\{N^{-1} \sum_{i\in A_N} \pi_i^{-1} y_i^2\right\} \quad = \quad O(n_{BN}^{-1})$$

and $\bar{y}_{HT}^2 = \mu^2 + O_p(n_{BN}^{-1/2})$ by the fourth moments of y_i. Therefore,

$$\hat{V}\{\bar{y}_{HT} - \mu\} \quad = \quad V\{\bar{y}_{HT} - \mu\} + O_p(n_{BN}^{-1.5}) \tag{1.3.47}$$

and result (1.3.45) follows from (1.3.46) and (1.3.47). ■

1.3.3 Functions of means

Theorems 1.3.2, 1.3.3, and 1.3.4 give the limiting distributions for means, but functions of means also occur frequently in the analysis of survey samples. It is a standard result that the limiting distribution of a continuous differentiable function of a sample mean is normal, provided that the standardized mean converges in distribution to a normal distribution.

Theorem 1.3.7. Let $\bar{\mathbf{x}}_n$ be a vector random variable with $E\{\bar{\mathbf{x}}_n\} = \boldsymbol{\mu}_x$ such that

$$n^{1/2}(\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x) \xrightarrow{\mathcal{L}} N(\mathbf{0},\ \boldsymbol{\Sigma}_{xx})$$

as $n \to \infty$. Let $g(\bar{\mathbf{x}}_n)$ be a function of $\bar{\mathbf{x}}$ that is continuous at $\boldsymbol{\mu}_x$ with a continuous derivative at $\boldsymbol{\mu}_x$. Then

$$n^{1/2}\left[g(\bar{\mathbf{x}}_n) - g(\boldsymbol{\mu}_x)\right] \xrightarrow{\mathcal{L}} N\left[\mathbf{0},\ \mathbf{h}_x(\boldsymbol{\mu}_x)\boldsymbol{\Sigma}_{xx}\mathbf{h}_x'(\boldsymbol{\mu}_x)\right]$$

as $n \to \infty$, where $\mathbf{h}_x(\boldsymbol{\mu}_x)$ is the row vector of derivatives of $g(\bar{\mathbf{x}})$ with respect to $\bar{\mathbf{x}}$ evaluated at $\bar{\mathbf{x}} = \boldsymbol{\mu}_x$.

Proof. By a Taylor expansion

$$g(\bar{\mathbf{x}}_n) \quad = \quad g(\boldsymbol{\mu}_x) + (\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)\mathbf{h}_x'(\boldsymbol{\mu}_x^*),$$

where $\boldsymbol{\mu}_x^*$ is on the line segment joining $\bar{\mathbf{x}}_n$ and $\boldsymbol{\mu}_x$. Now $\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x = O_p(n^{-1/2})$ and $\mathbf{h}_x(\mathbf{x})$ is continuous at $\mathbf{x} = \boldsymbol{\mu}_x$. Therefore, given $\delta > 0$ and $\varepsilon_x > 0$, there is some n_0 and a closed set B containing $\boldsymbol{\mu}_x$ as an interior point such that $\mathbf{h}_x(\mathbf{x})$ is uniformly continuous on B and $P\{\bar{\mathbf{x}}_n \in B$ and $\mid \bar{\mathbf{x}}_n - \boldsymbol{\mu}_x \mid < \varepsilon_x\} > 1 - \delta$ for $n > n_0$. Therefore, given $\delta > 0$ and an $\varepsilon_h > 0$, there is an $\varepsilon_x > 0$ and an n_0 such that

$$P\left\{\mid g(\bar{\mathbf{x}}_n) - g(\boldsymbol{\mu}_x) - (\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)\mathbf{h}_x'(\boldsymbol{\mu}_x) \mid > \varepsilon_h\right\} < \delta$$

for all $n > n_0$. The limiting normality follows because $n^{1/2}(\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)$ converges to a normal vector and $\mathbf{h}_x'(\boldsymbol{\mu}_x)$ is a fixed vector. ∎

Ratios of random variables, particularly ratios of two sample means, play a central role in survey sampling. Because of their importance, we give a separate theorem for the large-sample properties of ratios.

Theorem 1.3.8. Let a sequence of finite populations be created as samples from a superpopulation with finite fourth moments. Let $\mathbf{x}_j = (x_{1j},\ x_{2j})$ and assume that $\mu_{x1} \neq 0$, where $\boldsymbol{\mu}_x = (\mu_{x1},\ \mu_{x2})$ is the superpopulation mean. Assume that the sequence of designs is such that

$$n_{BN}^{1/2}N^{-1}(\hat{\mathbf{T}}_x - N\boldsymbol{\mu}_x) \xrightarrow{\mathcal{L}} N(\mathbf{0},\ \boldsymbol{\Sigma}_{xx}) \tag{1.3.48}$$

and

$$n_{BN}^{1/2}N^{-1}(\hat{\mathbf{T}}_x - \mathbf{T}_{xN}) \xrightarrow{\mathcal{L}} N(\mathbf{0},\ \mathbf{M}_{xx}), \tag{1.3.49}$$

where $\boldsymbol{\Sigma}_{xx}$ and $\mathbf{M}_{xx}$ are positive definite, $n_{BN} = E\{n_N \mid \mathcal{F}_N\}$,

$$\hat{\mathbf{T}}_x \quad = \quad \sum_{i \in A_N} \pi_i^{-1}\mathbf{x}_i,$$

and $\mathbf{T}_{xN} = N\bar{\mathbf{x}}_N$. Then

$$n_{BN}^{1/2}(\hat{R} - R_S) \xrightarrow{\mathcal{L}} N(\mathbf{0},\ \mathbf{h}_S\boldsymbol{\Sigma}_{xx}\mathbf{h}_S') \tag{1.3.50}$$

and

$$n_{BN}^{1/2}(\hat{R} - R_N) \xrightarrow{\mathcal{L}} N(\mathbf{0},\ \mathbf{h}_N\mathbf{M}_{xx}\mathbf{h}_N'), \tag{1.3.51}$$

where $\hat{R} = \hat{T}_{x1}^{-1}\hat{T}_{x2}$, $R_S = \mu_{x1}^{-1}\mu_{x2}$, $R_N = \bar{x}_{1N}^{-1}\bar{x}_{2N}$,

$$\mathbf{h}_S = \left(\frac{\partial R}{\partial x_1}, \frac{\partial R}{\partial x_2}\right) \mid \mathbf{x} = \boldsymbol{\mu}_x = \mu_{x1}^{-1}(1,\ -R_s)$$

and

$$\mathbf{h}_N = \left(\frac{\partial R}{\partial x_1}, \frac{\partial R}{\partial x_2}\right) \mid \mathbf{x} = \bar{\mathbf{x}}_N = \bar{x}_{1N}^{-1}(1,\ -R_N).$$

Let the designs be such that

$$[V\{\hat{\mathbf{T}}_x \mid \mathcal{F}_N\}]^{-1}\hat{V}_{HT}\{\hat{\mathbf{T}}_x\} - \mathbf{I} = O_p\left(n_{BN}^{-1/2}\right) \tag{1.3.52}$$

for any x-variable with finite fourth moments, where $\hat{V}_{HT}\{\hat{\mathbf{T}}_x\}$ is the Horvitz–Thompson variance estimator. Then

$$[\hat{V}_{HT}\{\hat{T}(\hat{d})\}]^{-1/2}(\hat{R} - R_N) \xrightarrow{\mathcal{L}} N(0,\ 1), \tag{1.3.53}$$

where $\hat{V}_{HT}\{\hat{T}(\hat{d})\}$ is the Horvitz–Thompson variance estimator calculated for

$$\hat{T}(\hat{d}) = \sum_{i \in A_N} \pi_i^{-1}\hat{d}_i,$$

and $\hat{d}_i = \hat{T}_{x1}^{-1}(x_{2i} - \hat{R}x_{1i})$.

Proof. Results (1.3.50) and (1.3.51) follow from (1.3.48), (1.3.49), and Theorem 1.3.7. The Taylor expansion for the estimator of the finite population ratio is

$$\begin{aligned}\hat{R} &= R_N + T_{x1,N}^{-1}(\hat{T}_{x2} - T_{x2,N}) - T_{x1,N}^{-2}T_{x2,N}(\hat{T}_{x1} - T_{x1,N}) + O_p(n_{BN}^{-1}) \\ &= R_N + T_{x1,N}^{-1}\left(\sum_{i \in A_N} \pi_i^{-1}e_i\right) + O_p(n_{BN}^{-1}),\end{aligned} \tag{1.3.54}$$

where $e_i = x_{2i} - R_N x_{1i}$. The remainder is $O_p(n_{BN}^{-1})$ because the second derivatives are continuous at $(T_{x1,N},\ T_{x2,N})$. Now

$$\hat{V}_{HT}\{\hat{T}(\hat{d})\} = \sum_{i,j \in A_N}\sum \pi_{ij}^{-1}(\pi_{ij} - \pi_i\pi_j)\pi_i^{-1}\hat{d}_i\pi_j^{-1}\hat{d}_j,$$

where

$$\hat{d}_i = d_i - \hat{T}_{x1}^{-1} T_{x1,N}^{-1} (\hat{T}_{x1} - T_{x1,N}) e_i + \hat{T}_{x1}^{-1} (\hat{R} - R_N) x_{1i},$$

and $d_i = T_{x1,N}^{-1} e_i$. Because $\boldsymbol{\Sigma}_{xx}$ is positive definite, $V\{\hat{T}(d)\}$ is the same order as $V\{\hat{T}_{x1}\}$. It follows from (1.3.52) that

$$[V\{\hat{T}(d)\}]^{-1} \hat{V}_{HT}\{\hat{T}(\hat{d})\} = 1 + O_p(n_{BN}^{-1/2}), \tag{1.3.55}$$

because, for example,

$$[V\{\hat{T}(d)\}]^{-1} \hat{T}_{x1}^{-2} (\hat{R} - R_N)^2 \sum_{i,j \in A_N} \sum g_{ij} x_{1i} x_{2j} = O_p(n_B^{-1}),$$

$$[V\{\hat{T}(d)\}]^{-1} \hat{T}_{x1}^{-1} T_{x1,N}^{-1} (\hat{T}_{x1} - T_{x1,N}) \sum_{i,j \in A_N} \sum g_{ij} d_i e_j = O_p(n_B^{-1/2}),$$

where $g_{ij} = \pi_{ij}^{-1} (\pi_{ij} - \pi_i \pi_j) \pi_i^{-1} \pi_j^{-1}$. Result (1.3.53) follows from (1.3.55) and (1.3.54). ■

Theorem 1.3.8 is for unconditional properties derived under the probability structure defined by sampling from a finite population that is a sample from a superpopulation. It is possible to prove a theorem analogous to Theorem 1.3.5 in which the result holds almost surely for the sequence of finite populations.

Observe that the error in the ratio estimator is approximated by a design linear estimator in e_i in expression (1.3.54). This approximation is what leads to the limiting normality in (1.3.53). Although the estimator is not exactly normally distributed and $\hat{V}_{HT}\{\hat{T}(\hat{d})\}$ is not exactly a multiple of a chi-square random variable, the limiting distribution of the "t-statistic" is $N(0, 1)$. This type of result will be used repeatedly for nonlinear functions of Horvitz–Thompson estimators.

Expression (1.3.53) provides an efficient way to compute the estimated variance of the ratio using

$$\hat{d}_i = \hat{T}_{x1}^{-1} (x_{2i} - \hat{R} x_{1i}).$$

The $\hat{d}_i$ is sometimes called the *estimated Taylor deviate.* In the notation of Theorem 1.3.7, the Taylor deviate is

$$d_i = \mathbf{h}_x(\boldsymbol{\mu}_x)(\mathbf{x}_i - \bar{\mathbf{x}}_\pi)'$$

and the estimated variance of $g(\hat{\mathbf{T}}_x)$ is the estimated variance of $\bar{d}_{HT}$ calculated with $\hat{d}_i$.

We can use Theorem 1.3.8 to define an estimator for the population mean of y by letting $(x_{1i},\ x_{2i}) = (1,\ y_i)$. Then we obtain

$$\bar{y}_\pi \quad = \quad \left(\sum_{i\in A}\pi_i^{-1}\right)^{-1}\sum_{i\in A}\pi_i^{-1}y_i \tag{1.3.56}$$

as an estimator of $\bar{y}_N$. We gave the unbiased estimator

$$\bar{y}_{HT} \quad = \quad N^{-1}\hat{T}_y \quad = \quad N^{-1}\sum_{j\in A}\pi_j^{-1}y_j \tag{1.3.57}$$

in (1.2.24). The estimators (1.3.56) and (1.3.57) are identical for many designs, including stratified sampling, but can differ considerably for designs such as Poisson sampling with unequal probabilities. In general, and under mild regularity conditions, $N^{-1}\hat{T}_y$ is design unbiased and design consistent, whereas $\bar{y}_\pi$ is only design consistent. However, $\bar{y}_\pi$ is location and scale invariant, whereas $N^{-1}\hat{T}_y$ is only scale invariant. See (1.2.29). The estimator (1.3.56) is sometimes called the *Hájek estimator*. See Hájek (1971).

The estimators (1.3.56) and (1.3.57) can be compared under models for the population. One superpopulation model is

$$\begin{aligned} y_i \quad &= \quad \beta_0 + \beta_1 x_i + e_i, \\ e_i \quad &\sim \quad ind(0,\ x_i^\alpha\sigma^2), \end{aligned} \tag{1.3.58}$$

where α is positive, the x_i are positive, e_j is independent of x_i for all i and j, and $\sim$ *ind* denotes distributed independently. Let $(x_1,\ x_2,\ \ldots,\ x_N)$ be a finite population of positive x values, let the finite population of y_i values be generated by model (1.3.58), and let a sample be selected with probabilities $\pi_i = n(\sum_{j\in U}x_j)^{-1}x_i$.

Then the conditional expected value of the finite population mean is

$$E\left\{\bar{y}_N \mid \bar{x}_N\right\} \quad = \quad \beta_0 + \beta_1\bar{x}_N. \tag{1.3.59}$$

For fixed-size designs, the conditional expectations of the estimators are

$$E\{N^{-1}\hat{T}_y \mid \mathbf{x}_A\} \quad = \quad \beta_0 N^{-1}\hat{N}_{HT} + \beta_1\bar{x}_N, \tag{1.3.60}$$

$$E\{\bar{y}_\pi \mid \mathbf{x}_A\} \quad = \quad \beta_0 + \hat{N}_{HT}^{-1}N\beta_1\bar{x}_N, \tag{1.3.61}$$

where $\hat{N}_{HT} = \sum_{i\in A}\pi_i^{-1}$ and $\mathbf{x}_A$ is the set of x values in the sample. If $\beta_0 = 0$, $N^{-1}\hat{T}_y$ is conditionally unbiased, and if $\beta_1 = 0$, $\bar{y}_\pi$ is conditionally unbiased, conditional on $\mathbf{x}_A$.

The conditional variances are

$$V\{N^{-1}\hat{T}_y \mid \mathbf{x}_A\} = N^{-2}\sum_{i\in A}\pi_i^{-2+\alpha}\sigma^2 \tag{1.3.62}$$

and

$$V\{\bar{y}_\pi \mid \mathbf{x}_A\} = \left(\sum_{i\in A}\pi_i^{-1}\right)^{-2}\sum_{i\in A}\pi_i^{-2+\alpha}\sigma^2. \tag{1.3.63}$$

Thus, the conditional variance of $\bar{y}_\pi$ can be larger or smaller than that of $N^{-1}\hat{T}_y$.

The design variance of $\bar{y}_{HT}$ is

$$V\{\bar{y}_{HT} \mid \mathcal{F}\} = V\left\{N^{-1}\sum_{i\in A}\pi_i^{-1}y_i \mid \mathcal{F}\right\} \tag{1.3.64}$$

and the design variance of the approximate distribution of $\bar{y}_\pi$ is

$$V\{\bar{y}_\pi \mid \mathcal{F}\} = V\left\{N^{-1}\sum_{i\in A}\pi_i^{-1}(y_i - \bar{y}_N) \mid \mathcal{F}\right\}. \tag{1.3.65}$$

Thus, as suggested by (1.3.64) and (1.3.65), $\bar{y}_{HT}$ will have smaller design variance than $\bar{y}_\pi$ if the ratio of y_i to π_i is nearly constant and $\bar{y}_\pi$ will have smaller design variance than $\bar{y}_{HT}$ if $y_i - \bar{y}_N$ is nearly a constant multiple of π_i. Also see Exercise 6.

Because $\bar{y}_\pi$ is location invariant, we generally begin estimation for more complex situations with $\bar{y}_\pi$. A regression estimator that is conditionally model unbiased, conditional on $\mathbf{x}_A$, is discussed in Chapter 2.

In the analysis of survey samples, subpopulations are often called *domains of study* or, simply, *domains*. Thus, in reporting unemployment rates, the rate might be reported for a domain composed of females aged 35 to 44. To study the properties of the estimated mean for a domain, let

$$\begin{aligned} y_{Di} &= y_i && \text{if element } i \text{ is in domain } D \\ &= 0 && \text{otherwise,} \\ z_{Di} &= 1 && \text{if element } i \text{ is in domain } D \\ &= 0 && \text{otherwise.} \end{aligned}$$

Then the estimator of the domain mean is the ratio estimator

$$\hat{\theta}_D = \bar{z}_{D\pi}^{-1}\bar{y}_{D\pi} = \hat{T}_{zD}^{-1}\hat{T}_{yD}, \tag{1.3.66}$$

where

$$(\hat{T}_{zD},\ \hat{T}_{zD}) \quad = \quad \sum_{i \in A} \pi_i^{-1}(z_{Di},\ y_{Di}).$$

The Horvitz–Thompson variance estimator of Theorem 1.3.8 is

$$\hat{V}_{LS}\{\hat{\theta}_D\} \quad = \quad \bar{z}_{D\pi}^{-2} \sum_{i,j \in A} \sum \pi_{ij}^{-1}(\pi_{ij} - \pi_i \pi_j)\pi_i^{-1}\hat{e}_i \pi_j^{-1}\hat{e}_j, \qquad (1.3.67)$$

where $\hat{e}_i = y_{Di} - \hat{\theta}_D z_{Di}$. Observe that $\hat{e}_i$ is zero if element i is not in the domain.

The properties of the estimated domain mean illustrate the care required in the use of large-sample results. Assume that we have a simple random sample from a finite population that is, in turn, a random sample from a normal distribution. Assume that the finite population correction can be ignored. Then the domain mean is the simple mean of the elements in the domain,

$$\hat{\theta}_D \quad = \quad \bar{y}_D \quad = \quad n_D^{-1} \sum_{i \in A_D} y_i, \qquad (1.3.68)$$

where $n_D = \Sigma_{i \in A} z_{Di}$ is the number of elements in domain D and A_D is the set of indices of elements in domain D. The variance estimator (1.3.67) becomes

$$\hat{V}_{LS}\{\hat{\theta}_D\} \quad = \quad n_D^{-2} n^2 \left[n(n-1)\right]^{-1} \sum_{i \in A_D} (y_i - \bar{y}_D)^2. \qquad (1.3.69)$$

Because the original sample is a simple random sample, the n_D elements selected from domain D are a simple random sample from that domain. Therefore,

$$t_{srs,D} \quad = \quad [\hat{V}_{srs}\{\bar{y}_D\}]^{-1/2}(\bar{y}_D - \mu_D) \qquad (1.3.70)$$

is, conditional on $n_D, n_D > 1$, distributed as Student's t with $n_D - 1$ degrees of freedom, where

$$\hat{V}_{srs}\{\bar{y}_D\} \quad = \quad [n_D(n_D - 1)]^{-1} \sum_{i \in A_D} (y_i - \bar{y}_D)^2 \qquad (1.3.71)$$

and μ_D is the population domain mean. If we use (1.3.69) to construct the estimated variance, we have

$$\hat{V}_{LS}\{\hat{\theta}_D\} \quad = \quad n_D^{-1}(n_D - 1)\hat{V}_{srs}\{\hat{\theta}_D\}. \qquad (1.3.72)$$

Thus, the estimator based on the large-sample approximation underestimates the variance and

$$[\hat{V}_{LS}\{\hat{\theta}_D\}]^{-1/2}(\hat{\theta}_D - \mu_D) \sim \left[n_D(n_D-1)^{-1}\right]^{1/2} t_{srs,D}, \tag{1.3.73}$$

where t_{n_D-1} is distributed as Student's t with $n_D - 1$ degrees of freedom. For small n_D, $N(0, 1)$ will be a very poor approximation for the distribution of (1.3.73).

The assumptions of Theorem 1.3.8 require the distributions of $\bar{z}_{D\pi}$ and $\bar{y}_{D\pi}$ to have small variances. This condition does not hold for the components of the domain mean if n_D is small, no matter how large the original sample.

1.3.4 Approximations for complex estimators

An estimator is often defined as the solution to a system of equations, where the solution may be implicit. In Theorem 1.3.9 we show that Taylor methods can be used to obtain an approximation to the distribution of such an estimator. Results are given for $\hat{\boldsymbol{\theta}}$ as an estimator of the finite population parameter and for $\hat{\boldsymbol{\theta}}$ as an estimator of the parameter of the superpopulation that generated the finite population.

The theorem contains a number of technical assumptions. They can be summarized as assumptions of existence of moments for the superpopulation, assumptions pertaining to the design, and assumptions about the functions defining the estimator. The design must be such that a central limit theorem holds for the Horvitz–Thompson estimator, and the function must be continuous with at least a continuous second derivative with respect to the parameter.

It is assumed that the estimator is consistent. See (1.3.80) and (1.3.81). The consistency assumption is required because some functions have more than one root. If the function $\mathbf{g}(\mathbf{x}, \boldsymbol{\theta})$ is the vector of derivatives of an objective function, it may be possible to use the properties of the objective function to prove (1.3.80) and (1.3.81). See, for example, Gallant (1987). The usual w_i of the theorem is π_i^{-1}, but alternative weights, some considered in Chapter 6, are possible. Equation (1.3.75), which defines the estimator, is sometimes called an *estimating equation*. See Godambe (1991).

Theorem 1.3.9. Let $\mathcal{F}_N = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ be the Nth finite population in the sequence $\{\mathcal{F}_N\}$, where $\{\mathbf{x}_i\}$ is a sequence of *iid* random variables with finite fourth moments. Assume that the sequence of designs is such that for any $\mathbf{x}_j$ with positive variance,

$$V\{n_{BN}^{1/2}N^{-1}(\hat{\mathbf{T}}_x - \mathbf{T}_{x,N}) \mid \mathcal{F}_N\} = \mathbf{V}_{T,xx,N}, \tag{1.3.74}$$

where $\mathbf{V}_{T,xx,N}$ is positive semidefinite almost surely, $\mathbf{T}_{x,N}$ is the population total of $\mathbf{x}$, $\hat{\mathbf{T}}_{\mathbf{x}}$ is the Horvitz–Thompson estimator of the total, and $n_{BN} = E\{n_N \mid \mathcal{F}_N\}$. Let an estimator $\hat{\boldsymbol{\theta}}$, be defined by

$$\sum_{i \in A} w_i \mathbf{g}(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) = \mathbf{0} \tag{1.3.75}$$

and let $\boldsymbol{\theta}_N$ satisfy

$$\sum_{i \in U} \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}_N) = \mathbf{0}, \tag{1.3.76}$$

where we have omitted the subscript N on U_N and A_N. Assume that $\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ for all $\boldsymbol{\theta}$ in a closed set $\mathcal{B}$ containing $\boldsymbol{\theta}^\circ$ as an interior point and all $\mathbf{x}_i$, where $\boldsymbol{\theta}^\circ$ satisfies

$$E\left\{\sum_{i \in U} \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}^\circ)\right\} = \mathbf{0}. \tag{1.3.77}$$

Assume that $\mathbf{H}(\mathbf{x}_i, \boldsymbol{\theta}) = \partial \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})/\partial \boldsymbol{\theta}'$ is continuous in $\boldsymbol{\theta}$ for all $\boldsymbol{\theta}$ in $\mathcal{B}$ and all $\mathbf{x}_i$. Assume for all $\boldsymbol{\theta}$ in $\mathcal{B}$ that

$$N^{-1} \sum_{i \in A} w_i \mathbf{H}(\mathbf{x}_i, \boldsymbol{\theta}) = N^{-1} \sum_{i \in U} \mathbf{H}(\mathbf{x}_i, \boldsymbol{\theta}) + O_p(n_{BN}^{-1/2}) \tag{1.3.78}$$

and

$$\lim_{N \to \infty} N^{-1} \sum_{i \in U} \mathbf{H}(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta}) \quad \text{a.s.},$$

where $\mathbf{H}(\boldsymbol{\theta})$ is nonsingular. Assume that

$$| \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}) | < K(\mathbf{x}_i) \tag{1.3.79}$$

for some $K(\mathbf{x})$ with finite fourth moment for all $\mathbf{x}_i$ and all $\boldsymbol{\theta}$ in $\mathcal{B}$. Assume that

$$p \lim_{N \to \infty} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ) = \mathbf{0} \tag{1.3.80}$$

and

$$p \lim_{N \to \infty} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N) \mid \mathcal{F}_N = \mathbf{0} \quad \text{a.s.} \tag{1.3.81}$$

Then

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ = \left(\sum_{i \in A} w_i \mathbf{H}(\mathbf{x}_i, \boldsymbol{\theta}^\circ)\right)^{-1} \sum_{i \in A} w_i \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}^\circ) + o_p(n_{BN}^{-1/2}) \tag{1.3.82}$$

and

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N = \left(\sum_{i\in A} w_i \mathbf{H}(\mathbf{x}_i, \boldsymbol{\theta}_N)\right)^{-1} \sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}_N) + o_p(n_{BN}^{-1/2}). \tag{1.3.83}$$

Let $\mathbf{V}_{t,xx,N}$ denote the conditional variance of $N^{-1}(\hat{\mathbf{T}}_x - \mathbf{T}_{x,N})$ conditional on $\mathcal{F}_N$, and assume that

$$\mathbf{V}_{t,xx,N}^{-1/2} N^{-1}(\hat{\mathbf{T}}_x - \mathbf{T}_{x,N}) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}) \quad \text{a.s.} \tag{1.3.84}$$

and

$$\lim_{N\to\infty} (n_{BN}\mathbf{V}_{t,xx,N} - \boldsymbol{\Sigma}_{t,xx}) = \mathbf{0} \quad \text{a.s.}, \tag{1.3.85}$$

where $\boldsymbol{\Sigma}_{t,xx}$ is positive definite for any $\mathbf{x}_j$ with positive definite covariance matrix. Then

$$[V_\infty\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ\}]^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}) \tag{1.3.86}$$

and

$$[V_\infty\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N \mid \mathcal{F}_N\}]^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}) \quad \text{a.s.}, \tag{1.3.87}$$

where

$$\begin{aligned} V_\infty\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N \mid \mathcal{F}_N\} &= \mathbf{H}^{-1}(\boldsymbol{\theta}_N)\mathbf{V}_{t,gg,N}\mathbf{H}^{-1}(\boldsymbol{\theta}_N), \\ \mathbf{H}(\boldsymbol{\theta}_N) &= N^{-1}\sum_{i\in U} \mathbf{H}(\mathbf{x}_i, \boldsymbol{\theta}_N), \\ \mathbf{V}_{t,gg,N} &= V\left\{N^{-1}\sum_{i\in A} w_i g(\mathbf{x}_i, \boldsymbol{\theta}_N) \mid \mathcal{F}_N\right\}, \\ V_\infty\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ\} &= \mathbf{H}^{-1}(\boldsymbol{\theta}^\circ)(N^{-1}\boldsymbol{\Sigma}_{gg} + \mathbf{V}_{t,gg,N}), \mathbf{H}^{-1}(\boldsymbol{\theta}^\circ), \end{aligned}$$

and $\boldsymbol{\Sigma}_{gg} = E\{\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}^\circ)\mathbf{g}'(\mathbf{x}_i, \boldsymbol{\theta}^\circ)\}$.

Proof. For $\hat{\boldsymbol{\theta}} \in \mathcal{B}$, by a Taylor expansion,

$$\begin{aligned} N^{-1}\sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) &= N^{-1}\sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}^\circ) \\ &\quad + N^{-1}\sum_{i\in A} w_i \mathbf{H}(\mathbf{x}_i, \boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ) \end{aligned}$$

$$= N^{-1}\sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i,\ \boldsymbol{\theta}^\circ) + N^{-1}\sum_{i\in A} w_i \mathbf{H}(\mathbf{x}_i,\ \boldsymbol{\theta}^\circ)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ) + o_p(|\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ\ |), \tag{1.3.88}$$

where $\boldsymbol{\theta}^*$ is between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^\circ$. The continuity of $\mathbf{H}(\mathbf{x}_i,\ \boldsymbol{\theta})$, (1.3.80), and (1.3.78) were used to obtain the second equality. Given $\epsilon > 0$, by (1.3.80), there is an n_o such that for $n > n_o$, $P\{\hat{\boldsymbol{\theta}} \in \mathcal{B}\} > 1 - \epsilon$. Therefore, result (1.3.88) holds in general. Now, by (1.3.79) and (1.3.74),

$$N^{-1}\sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i,\ \boldsymbol{\theta}^\circ) - N^{-1}\sum_{i\in U} \mathbf{g}(\mathbf{x}_i,\ \boldsymbol{\theta}^\circ) \ = \ O_p(n_{BN}^{-1/2})$$

and by (1.3.79),

$$V\left\{N^{-1}\sum_{i\in U} \mathbf{g}(\mathbf{x}_i,\ \boldsymbol{\theta}^\circ)\right\} \ = \ O(N^{-1}).$$

Therefore,

$$N^{-1}\sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i,\ \boldsymbol{\theta}^\circ) \ = \ O_p(n_{BN}^{-1/2}) \tag{1.3.89}$$

and result (1.3.82) is proven. Also see Exercise 31.

Result (1.3.83) follows by analogous arguments.

By (1.3.83) and (1.3.78),

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N \mid \mathcal{F}_N = \left(\sum_{i\in U} \mathbf{H}(\mathbf{x}_i,\ \boldsymbol{\theta}_N)\right)^{-1} \sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i,\ \boldsymbol{\theta}_N) + o_p(n_{BN}^{-1/2}) \quad \text{a.s.} \tag{1.3.90}$$

and result (1.3.87) follows from (1.3.79) and the independence assumption. By (1.3.79) and the Lindeberg Central Limit Theorem,

$$N^{-1/2}\sum_{i\in U} \mathbf{g}(\mathbf{x}_i,\ \boldsymbol{\theta}^\circ) \xrightarrow{\mathcal{L}} N(\mathbf{0},\ \boldsymbol{\Sigma}_{gg})$$

and, by (1.3.78),

$$N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ) \xrightarrow{\mathcal{L}} N\left[\mathbf{0},\ \mathbf{H}^{-1}(\boldsymbol{\theta}^\circ)\boldsymbol{\Sigma}_{gg}\mathbf{H}^{-1}(\boldsymbol{\theta}^\circ)\right]. \tag{1.3.91}$$

Result (1.3.86) follows from (1.3.90) and (1.3.91) by Theorem 1.3.6. ■

To apply Theorem 1.3.9, we require estimators of the variances. Estimators are obtained by substituting estimators for unknown parameters.

Corollary 1.3.9.1. Let the assumptions of Theorem 1.3.9 hold. Then

$$[\hat{V}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N \mid \mathcal{F}_N\}]^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}) \quad \text{a.s.}$$

and

$$[\hat{V}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ\}]^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}),$$

where

$$\hat{V}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N \mid \mathcal{F}_N\} = \hat{\mathbf{T}}_H^{-1}\hat{V}_{HT}\left\{\sum_{i\in A}\mathbf{g}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})\right\}\hat{\mathbf{T}}_H^{-1},$$

$$\hat{V}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ\} = \hat{V}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N \mid \mathcal{F}_N\} + \hat{\mathbf{T}}_H^{-1}N\hat{\boldsymbol{\Sigma}}_{gg}\hat{\mathbf{T}}_H^{-1},$$

$$\hat{\mathbf{T}}_H = \sum_{i\in A} w_i \mathbf{H}(\mathbf{x}_i, \hat{\boldsymbol{\theta}}),$$

and

$$N\hat{\boldsymbol{\Sigma}}_{gg} = \sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})\mathbf{g}'(\mathbf{x}_i, \hat{\boldsymbol{\theta}}).$$

Proof. By the assumptions that $\mathbf{H}(\mathbf{x}_i, \boldsymbol{\theta})$ and $\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})$ are continuous in $\boldsymbol{\theta}$,

$$N^{-1}\sum_{i\in A} w_i \mathbf{H}(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) - N^{-1}\sum_{i\in A} w_i \mathbf{H}(\mathbf{x}_i, \boldsymbol{\theta}^\circ) = o_p(1),$$

$$N^{-1}\sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) - N^{-1}\sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}^\circ) = O_p(n_{BN}^{-1/2}),$$

and

$$\hat{\boldsymbol{\Sigma}}_{gg} - N^{-1}\sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}^\circ)\mathbf{g}'(\mathbf{x}_i, \boldsymbol{\theta}^\circ) = O_p(n_{BN}^{-1/2}).$$

Therefore, by (1.3.78),

$$N^{-1}\sum_{i\in A} w_i \mathbf{H}(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) - N^{-1}\sum_{i\in U} \mathbf{H}(\mathbf{x}_i, \boldsymbol{\theta}^\circ) = o_p(1).$$

Furthermore, by (1.3.79),

$$N^{-1}\sum_{i\in A} w_i \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}^\circ) - N^{-1}\sum_{i\in U} \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}^\circ) = O_p(n_{BN}^{-1/2})$$

and

$$\hat{\boldsymbol{\Sigma}}_{gg} - \boldsymbol{\Sigma}_{gg} = O_p(n_{BN}^{-1/2}).$$

The conclusions then follow because the estimators of the variances are consistent estimators. ■

1.3.5 Quantiles

Means, totals, and functions of means are the most common statistics, but estimators of the quantiles of the distribution function are also important. Let y be the variable of interest and define the finite population distribution function by

$$F_{y,N}(a) = N^{-1}\sum_{i=1}^{N} d_{ai}, \tag{1.3.92}$$

where

$$\begin{aligned} d_{ai} &= 1 \quad \text{if } y_i \leq a \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Given a sample, an estimator of the distribution function at point a is the sample mean of the indicator function

$$\hat{F}_y(a) = \bar{d}_{a\pi} = \left(\sum_{i\in A}\pi_i^{-1}\right)^{-1}\sum_{i\in A}\pi_i^{-1}d_{ai}. \tag{1.3.93}$$

The finite population quantile is defined as

$$\xi_{b,N} =: Q_{y,N}(b) = \inf\{a : F_{y,N}(a) \geq b\} \tag{1.3.94}$$

and the sample quantile by

$$\hat{\xi}_b =: \hat{Q}_y(b) = \inf\{a : \hat{F}_y(a) \geq b\}. \tag{1.3.95}$$

Estimated quantiles are not simple functions of means, and therefore the results of Section 1.3.3 are not applicable. However, the relationship between the distribution function and the quantile function can be exploited to obtain useful results.

Let $\hat{s}_{ca}^2$ be the estimated variance of $\hat{F}_y(a)$ and assume that the sample is large enough so that $\hat{F}_y(a)$ can be treated as being normally distributed. Then the hypothesis that $F_y(a) = b$ will be accepted at the α level if $\hat{F}(a)$ falls in the interval

$$(b - t_\alpha \hat{s}_{ca},\ b + t_\alpha \hat{s}_{ca}), \tag{1.3.96}$$

where t_α is the α percentage point of the normal distribution. If $\hat{F}_y(a)$ is in the interval defined in (1.3.96), then

$$\hat{Q}_y(b - t_\alpha \hat{s}_{ca}) \leq Q_y(b) \leq \hat{Q}_y(b + t_\alpha \hat{s}_{ca}), \tag{1.3.97}$$

where $F(a) = b$. Therefore, $[\hat{Q}_y(b - t_\alpha \hat{s}_{ca}),\ \hat{Q}(b + t_\alpha \hat{s}_{ca})]$ is a $1 - \alpha$ confidence interval for $Q_y(b)$. Intervals of this type are sometimes called *test inversion intervals*. The interval (1.3.97) is also called the *Woodruff interval* in the survey sampling literature. See Woodruff (1952).

Using a plot of the distribution function, one can see that shifting the function up by an amount δ will shift the quantile left by an amount approximately equal to δ divided by the slope of the distribution function. This local approximation can be used to approximate the distribution of a quantile. For simple random samples from a distribution with a density, the limiting distribution of a quantile associated with a positive part of the density is normal because the error in the quantile can be written

$$\hat{\xi}_b - \xi_b \ = \ [f_y(a)]^{-1} (b - \bar{d}_{a,\pi}) + o_p(n^{-1/2}), \qquad (1.3.98)$$

where $\xi_b = a$ and $f_y(a)$ is the density of y evaluated at a. Equation (1.3.98) is called the *Bahadur representation*. See Bahadur (1966), Ghosh (1971), and David (1981, Section 9.2). Francisco and Fuller (1991) extended representation (1.3.98) to a more general class of samples and used the representation to show that sample quantiles for complex samples are normally distributed in the limit.

Theorem 1.3.10. Let a sequence of finite populations be created as samples from a superpopulation with cumulative distribution function $F_y(\cdot)$ and finite fourth moments. Let $\xi_b^\circ = a^\circ$ be the bth quantile. Assume that the cumulative distribution function $F_y(a)$ is continuous with a continuous positive derivative on a closed interval B containing a° as an interior point. Assume that the sequence of designs is such that

$$n_{BN}^{1/2} N^{-1} (\hat{T}_x - N\mu_x) \xrightarrow{\mathcal{L}} N(0,\ \sigma_{xx}),$$

$$n_{BN}^{1/2} N^{-1} (\hat{T}_x - N\bar{x}_N) \xrightarrow{\mathcal{L}} N(0,\ M_{xx}),$$

$$[V\{\hat{T}_x \mid \mathcal{F}_N\}]^{-1} \hat{V}_{HT}\{\hat{T}_x \mid \mathcal{F}_N\} - 1 \ = \ O_p(n_{BN}^{-1/2}),$$

$$[V\{\bar{x}_\pi\}]^{-1} \hat{V}\{\bar{x}_\pi\} - 1 \ = \ O_p(n_{BN}^{-1/2}),$$

for any x with positive variance and fourth moment, where $n_{BN} = E\{n_N\}$, $\hat{V}_{HT}\{\hat{T}_x \mid \mathcal{F}_N\}$ is the Horvitz–Thompson estimator of the variance of $\hat{T}_x - T_x$ given $\mathcal{F}_N$, $\hat{V}\{\bar{x}_\pi\}$ is an estimator of the unconditional variance of $\bar{x}_\pi - \mu_x$, and μ_x is the superpopulation mean. Assume that $n_N V\{\hat{F}_y(a)\}$ and $n_N V\{\hat{F}_y(a) - F_{y,N}(a) \mid \mathcal{F}_N\}$ are positive and continuous in a for $a \in B$. Assume that

$$V\{\hat{F}_y(a + \delta) - \hat{F}_y(a)\} \ \leq \ C n_N^{-1} \mid \delta \mid$$

for some $0 < C < \infty$, for all N, and for all a and $a + \delta$ in B. Then

$$\hat{s}_{ca}^{-1}\hat{f}_y(\hat{\xi}_b)(\hat{\xi}_b - a^\circ) \xrightarrow{\mathcal{L}} N(0,\ 1), \tag{1.3.99}$$

where $\hat{s}_{ca}^2 = \hat{V}\{\hat{F}_y(a)\}$,

$$\hat{f}_y(\hat{\xi}_b) = (2t_\alpha \hat{s}_{ca})[\hat{Q}(b + t_\alpha \hat{s}_{ca}) - \hat{Q}(b - t_\alpha \hat{s}_{ca})]^{-1},$$

$\hat{a} = \hat{\xi}_b$, t_α is defined by $\Phi(t_\alpha) = 1 - 0.5\alpha$, and $\Phi(\cdot)$ is the distribution function of a standard normal random variable.

Also,

$$\hat{s}_{HT,ca}^{-1}\hat{f}_y(\hat{\xi}_b)(\hat{\xi}_b - \xi_{b,N}) \xrightarrow{\mathcal{L}} N(0,1),$$

where $\hat{s}_{HT,ca}^2 = \hat{V}\{\hat{F}_y(a) - F_{y,N}(a) \mid \mathcal{F}_N\}$.

Proof. Omitted. See Francisco and Fuller (1991) and Shao (1994). ■

In Theorem 1.3.10, the ratio of the difference between two values of the sample distribution function to the distance between the points defining the values is used to estimate the density. The use of $t_\alpha = 2$ in (1.3.99) to estimate $f_y(a)$ seems to work well in practice. The estimator of $f_y(a)$ in (1.3.99) can be viewed as a regression in the order statistics for order statistics "close" to $\hat{\xi}_b$.

Let $y_{(r)}$ be the largest order statistic less than $\hat{Q}(\hat{\xi}_b - t_\alpha \hat{s}_{ca})$, let $y_{(m)}$ be the smallest order statistic greater than $\hat{Q}(\hat{\xi}_b + t_\alpha \hat{s}_{ca})$, and let a "smoothed" estimator of the distribution function of $y_{(i)}$ be

$$z_i = 0.5[\hat{F}(y_{(i)}) + \hat{F}(y_{(i-1)})]. \tag{1.3.100}$$

Let $\hat{\theta}_0$ and $\hat{\theta}_1$ be the regression coefficients obtained in a weighted regression of z_i on $(1,\ y_{(i)})$ for $i = r,\ r+1,\ \ldots,\ m$. Then $\hat{\theta}_1$ is an estimator of $f_y(\xi_b)$ and

$$\tilde{\xi}_b = \hat{\theta}_1^{-1}(b - \hat{\theta}_0) \tag{1.3.101}$$

is a smoothed estimator of ξ_b. If only $y_{(m)}$ and $y_{(r)}$ are used in the regression

$$\hat{\theta}_1 = (y_{(m)} - y_{(r)})^{-1}(z_m - z_r),$$

$\hat{\theta}_0 = z_k - \hat{\theta}_1 y_{(k)}$, and

$$\hat{\xi}_b = \bar{y}_{(r)} + (z_m - z_r)^{-1}(y_{(m)} - y_{(r)})(b - z_r).$$

There are many smoothed estimators of quantiles. See Silverman (1986) and Scott (1992).

1.4 METHODS OF UNEQUAL PROBABILITY SAMPLE SELECTION

The literature contains numerous procedures for the selection of nonreplacement unequal probability samples. The number of procedures is indicative of the difficulty of constructing a completely general procedure that is not extremely cumbersome computationally. We consider only sampling schemes where selection is not a function of the y values. For selection procedures that are functions of y, see Thompson and Seber (1996).

Selection procedures have been classified by Carroll and Hartley (1964) as draw-by-draw methods, mass draw procedures wherein samples are rejected if duplication occurs, and systematic procedures. The draw-by-draw and mass draw methods require computation of "working probabilities" if the probability of selection is to be maintained at the values specified. The working probabilities are typically given as the solutions to a system of N, Nn, or $N(n-1)$ equations. Some procedures have been demonstrated to be superior to replacement sampling for $n = 2$ (Fellegi, 1963), while others are justified on the basis of joint probabilities that guarantee nonnegative estimators of variance (Hanurav, 1967; Vijayan, 1968). Alternative procedures have been discussed by Jessen (1969) and Rao (1978), and procedures have been reviewed extensively by Brewer and Hanif (1983) and Tillé (2006).

Perhaps the most common method of selecting an unequal probability sample is the systematic procedure described in Section 1.2.4. The systematic procedure is easy to implement but has the disadvantage that no design-unbiased estimator of the variance is available.

The following two draw-by-draw methods of selecting a sample of size 2 yield the same joint inclusion probabilities. The first was suggested by Brewer (1963a) and the second by Durbin (1967). Let p_i be a set of positive numbers (probabilities) with the properties that $\Sigma_{i=1}^{N} p_i = 1$ and $p_i < 0.5$ for all i. The selection probability is then $\pi_i = 2p_i$.

Procedure 1

1. Select a unit with probability q_{1j}, where

$$q_{1j} = \left(\sum_{i=1}^{N} (1-2p_i)^{-1} p_i (1-p_i) \right)^{-1} (1-2p_j)^{-1} p_j (1-p_j). \tag{1.4.1}$$

2. Select a second unit with probability

$$q_{2j} = (1-p_{i(1)})^{-1} p_j, \tag{1.4.2}$$

where $p_{i(1)}$ is the value of p for the unit selected at the first draw.

Procedure 2

1. Select a unit with probability p_i.
2. Select a second unit with probability $p_i^{-1}\pi_{ij}$, $j \neq i$, where i is the unit selected on the first draw and

$$\pi_{ij} = \frac{\pi_i \pi_j}{2(1+A)} \left(\frac{1}{1-\pi_i} + \frac{1}{1-\pi_j} \right), \tag{1.4.3}$$

where

$$A = \frac{1}{2} \sum_{i=1}^{N} \frac{\pi_i}{1-\pi_i} = \sum_{i=1}^{N} \frac{p_i}{1-2p_i}. \tag{1.4.4}$$

For both procedures, the joint probability is given by (1.4.3) and the total probability of selecting unit i is $\pi_i = 2p_i$. Under selection procedure 2, the probability that the unit is selected on the first draw is p_i and the probability that it is selected on the second draw is p_i. Fuller (1971) gave the following motivation for the joint probabilities (1.4.3).

Assume that the population of N values of $p_i^{-1}y_i$ is a random sample from a normal population with variance σ^2. Then considering the population of all such populations, the variance of the Yates–Grundy–Sen estimated variance for samples of size 2 is

$$8\sigma^4 \sum_{i<j}^{N} \pi_{ij}^{-1} (\pi_i \pi_j - \pi_{ij})^2. \tag{1.4.5}$$

Therefore, under this model, minimization of the summation with respect to π_{ij} will result in a minimum variance for the estimated variance. Since minimization of this expression leads to a system of nonlinear equations for the π_{ij}, consider the approximation obtained by replacing the π_{ij} in the denominator by $\pi_i \pi_j$.

For $n = 2$ those π_{ij} that minimize

$$\sum_{i<j}^{N} \frac{(\pi_i \pi_j - \pi_{ij})^2}{\pi_i \pi_j} \tag{1.4.6}$$

subject to the restrictions that

$$\sum_{\substack{j=1 \\ j \neq i}}^{N} \pi_{ij} = \pi_i \qquad \text{for all} \quad i \tag{1.4.7}$$

are the π_{ij} of (1.4.3). The π_{ij} of (1.4.3) are positive, and furthermore, for $0 < \pi_i < 1.0$,

$$\frac{1}{1-\pi_i} + \frac{1}{1-\pi_j} = \frac{\pi_i}{1-\pi_i} + \frac{\pi_j}{1-\pi_j} + 2 < 2 + 2A$$

and hence $\pi_{ij} < \pi_i \pi_j$. Therefore, the joint probabilities (1.4.3) permit the construction of an unbiased nonnegative estimator of variance. The following theorem demonstrates that the sampling scheme is always more efficient than replacement sampling.

Theorem 1.4.1. The variance of the Horvitz–Thompson estimator for the sampling scheme with joint probabilities (1.4.3) is never greater than that of estimator (1.2.70) for replacement sampling, equality holding only if all $(z_i - z_j)^2 = 0$, where $z_i = y_i \pi_i^{-1}$.

Proof. Using the variance expression (1.2.28), the variance of the Horvitz–Thompson estimated total is

$$V_N(\hat{Y}) = \sum_{i<j} \left[\pi_i \pi_j - \frac{\pi_i \pi_j}{2(1+A)} \left(\frac{1}{1-\pi_i} + \frac{1}{1-\pi_j} \right) \right] (z_i - z_j)^2$$

and the variance of the replacement sampling estimator (1.2.70) is

$$V_R(\hat{Y}) = \frac{1}{2} \sum_{i<j} \pi_i \pi_j (z_i - z_j)^2 = \sum_{j=1}^{N} \pi_j \left(z_j - 0.5Y \right)^2 .$$

Then

$$V_R - V_N = \frac{1}{1+A} \left[-(1+A) V_R(\hat{Y}) + \frac{1}{2} \sum_i \frac{\pi_i}{1-\pi_i} \sum_j \pi_j (z_i - z_j)^2 \right]$$

$$= \frac{1}{1+A} \left[\sum_i \frac{\pi_i^2}{1-\pi_i} (z_i - 0.5Y)^2 \right] \geq 0,$$

equality holding only if all $z_i \equiv 0.5Y$. ∎

Two procedures that maintain inclusion probabilities equal to np_i for $n > 2$ are that of Brewer (1963a) and that proposed by Rao (1965) and Sampford (1967). In the Brewer (1963a) procedure, the first selection for a sample of size n is made with probability

$$q_{jN} = \left[\sum_{i=1}^{N} (1 - np_i) p_i (1 - p_i) \right]^{-1} (1 - np_j)^{-1} p_j (1 - p_j),$$

the next with probabilities proportional to

$$[1-(n-1)p_j]^{-1}\, p_j(1-p_j),$$

and so on.

Another way to select unequal probability samples is to select a replacement sample and reject the sample if the sample contains duplicates. Hájek (1964) studied this procedure. To illustrate, consider the selection of a sample of size 2 from a population of size N with draw probabilities p_i. The total probability of selecting the sample is

$$1 \;=\; \left(\sum_{i=1}^{N} p_i\right)^2$$

and the probabilities of selecting one of the units twice is

$$P\{\text{repeated selection}\} \;=\; \sum_{i=1}^{N} p_i^2.$$

Thus, the joint probability of element i and element j, $i \neq j$, appearing in a sample where samples with repeated elements are rejected is

$$\pi_{ij|NR} \;=\; P\{(i,\, j) \in A \mid NR\} \;=\; \left(1-\sum_{k=1}^{N} p_k^2\right)^{-1} p_i p_j,$$

where NR denotes no repeated elements in the sample. The probability that element i appears in the sample is

$$\pi_{i|NR} \;=\; \sum_{\substack{j=1 \\ j\neq i}}^{N} \pi_{ij} \;=\; \left(1-\sum_{k=1}^{N} p_k^2\right)^{-1} (1-p_i)\, p_i.$$

If a nonreplacement sample with selection probabilities close to the specified π_i is desired, working probabilities must be specified. Hájek (1964) suggested approximate p_i, and Carroll and Hartley (1964) gave an iterative procedure, described by Brewer and Hanif (1983), for determining working probabilities. Chen, Dempster, and Liu (1994) give a computational algorithm that can be used for sample selection. For a complete discussion, see Tillé (2006, Chapter 5). Also see Section 3.4.

1.5 REFERENCES

Sections 1.1, 1.2. Brewer (1963b), Cochran (1946, 1977), Goldberger (1962), Graybill (1976), Hansen and Hurwitz (1943), Horvitz and

Thompson (1952), Narain (1951), Royall (1970), Sen (1953), Stuart and Ord (1991), Yates (1948), Yates and Grundy (1953).

Section 1.3. Bickel and Freedman (1984), Binder (1983), Blight (1973), Cochran (1946, 1977), Francisco (1987), Francisco and Fuller (1991), Fuller (1975, 1987b, 1996), Hájek (1960), Hannan (1962), Isaki and Fuller (1982), Krewski and Rao (1981), Madow (1948), Madow and Madow (1944), Papageorgiou and Karakostas (1998), Rao and Wu (1987), R. R. Rao (1962), Rubin-Bleuer and Kratina (2005), Sen (1988), Shao (1994), Thompson (1997), Woodruff (1952, 1971), Xiang (1994).

Section 1.4. Brewer (1963a), Brewer and Hanif (1983), Carroll and Hartley (1964), Durbin (1967), Fellegi (1963), Fuller (1971), Hájek (1964), Hanurav (1967), Hedayat and Sinha (1991), Jessen (1969), Rao (1965, 1978), Rao, Hartley, and Cochran (1962), Sampford (1967), Vijayan (1968), Yates and Grundy (1953).

1.6 EXERCISES

1. (Section 1.2.1) Let $\mathbf{d} = (I_1, I_2, \ldots, I_N)$, as defined in (1.2.4). Show that a matrix expression for the variance of the design linear estimator $\hat{\theta}$ of (1.2.17) is

$$V\{\hat{\theta} \mid \mathcal{F}\} = \mathbf{y}_N \mathbf{W}_N \boldsymbol{\Sigma}_{dd} \mathbf{W}_N \mathbf{y}'_N,$$

where $\mathbf{y}_N = (y_1, y_2, \ldots, y_N)$, $\mathbf{W}_N = \text{diag}(w_1, w_2, \ldots, w_N)$ is a diagonal matrix whose diagonal elements starting in the upper left corner are $w_1, w_2, \ldots, w_N$, and $\boldsymbol{\Sigma}_{dd}$ is the covariance matrix of $\mathbf{d}$.

2. (Section 1.2.4) Derive the joint probabilities of selection for systematic samples of size 3 selected from the population of Table 1.1 with the measures of size of Table 1.1.

3. (Section 1.2.4) Assume that a population satisfies

$$y_t = \sin 2\pi k^{-1} t$$

for $t = 1, 2, \ldots, N$. Give the variance of the sample mean of a systematic sample of size 10 as an estimator of the population mean for a population with $k = 6$ and $N = 60$. Compare this to the variance of the mean of a simple random nonreplacement sample and to the variance of the mean of a stratified sample, where the population is divided into

two equal-sized strata with the smallest 30 indices in the first stratum. How do the results change if the sample size is 12?

4. (Section 1.2.3) Let a stratified population be of the form described in Section 1.2.3 with H strata of sizes $N_1, N_2, \ldots, N_H$. Find the optimal allocation to strata to estimate the linear function

$$\theta = \sum_{h=1}^{H} \alpha_h \bar{y}_{Nh},$$

where α_h, $h = 1, 2, \ldots, H$, are fixed constants. Assume equal costs for observations in the strata.

5. (Section 1.2, 1.3) Consider the following sampling scheme. A simple random sample of n households is selected from N households. The ith household contains M_i family members. In each household selected, one family member is selected at random and interviewed. Give the probability that person ij (the jth person in the ith household) is interviewed. Define the Horvitz–Thompson estimator of the total of y. Give the joint probability that any two people appear in the sample. Is it possible to construct an unbiased estimator of the variance of the Horvitz–Thompson estimator?

Consider the estimator of the variance,

$$\hat{V}\{\bar{y}_\pi \mid \mathcal{F}_N\} = n(n-1)^{-1} \left(\sum_{t=1}^{n} M_t \right)^{-2} \sum_{t=1}^{n} M_t^2 (y_{tj} - \bar{y}_\pi)^2,$$

where

$$\bar{y}_\pi = \left(\sum_{t=1}^{n} M_t \right)^{-1} \sum_{t=1}^{n} M_t y_{tj}$$

and y_{ij} is the value observed for person ij. Assume that the household size satisfies $1 \le M_t \le K$ for some K and assume that the finite population is a random sample from a superpopulation, where (y_{tj}, M_t) has a distribution with finite fourth moments and (y_{tj}, M_t) is independent of (y_{ij}, M_i) for $t \neq i$. Show that

$$n[\hat{V}\{\bar{y}_\pi \mid \mathcal{F}_N\} - V\{\bar{y}_\pi \mid \mathcal{F}_N\}] = o_{p(1)} \quad \text{a.s.}$$

as $N \to \infty$, $n \to \infty$, and $N^{-1}n \to 0$, where

$$V\{\bar{y}_\pi \mid \mathcal{F}_N\} = E\left\{ \left[\bar{y}_\pi - \left(\sum_{i=1}^{N} M_i \right)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ij} \right]^2 \middle| \mathcal{F}_N \right\} \quad \text{a.s.}$$

6. (Section 1.3.3) Consider a population of x values with $x_i > 0$ for all i. Let model (1.3.58) hold with $E\{\bar{y}_N \mid \bar{x}_N\} = \bar{x}_N$, $\sigma^2 = 1$, and $\alpha = 2$. Let samples be selected from a finite population generated by (1.3.58) with π_i proportional to x_i. Using the approximate design variances, for what values of β_0 and β_1 is $V\{\bar{y}_\pi \mid \mathcal{F}\} < V\{\bar{y}_{HT} \mid \mathcal{F}\}$? You may consider the set of finite populations with common $(x_1, x_2, \ldots, x_N)$.

7. (Section 1.2.4) In Section 1.2.4 it is stated that the sample mean for a systematic sample with equal probabilities for samples of unequal sizes is biased for the population mean. Derive the bias and construct an unbiased estimator of the population mean. Assign probabilities to the two types of samples so that the sample mean is unbiased for the population mean.

8. (Section 1.2) Consider a population of size 9 that has been divided into two strata of size 4 and 5, respectively. Assume that a stratified sample of size 5 is to be selected, with two in stratum 1 and three in stratum 2. Let $\mathbf{d}$ be the nine-dimensional vector of indicator variables defined in (1.2.4). Give the mean and covariance matrix of $\mathbf{d}$.

9. (Section 1.2.5) Assume that a replacement sample of size 3 is selected from a population of size N with draw probabilities $(p_1, p_2, \ldots, p_N)$.

   (a) What is the probability that element i appears in the sample three times?

   (b) What is the probability that element i is observed given that the sample contains only one distinct unit?

   (c) What is the probability that element i is selected twice?

   (d) What is the probability that element i is selected twice given that some element was selected twice?

   (e) What is the probability that element i appears in the sample at least once?

   (f) What is the probability that element i appears in the sample given that the sample contains three distinct units?

10. (Section 1.2) Consider a design for a population of size N, where the design has $N+1$ possible samples. N of the samples are of size 1, where each sample contains one of the possible N elements. One sample is of size N, containing all elements in the population. Each of the $N + 1$ possible samples is given an equal probability of selection.

    (a) What is the probability that element j is included in the sample?

(b) What is the expected sample size?

(c) What is the variance of the sample size?

(d) If the finite population is a realization of $NI(0,\ \sigma^2)$ random variables, what is the variance (over all populations and samples) of $\hat{T}_y - T_y$, where $\hat{T}_y$ is the Horvitz–Thompson estimator of the finite population total?

(e) Compare the variance of the estimated total of part (d) with the variance of $N(\bar{y}_n - \bar{y}_N)$ for a simple random sample of size n.

(f) Consider the estimator that conditions on sample size

$$\tilde{T}_y \ = \ Nn^{-1}\sum_{i\in A} y_i,$$

where n is the realized sample size. Show that this estimator is design unbiased for T_y.

(g) Give the variance of $\hat{T}_y - T_y$ of part (f) under the conditions of part (d).

11. (Section 1.2) Assume an R-person list, where the ith person appears on the list r_i times. The total size of the list is N. Assume that a simple random nonreplacement sample of n lines is selected from the list. For each line selected, a person's characteristic, denoted by y_i, the total number of lines for person i, denoted by r_i, and the number of times that person i occurs in the sample, denoted by t_i, are determined. Assume that r_i is known only for the sample.

(a) Give an estimator for the number of people on the list.

(b) Give an estimator for the total of y.

12. (Section 1.2) The possible samples of size 3 selected from a population of size 5 are enumerated in Table 1.3. The table also contains probabilities of selection for a particular design.

(a) Compute the probabilities of selection, π_i, for $i = 1,\ 2,\ \ldots,\ 10$.

(b) Compute the joint probabilities of selection, π_{ij}, for all possible pairs.

(c) Compute the joint probability of selection for each pair of pairs.

(d) Assume that a population of finite populations of size 5 is such that each finite population is a sample of 5 $NI(\mu,\ \sigma^2)$ random variables. What are the mean and variance of the Horvitz–Thompson

Table 1.3 Design for Samples of Size 3 from a Population of Size 5

Sample	Sample Elements	Prob. of Sample	Sample	Sample Elements	Prob. of Sample
1	1,2,3	0.06	6	1,4,5	0.10
2	1,2,4	0.07	7	2,3,4	0.11
3	1,2,5	0.08	8	2,3,5	0.12
4	1,3,4	0.09	9	2,4,5	0.13
5	1,3,5	0.10	10	3,4,5	0.14

estimator of the total of the finite population when the sample is selected according to the design of the table?

(e) Under the assumptions of part (d), find the mean and variance of the variance estimator (1.2.33).

(f) Under the assumptions of part (d), find the mean and variance of $\hat{\boldsymbol{\theta}}_k - \bar{y}_N$, $k = 1, 2$, where

$$\hat{\boldsymbol{\theta}}_1 = N^{-1} \sum_{i \in A} \pi_i^{-1} y_i$$

and

$$\hat{\boldsymbol{\theta}}_2 = \left(\sum_{i \in A} \pi_i^{-1} \right)^{-1} \left(\sum_{i \in A} \pi_i^{-1} y_i \right).$$

13. (Section 1.2) [Sirken (2001)] Let a population be composed of N units with integer measures of size m_i, $i = 1, 2, \ldots, N$. Let $M_0 = 0$ and let $M_j = \Sigma_{i=1}^{j} m_i$, $j = 1, 2, \ldots, N$. Consider two sampling procedures:

(a) A replacement simple random sample of n integers is selected from the set $\{1, 2, \ldots, M_N\}$. If the selected integer, denoted by k, satisfies

$$M_{i-1} < k \leq M_i,$$

element i is in the sample.

(b) A nonreplacement simple random sample of n integers is selected from the set $\{1, 2, \ldots, M_N\}$. The rule for identifying selected units is the same as for procedure (a).

For each procedure:

i. Determine the probability that element i is selected for the sample exactly once.

ii. Determine the probability that element i is selected for the sample at least once.

iii. Determine the joint probability that elements i and k appear together in the sample.

14. (Section 1.2.2) Let a finite population of size N be a random sample from an infinite population satisfying the model

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + e_i, \\ e_i &\sim NI(0,\ x_i\sigma^2), \end{aligned}$$

where x_i is distributed as a multiple of a chi-square random variable with d, $d \geq 3$, degrees of freedom, and e_i is independent of x_j for all i and j. Let a Poisson sample of expected size n_B be selected with the selection probability for element i proportional to x_i. What is the expected value of

$$n_B^{-1} \sum_{i \in U} x_i \left(\sum_{i \in A} x_i^{-1},\ \sum_{i \in A} y_i,\ \sum_{i \in A} x_i^{-2} y_i,\ \sum_{i \in A} x_i^{-1} y_i \right)?$$

15. (Section 1.2.2) Assume that a sample of n elements is selected using Poisson sampling with probabilities π_i, $i = 1, 2, \ldots, N$. Find the design variance of the linear function

$$\hat{\theta} = \sum_{i \in A} g_i y_i,$$

where the g_i, $i = 1, 2, \ldots, N$, are fixed coefficients. Determine an estimator of the variance of $\hat{\theta}$.

16. (Section 1.3) Let a sequence of populations of size N be selected from a distribution with mean μ and variance σ^2. Give an example of a sequence (N, n_N) such that

$$N^{-1} \left(\sum_{i \in A} y_i + (N - n_N) n_N^{-1} \sum_{i \in A} y_i - \sum_{i \in U} y_i \right) = O_p(n_N^{-\beta}),$$

where $\beta > 0.5$.

17. (Section 1.3.3) Let $y_i \sim NI(0,\ 1)$ and define x_i by

$$\begin{aligned} x_i &= y_i && \text{with probability } 0.5 \\ &= -y_i && \text{with probability } 0.5, \end{aligned}$$

where the event defining x_i is independent of y_i. Let $(\bar{x},\ \bar{y}) = n^{-1}\sum_{i=1}^{n}(x_i,\ y_i)$.

(a) Prove that

$$n^{1/2}(\bar{x},\ \bar{y}) \xrightarrow{\mathcal{L}} N(\mathbf{0},\ \mathbf{I}).$$

(b) Does the conditional distribution of $\bar{x}$ given $\bar{y}$ converge to a normal distribution almost surely?

(c) Let

$$\begin{aligned} \bar{z} &= \bar{y} && \text{with probability } 0.5 \\ &= -\bar{y} && \text{with probability } 0.5, \end{aligned}$$

where the event defining $\bar{z}$ is independent of $\bar{y}$. Show that $\bar{z}$ is a normal random variable. Is the conditional distribution of $\bar{z}$ given $\bar{y}$ normal?

18. (Section 1.3) Assume that a finite population of size N is a realization of N binomial trials with probability of success equal to p. Let the finite population proportion be p_N. Assume that a sample of size n is selected with replacement from the finite population. Show that the variance of the sample proportion, $\hat{p}$, as an estimator of the infinite population proportion is

$$V\{\hat{p} - p\} = N^{-1}n^{-2}\left[(n-1)N + n^2\right]p(1-p),$$

where $\hat{p}$ is the replacement estimator of the mean obtained by dividing the estimated total (1.2.66) by N.

19. (Section 1.2.2) Assume that a Poisson sample is selected with known probabilities π_i, where the π_i differ. Let n_B be the expected sample size. Find the design mean and variance of the estimator

$$\hat{T}_y = Nn_B^{-1}\sum_{i\in A} y_i.$$

Is it possible to construct a design-unbiased estimator of the design variance of $\hat{T}_y$?

20. (Section 1.2) Assume that a simple random sample of size n is selected from a population of size N and then a simple random sample of size m

is selected from the remaining $N - n$. Show that the $n + m$ elements constitute a simple random sample from the population of size N. What is $C\{\bar{y}_n, \bar{y}_m \mid \mathcal{F}\}$, where $\bar{y}_n$ is the mean of the first n elements and $\bar{y}_m$ is the mean of the second m elements?

21. (Section 1.3) Show that the assumptions

    (a) $K_L < \pi_i < K_U$

    (b) $\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \mid y_i \mid^{2+\delta} = M_{2\delta} > 0$

    for positive constants δ, K_L and K_U are sufficient for

$$\lim_{N\to\infty} \sup_{l\le k\le N} \left[\sum_{i=1}^{N} y_i^2 \pi_i (1 - \pi_i)\right]^{-1} y_k^2 = 0.$$

22. (Section 1.3) Consider a sequence of finite populations composed of H_N strata. Assume that random samples of size 2 are selected from each stratum. Do the μ_h need to be bounded for the results of Theorem 1.3.2 to hold?

23. (Section 1.2.2) Assume that the data in Table 1.4 are a Poisson sample selected with the probabilities given in the table.

    (a) Estimate the fraction of managers who are over 50. Estimate the variance of your estimator.

    (b) Estimate the fraction of employees who have a manager over 50. Estimate the variance of your estimator.

    (c) Estimate the population covariance between age of manager and number of employees for the population of managers.

24. (Section 1.3.3) Consider the estimator $\bar{\pi}_N \hat{R}_{\pi y}$, where

$$\hat{R}_{\pi y} = \left(\sum_{i\in A} \pi_i^{-1} \pi_i\right)^{-1} \sum_{i\in A} \pi_i^{-1} y_i.$$

    The denominator of the ratio is n, but the summation expression emphasizes the fact that n is an estimator of $N\bar{\pi}_N$. Thus, under the assumptions of Theorem 1.3.8, $\bar{\pi}_N \hat{R}_{\pi y} = \bar{y}_N + O_p(n_{BN}^{-1/2})$. Find $E\{\bar{\pi}_N \hat{R}_{\pi y} \mid \pi_n\}$ under model (1.3.58), where π_n is the set of π_i in A. Assume that $\alpha = 2$ and $\beta_0 = 0$ in model (1.3.58). Find the best linear unbiased estimator

Table 1.4 Poisson Sample of Managers

Probability	Age of Manager	Number of Employees	Probability	Age of Manager	Number of Employees
0.016	35	10	0.070	50	31
0.016	36	4	0.100	51	47
0.036	41	15	0.100	56	55
0.040	45	25	0.120	62	41
0.024	45	8	0.100	64	50

of β_1, conditional on π_n. What is the best linear unbiased predictor of T_y?

25. (Section 1.3.2) In the proof of Theorem 1.3.4 we demonstrated that for a sequence of Bernoulli samples there is a corresponding sequence of simple random samples such that the difference between the two means is $O_p(n_B^{-3/2})$. Given a sequence of simple random samples, construct a corresponding sequence of Bernoulli samples such that the difference between the two means is $O_p(n_B^{-3/2})$.

26. (Section 1.3.1) Prove the following.

    Result. Let $\{X_n\}$ be a sequence of random variables such that

    $$\begin{aligned} E\{X_n\} &= 0, \\ V_n\{X_n\} &= O_p(n^{-\alpha}), \end{aligned}$$

    where $V_n\{X_n\}$ is the sequence of variances of X_n and $\alpha > 0$. Then

    $$X_n = O_p(n^{-0.5\alpha}).$$

27. (Section 1.2.5) Let a population of size N be given and denoted by $\mathcal{F}_N$. Let a second finite population of size nN be created by replicating each of the original observations n times. Denote the second population by $\mathcal{F}_{nN}$. Is $V\{\bar{x}_{rn} - \bar{x}_N \mid \mathcal{F}_N\}$ for an equal probability replacement sample of size n selected from $\mathcal{F}_N$ the same as the variance of $V\{\bar{x}_n - \bar{x}_N \mid \mathcal{F}_{nN}\}$ for a simple random nonreplacement sample of size n selected from $\mathcal{F}_{nN}$? The statistic $\bar{x}_{rN}$ for the replacement sample is the mean of the y for the n draws, not the mean of distinct units.

28. (Section 1.2.3) Show that the estimator (1.2.56) is the Horvitz–Thompson variance estimator.

29. (Section 1.2.1) Show that

$$\begin{aligned} S^2 &= (N-1)^{-1}\sum_{i\in U}(y_i - \bar{y}_N)^2 \\ &= 0.5N^{-1}(N-1)^{-1}\sum_{i\in U}\sum_{j\in U}(y_i - y_j)^2. \end{aligned}$$

Hence, show that expression (1.2.28) for simple random sampling is

$$V\{N\bar{y}_n \mid \mathcal{F}\} = N^2(n^{-1} - N^{-1})S^2.$$

30. (Section 1.2.6) In the example in the text, the selection of a sample of size 3 from a population of size 6 led to selection probabilities of (9/16, 8/16, 7/16, 7/16, 8/16, 9/16). What is the joint selection probability of units 1 and 2? Of units 3 and 4? Of units 1 and 6?

31. (Section 1.3.1) Prove:

Lemma 1.6.1. If $\hat{\theta}_n = B_n + o_p(|\ \hat{\theta}_n\ |)$, then $\hat{\theta}_n = O_p(\ |B_n|\ )$.

32. (Section 1.2.1) Let A_1 be the indexes of a simple random nonreplacement sample of size n_1 selected from a finite population of size N. Let A_2 be the indexes of a simple random sample of size n_2 selected from the remaining $N - n_1$ elements. Let $\bar{y}_1$ be the mean for sample A_1 and $\bar{y}_2$ be the mean for sample A_2. What is $C\{\bar{y}_1, \bar{y}_2 \mid \mathcal{F}\}$?

33. (Section 1.4) Let a population of size N have assigned probabilities $(p_1, p_2, \ldots, p_N)$ and consider the following successive selection scheme. At step 1 a unit is selected from the N units with probability p_i. At step 2 a unit is selected from the remaining $N-1$ units with probability $p_j(1-p_i)^{-1}$. What is the probability that unit j is in a sample of size 2? What is the probability that units j and k will be in a sample of size 2? Rosen (1972) has studied this selection scheme for n selections.

34. (Section 1.3.2) In Theorem 1.3.4 it is asserted that

$$E\{(n_o^{-1} - n_B^{-1})^2 \mid n_o > 0\} = O(n_B^{-3}).$$

Show that the conditions of Theorem 5.4.4 of Fuller (1996) are satisfied for $n_o^{-2}n_B^2$ and hence that the conditions of Theorem 5.4.3 of Fuller (1996) are satisfied for $(n_o^{-1}n_B - 1)^2$. *Hint*: Let x of Theorem 5.4.4 be $n_B^{-1}n_o$.

35. (Section 1.3) Prove:

Theorem 1.6.1. Let $\{y_i\}$ be a sequence of fixed numbers and let $\mathcal{F}_N = \{y_1, y_2, \ldots, y_N\}$. Assume that

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} y_i = \mu$$

and

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} |\, y_i \,|^{1+\epsilon} = K_\epsilon$$

for some $\epsilon > 0$, where μ and K_ϵ are finite. Let $\{\pi_i\}$ be a sequence of probabilities with $0 < c_s < \pi_i < c_g < 1$. Let a sequence of Poisson samples be defined with selection probabilities π_i, where $A_{N-1} \subseteq A_N$. Then

$$\lim_{N\to\infty} N^{-1} \sum_{i\in A_N} \pi_i^{-1} y_i = \mu \quad \text{a.s.} \tag{1.6.1}$$

and

$$\lim_{N\to\infty} \left(\sum_{i\in A_N} \pi_i^{-1} \right)^{-1} \sum_{i\in A_N} \pi_i^{-1} y_i = \mu \quad \text{a.s.} \tag{1.6.2}$$

36. (Section 1.3) Let two finite populations of size N_1 and N_2 be realizations of *iid* random variables from a distribution $F(y)$. Let a simple random sample of size n_1 be selected from N_1 and a simple random sample of size n_2 be selected from N_2. Show that the sample of $n_1 + n_2$ elements can be treated as a simple random sample from the population of size $N_1 + N_2$.

37. (Section 1.2.7) Let $y_1, y_2, \ldots, y_n$ be independent random variables with $y_i \sim (\mu, \sigma_i^2)$. Show that

$$E\left\{ (n-1)^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \right\} = n^{-1} \sum_{i=1}^{n} \sigma_i^2.$$

38. (Section 1.3) Prove:

Result 1.6.1. Let $y_1, y_2, \ldots,$ be a sequence of real numbers. Let $\mathcal{F}_N = (y_1, y_2, \ldots, y_N)$ be a sequence of populations, and let (1.3.20) and

(1.3.21) hold. Let a sequence of samples be selected with probabilities π_i and joint probabilities $\pi_{ij,N}$, where $\pi_{ij,N} \leq \pi_i \pi_j$ for all i and j in $U_N, i \neq j$, and all N. Then

$$V\{N^{-1}(\hat{T}_y - T_y) \mid \mathcal{F}_N\} = O(n_{BN}^{-1}),$$

where n_{BN} is the expected sample size for population N and $\hat{T}_y$ is the Horvitz–Thompson estimator of the total.

39. (Section 1.4) Let a sample of size n be selected in the following way. The first element is selected with probability p_i, where $\Sigma_{i=1}^{N} p_i = 1$. Then $n-1$ elements are selected as a simple random sample from the remaining $N-1$ elements. What is the total probability, π_i, that element i is included in the sample? What is the probability that elements i and j appear in the sample? What is the probability that the n elements $i_1, i_2, \ldots, i_n$ form the sample? See Midzuno (1952).

40. (Section 1.2.8) Assume simple random sampling at each of the two stages of a two-stage sample. Are there population configurations such that $\pi_{(ij)(km)}$ of (1.2.76) is the same for all ij and km, $ij \neq km$?

41. (Section 1.3.1) Consider a sequence of populations $\{\mathcal{F}_N\}$ created as the first N elements of the sequence $\{y_1, y_2, \ldots\}$. Assume that the $\mid y_i \mid$ are bounded and that S_{yN}^2 converges to a positive quantity. Let a systematic sample be selected from the Nth population with a rate of K^{-1} for all N. Is $\bar{y}_n$ design consistent for $\bar{y}_N$? Explain.

42. (Section 1.2.4) Assume that a population of size 10 is generated by the autoregressive model of (1.2.61) with $\rho = 0.9$. Assume that a systematic sample of size 2 is selected and that the selected elements are $A = [i, j]$. Give

$$V\{\bar{y}_2 - \bar{y}_N \mid A = [2, 7]\}$$

and

$$V\{\bar{y}_2 - \bar{y}_N \mid A = [3, 8]\}.$$

Derive the variance of the best linear unbiased predictor of $\bar{y}_N$ for each of the situations $A = [2, 7]$ and $A = [3, 8]$. Give the variance of the predictor.

43. (Section 1.2) Let a population of size $2N$ be divided into two groups of size N. Let a group be selected at random (with probability equal to one-half), and let one unit be selected at random from the selected group. Let a simple random sample of size 2 be selected from the other

group to form a sample of size 3 from the original population. Give the inclusion probabilities and joint inclusion probabilities for this selection scheme.

44. (Section 1.2.5) Assume that a replacement sample of size 2 is selected from a population of size 4 with probability $p_i = 0.25$ at each draw. What is the relative efficiency of estimator (1.2.66) to estimator (1.2.71)?

45. (Section 1.2.1) Let $y_i, i = 1, 2, \ldots, n$, be independent (μ, σ_i^2) random variables and let

$$\bar{y} = n^{-1} \sum_{i=1}^{n} y_i.$$

Show that

$$\hat{V}\{\bar{y}\} = n^{-1}(n-1)^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

is unbiased for $V\{\bar{y}\}$.

46. (Section 1.3.1) Let $\hat{\theta} = \Sigma_{i=1}^{n} w_i y_i$, where $\Sigma_{i=1}^{n} w_i^4 = O(n^{-3})$, $\Sigma_{i=1}^{n} w_i = 1$ for all n, the w_i are fixed, and the $y_i, i = 1, 2, \ldots$, are independent (μ, σ_i^2) random variables with bounded fourth moments. Show that

$$E\{\hat{V}(\hat{\theta})\} = V\{\hat{\theta}\} + O(n^{-2}),$$

where

$$\hat{V}\{\hat{\theta}\} = \left(1 - \sum_{i=1}^{n} w_i^2\right)^{-1} \sum_{i=1}^{n} w_i^2 (y_i - \hat{\theta})^2.$$

47. (Section 1.3) Let $(y_1, y_2, \ldots, y_n)$ be a simple random sample from a population with $y_i > 0$ for all i and finite fourth moment. Let an estimator of the coefficient of variation be

$$\hat{\theta} = \bar{y}^{-1} s_y,$$

where $\theta = \bar{y}_N S_{y,N}$ is the coefficient of variation. Using a Taylor expansion, find the variance of the approximate distribution of $n^{0.5}(\hat{\theta} - \theta)$.

48. (Section 1.2) Let $(y_1, y_2, \ldots, y_N) = \mathbf{y}'$ be a vector of $iid(\mu, \sigma^2)$ random variables. What are the conditional mean and covariance matrix of $\mathbf{y}$ conditional on $\Sigma_{i=1}^{N} y = T$? What are the conditional mean and covariance matrix of $\mathbf{y}$ conditional on (T, S^2), where

$$S^2 = (N-1)^{-1} \sum_{i=1}^{N} (y_i - \bar{y}_N)^2?$$

49. (Section 1.3) To evaluate M_5 of Theorem 1.3.2, show that for a population with zero mean and finite fourth moment,

$$E\{\bar{y}^4\} = n^{-3}[E\{y^4\} + 3(n-1)\sigma^4]$$

and

$$C\{\sum_{i=1}^{n} y_i^2,\ n\bar{y}^2\} = E\{y^4\} - \sigma^4.$$

Hint: See Fuller (1996, p. 241).

50. (Section 1.3) In Exercise 13, procedure (b) consisted of a nonreplacement sample of n integers. Let an estimator of the total of y be

$$\hat{T}_{y,r} = n^{-1}T_m^{-1}\sum_{d=1}^{n} m_d^{-1}y_d,$$

where d is the index for draw, (m_d, y_d) is the (measure of size, y value) obtained on the dth draw, and the vector of totals is

$$(T_m, T_y) = \sum_{i=1}^{N}(m_i, y_i).$$

(a) Show that $\hat{T}_{y,r}$ is design unbiased for T_y and give an expression for $V\{\hat{T}_{y,r} - T_y \mid \mathcal{F}\}$. Give an estimator of $V\{\hat{T}_{y,r} - T_y \mid \mathcal{F}\}$. *Hint*: Let $z_i = m_i^{-1}y_i$ and consider the population composed of m_1 values of $m_1^{-1}y_1, m_2$ values of $m_2^{-1}y_2, \ldots, m_N$ values of $m_N^{-1}y_N$.

(b) Consider a sequence of pairs of real numbers $\{m_i, y_i\}$, where the m_i are positive integers. Assume that:

(i) The m_i are bounded.

(ii) $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} y_i = \mu_y.$

(iii) $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}(y_i - \bar{y}_N)^2 = \sigma_y^2 > 0.$

(iv) $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} \mid y_i - \bar{y}_N \mid^{2+\delta} = K_3 < \infty.$

Show that

$$n^{1/2}(N^{-1}\hat{T}_{y,r} - \bar{y}_N) \xrightarrow{\mathcal{L}} N(0, V_{11}),$$

where

$$V_{11} = \lim_{N\to\infty} nT_m^{-2}V\{\hat{T}_{y,r} - T_y \mid \mathcal{F}_N\}.$$

1.7 APPENDIX 1A: SOME ORDER CONCEPTS

There are exact distributional results available for only a few of the statistics associated with survey sampling. For most, approximations based on large-sample theory are required. Concepts of relative magnitude or *order of magnitude* are useful in deriving those approximations. The following material is from Fuller (1996, Chapter 5). Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be sequences of real numbers, let $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ be sequences of positive real numbers, and let $\{X_n\}_{n=1}^{\infty}$ and $\{Y_n\}_{n=1}^{\infty}$ be sequences of random variables.

Definition 1.7.1. We say that a_n is of smaller order than g_n and write

$$a_n = o(g_n)$$

if

$$\lim_{N \to \infty} g_n^{-1} a_n = 0.$$

Definition 1.7.2. We say that a_n is at most of order g_n and write

$$a_n = O(g_n)$$

if there exists a real number M such that $g_n^{-1} \mid a_n \mid \leq M$ for all n.

Using the definitions of order and the properties of limits, one can prove:

1. If $a_n = o(f_n)$ and $b_n = o(g_n)$, then

$$\begin{aligned} a_n b_n &= o(f_n g_n), \\ \mid a_n \mid^s &= o(f_n^s) \quad \text{for } s > 0, \\ a_n + b_n &= o(\max\{f_n, g_n\}). \end{aligned}$$

2. If $a_n = O(f_n)$ and $b_n = O(g_n)$, then

$$\begin{aligned} a_n b_n &= O(f_n g_n), \\ \mid a_n \mid^s &= O(f_n^s) \quad \text{for } s \geq 0, \\ a_n + b_n &= O(\max\{f_n, g_n\}). \end{aligned}$$

3. If $a_n = o(f_n)$ and $b_n = O(g_n)$, then

$$a_n b_n = o(f_n g_n).$$

The concepts of order for random variables, introduced by Mann and Wald (1943), are closely related to *convergence in probability*.

Definition 1.7.3. The sequence of random variables $\{X_n\}$ converges in probability to the random variable X, written

$$p\lim X_n = X$$

(the *probability limit* of X_n is X), if for every $\epsilon > 0$

$$\lim_{n\to\infty} P\{|\, X_n - X \,| > \epsilon\} = 0.$$

Definition 1.7.4. We say that X_n is of smaller order in probability than g_n and write

$$X_n = o_p(g_n)$$

if

$$p\lim g_n^{-1} X_n = 0.$$

Definition 1.7.5. We say that X_n is at most of order in probability g_n (or bounded in probability by g_n) and write

$$X_n = O_p(g_n)$$

if for every $\epsilon > 0$ there exists a positive real number M_ϵ such that

$$P\{|\, X_n \,| \geq M_\epsilon g_n\} \leq \epsilon$$

for all n.

Analogous definitions hold for vectors.

Definition 1.7.6. If $\mathbf{X}_n$ is a k-dimensional random variable, $\mathbf{X}_n$ is at most of order in probability g_n and we write

$$\mathbf{X}_n = O_p(g_n)$$

if for every $\epsilon > 0$ there exists a positive real number M_ϵ such that

$$P\{|\ X_{jn}\ | \geq M_\epsilon g_n\} \leq \epsilon, \qquad j = 1, 2, \ldots, k,$$

for all n.

Definition 1.7.7. We say that X_n is of smaller order in probability than g_n and write

$$\mathbf{X}_n = o_p(g_n)$$

if for every $\epsilon > 0$ and $\delta > 0$ there exists an N such that for all $n > N$,

$$P\{|\ X_{jn}\ | > \epsilon g_n\} < \delta, \qquad j = 1, 2, \ldots, k.$$

Order operations for sequences of random variables are similar to those for sequences of real numbers; thus:

1. If $X_n = o_p(f_n)$ and $Y_n = o_p(g_n)$, then

$$\begin{aligned} X_n Y_n &= o_p(f_n g_n), \\ |\ X_n\ |^s &= o_p(f_n^s) \quad \text{for } s > 0, \\ X_n + Y_n &= o_p(\max\{f_n, g_n\}). \end{aligned}$$

2. If $X_n = O_p(f_n)$ and $Y_n = O_p(g_n)$, then

$$\begin{aligned} X_n Y_n &= O_p(f_n g_n), \\ |\ X_n\ |^s &= O_p(f_n^s) \quad \text{for } s \geq 0, \\ X_n + Y_n &= O_p(\max\{f_n, g_n\}). \end{aligned}$$

3. If $X_n = o_p(f_n)$ and $Y_n = O_p(g_n)$, then

$$X_n Y_n = o_p(f_n g_n).$$

One of the most useful tools for establishing the order in probability of random variables is *Chebyshev's inequality.*

Theorem 1.7.1. Let $r > 0$ and let X be a random variable such that $E\{|\ X\ |^r\} < \infty$. Then for every $\epsilon > 0$ and finite A,

$$P\{|\ X - A\ | \geq \epsilon\} \leq \frac{E\{|\ X - A\ |^r\}}{\epsilon^r}.$$

It follows from Chebyshev's inequality that any random variable with finite variance is bounded in probability by the square root of its second moment about the origin.

Corollary 1.7.1.1. If $\{X_n\}$ is a sequence of random variables such that

$$E\{X_n^2\} = O(a_n^2),$$

then

$$X_n = O_p(a_n).$$