Some Sampling Designs and Estimation Problems

Hassan Lakkis

Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/mathstat_etds

Part of the Mathematics Commons, and the Statistics and Probability Commons

Recommended Citation
https://digitalcommons.odu.edu/mathstat_etds/98

This Dissertation is brought to you for free and open access by the Mathematics & Statistics at ODU Digital Commons. It has been accepted for inclusion in Mathematics & Statistics Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.
SOME SAMPLING DESIGNS AND ESTIMATION PROBLEMS

by

HASSAN LAKKIS

B.S., Lebanese University, 1986
M.S., Old Dominion University, 1990

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY
Computational and Applied Mathematics

OLD DOMINION UNIVERSITY
August, 1994

Approved by:

__________________________
Ram C. Dahiya (Director)
In the first chapter we review some standard estimators in sampling from a finite population, and some design-based estimators in sampling from a continuous universe.

In concert with the theory initiated by professor Douglas Robson (personal communication) and later presented by Cordy (1993), we consider design-based variance estimation for probability sampling from a continuous and spatially distributed universe. Using this theory in chapter two, the sampling design of one random point from each cell of a translated grid is investigated and the problem of edge effects on estimation is illustrated with examples. Also in chapter four, standard systematic sampling methods from a finite population are reviewed. Then, for systematic samples drawn from a continuous universe, a new approach for investigating the estimators of the parameters of interest is introduced. This new approach can be useful for deriving unbiased variance estimators for many spatial systematic sampling methods and allows for proposing new efficient systematic sampling designs. For these systematic sampling designs, we present the estimator of the population total and the estimator of the variance for a population with one dimension, and we derive in general these estimators for n-dimensional population.
Furthermore in chapter five, a mean-balanced sample of size two from each cell of a translated grid is investigated. Then an unbiased estimator of the population total is presented. Also, explicit formulas for the inclusion density functions are derived.

The problem of sampling from an ordered population of size $N$ is considered in the third chapter. Let the response variable of interest be denoted by $y_1, y_2, ..., y_N$ and the auxiliary variable be denoted by $x_1, x_2, ..., x_N$. Assuming that the values of the $x$ variable are known for all units in the population, a sampling design is developed which accounts for the ordering of the units. Here, the probability for the $i^{th}$ unit to be selected in the sample depends on $x_i$. Based on this design two unbiased estimators of the population total, $T = \sum_{i=1}^{N} y_i$, are presented and exact expressions of the variances of these estimators are derived. Also non-negative unbiased estimators of the variances are provided. This is a new sampling design that does not exist in the literature.
Dedicated

to

my mother
ACKNOWLEDGEMENTS

I would like to thank Dr. Douglas Robson for his help in guiding me through this research. His continuous support and genuine care will never be forgotten. I would also like to thank Dr. Ram Dahiya for his extremely valuable advice and constant encouragement which helped me through this dissertation. I would like to express my appreciation and gratitude to my dissertation committee, Dr. Narasinga Rao Chaganty and Dr. Dayanand Naik for their helpful suggestions and comments.

Many thanks for Dr. Cynthia Jones for introducing me to Dr. Robson and for providing the financial support during my graduate study.

Finally, my heartfelt thanks to my parents and my wife for their love and support.
Contents

1 REVIEW ................................. 1

1.1 Sampling From a Finite Universe ................................................. 1

1.1.1 Sampling Design ............................................................ 1

1.1.2 Inclusion Probabilities ..................................................... 3

1.1.3 Results Related to the Membership Indicators ...................... 5

1.1.4 Horvitz-Thompson Estimator and its Statistical Properties . 6

1.2 Sampling from a Continuous Universe ..................................... 9

2 VARIANCE ESTIMATION IN STRATIFIED SPATIAL SAM-
PLING AND THE EDGE EFFECT PROBLEM ............................... 13

2.1 Introduction ............................................................................ 13

2.2 Sampling Universe .............................................................. 14

2.3 Uniform Random Sampling .................................................. 19

2.4 Stratified Random Sampling ................................................ 19

2.5 Three-step Stratified Sampling .............................................. 20

2.5.1 Calculation of the Inclusion Density Functions ................... 24

2.6 Edge Effect on the Estimators .............................................. 31

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
3 SAMPLING FROM AN ORDERED POPULATION WITH A GIVEN AUXILIARY VARIABLE

3.1 Introduction ................................................................. 42
3.2 Sampling Procedure ...................................................... 45
3.3 Numerical Example ...................................................... 46
3.4 Estimation ................................................................. 48
  3.4.1 Calculations of the Inclusion Probability Functions ......... 51
  3.4.2 Calculations of the Covariance Function ....................... 67
3.5 Estimation Using Sampling from a Continuous Universe ....... 72
3.6 Comparison of $\hat{T}_1$ and $\hat{T}_2$ ................................. 74
3.7 Comparison of the Variance Estimators of $\hat{T}_1$ ................. 74
3.8 Comparison of Fixed Order Stratified Sampling Procedure to
  Other Sampling Procedures ............................................... 76
  3.8.1 Group I Simulations ............................................... 78
  3.8.2 Group II Simulations ............................................... 83
  3.8.3 Group III Simulations ............................................... 87
### 4 VARIANCE ESTIMATION IN SYSTEMATIC SPATIAL SAMPLING

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>91</td>
</tr>
<tr>
<td>4.2</td>
<td>Systematic Sampling from a Finite Population</td>
<td>92</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Systematic Sampling with one Random Start from a Finite Population</td>
<td>92</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Systematic Sampling with Multiple Random Starts from a Finite Population</td>
<td>96</td>
</tr>
<tr>
<td>4.3</td>
<td>Systematic Sampling from a Continuous Universe</td>
<td>99</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Systematic Sampling with one Random Start from a Continuous Universe</td>
<td>101</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Systematic Spatial Sampling with Multiple Random Starts from a Spatial Universe</td>
<td>105</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Systematic Sampling with Multiple Restricted Random Starts from a Spatial Universe</td>
<td>107</td>
</tr>
</tbody>
</table>

### 5 VARIANCE ESTIMATION IN A MEAN BALANCED SPATIAL SAMPLING DESIGN

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>114</td>
</tr>
<tr>
<td>5.2</td>
<td>The Mean Balanced Design from a One-dimensional Space</td>
<td>116</td>
</tr>
<tr>
<td>5.3</td>
<td>The Mean Balanced Design from an n-dimensional Space</td>
<td>125</td>
</tr>
</tbody>
</table>
List of Figures

2.1 Example of $D$ with a known boundary and its corresponding $U$ in a one-dimensional space ($U = D$).......................... 15

2.2 Example of region $D$ in a two-dimensional space and its corresponding $U$ when the boundary of $D$ is known prior to sampling........ 16

2.3 Example of region $D$ in a two-dimensional space and its corresponding $U$ when the boundary of $D$ is not known prior to sampling... 17

2.4 ($tss$) from a two-dimensional grid before the sampled points outside $U$ are reduced............................................. 21

2.5 ($tss$) from a two-dimensional grid after the coordinates of the sample points 1-3, that fall outside $U$ in Figure 2.4, are reduced by modulus $M_1$ and $M_2$ respectively. The resulting points are denoted by 1-3 in this figure............................................. 21

2.6 The domain $D$ is the circle with center at (2,2) and radius 1, and its corresponding $U_C$ as explained in (i) of Example 2............. 37

2.7 The domain $D$ is the circle with center at (2,2) and radius 1, and its corresponding universe $U$ as explained in (ii) of Example 2........ 39
2.8 Example of sampling from a subpopulation $U$, where the population
is the circle. ................................................. 41

4.1 Example of $U$ and $U'$ in a two-dimensional space where $U'$ contains
the three cells in the lower left of the first row of $U$. .................. 110

4.2 Example of $U$ and $U'$ in a two-dimensional space where $U$ is the large
rectangle and $U'$ is the rectangle in the lower left of $U$ with three of
its vertices are at $2L_2$, $O$ and $3L_1$ ($n = n' = 2$). .................. 112

5.1 Example of (mbs) in a two-dimensional space before reducing the
sample points to fall inside $U$. ................................. 115
List of Tables

2.1 Comparison between the two methods of estimation explained in (i) and (ii) of Example 1 ........................ 35

2.2 Comparison between the two methods of estimation explained in (i) and (ii) of Example 2 .................. 40

3.1 Selection of $n = 4$ random arcs from a population of $N = 8$ units .......................... 47

3.2 Empirical average of the sampling variances of the estimators of the population totals and the efficiencies (in brackets) as explained in Group I Simulations .......................... 81

3.3 Empirical average of the sampling variances of the estimators of the population totals and the efficiencies (in brackets) as explained in Group II Simulations .............................................. 85

3.4 Empirical average of the sampling variances of the estimators of the population totals and the efficiencies (in brackets) as explained in Group III Simulations .............................................. 89
3.5 Empirical average of the sampling variances of the estimators of the population totals and the efficiencies (in brackets) as explained in Group III Simulations. The units are arranged such that $z_i$ are in peak ordering.
Chapter 1

REVIEW

In this chapter, we review the literature of sampling theory. First section deals with sampling from a finite universe while the second section deals with sampling from a continuous universe.

1.1 Sampling From a Finite Universe

In this section we present the basic theory of sampling design and inference in sampling from finite populations. Let \( U = \{u_1, u_2, \ldots, u_N\} \) denote a finite population of \( N \) distinct and identifiable units. We refer to \( N \), generally assumed to be known, as the size of the finite population or finite universe \( U \). The elements of \( U \) namely \( u_1, u_2, \ldots, u_N \) are known as the sampling units. Sometimes we refer to the sampling units simply by their indices. Thus, we may talk about \( u_i \) as \( i \) and represent the population as \( U = \{1, 2, \ldots, N\} \).

1.1.1 Sampling Design

Given a sample selection scheme, we denote the probability of selecting a specified sample \( s \) by \( p(s) \). In other words, we assume that there is a function \( p(.) \)
such that $p(s)$ gives the probability of selecting $s$ under the scheme in use. The function $p(.)$ will be called the sampling design. We will regard any sample $s$ as the outcome of a set-valued random variable $S$, whose probability distribution is specified by the function $p(.)$. Let $S$ be the set of all samples, $s$ (thus $S$ is the set of $2^N$ subsets of $U$, if we include the empty set as well as $U$ itself). Then we have

$$Pr(S = s) = p(s)$$

for any $s \in S$. Because $p(s)$ is a probability distribution on $S$, we have

$$p(s) \geq 0 \quad \text{for all } s \in S$$

and

$$\sum_{s \in S} p(s) = 1.$$ 

Note that many of the $2^N$ sets contained in $S$ may in fact have a zero probability. Let $S_0$ be the set of all possible samples, that is, the set of all $s \in S$ with $p(s) > 0$. The elements of $S_0$ are the only ones that can be drawn for a specified design. The sample size will be denoted by $n$ and it can be random or fixed. In this review we consider the sample $s$ to be a (unordered) subset of the universe. However, many sampling designs produce samples of ordered elements. The information about the orders of the selected elements can be dropped and a selected sample can be reduced to a (unordered) subset of elements of the universe. The interested reader about this subject is referred to Cassel, Sarndal, and Wretman (1977).
1.1.2 Inclusion Probabilities

An interesting feature of a finite population of $N$ elements is that the elements can be given different probabilities of inclusion in the sample. The sampling statistician often takes advantage of the identifiability of the elements by deliberately attaching different inclusion probabilities to the various elements. This is one way to obtain more accurate estimators. Suppose that a certain sampling design has been fixed. That is, $p(s)$, the probability of selecting $s$, has a given mathematical form. The inclusion of a given element $k$ in a sample is a random event indicated by the random variable $I_k$, defined as

$$I_k = \begin{cases} 
1 & \text{if } k \in S \\
0 & \text{if not.}
\end{cases}$$

Note that $I_k = I_k(S)$ is a function of the random variable $S$. We call $I_k$ the sample membership indicator of element $k$.

The probability that element $k$ will be included in the sample, denoted by $\pi_k$, is obtained from the given design $p(.)$ as follows:

$$\pi_k = Pr(k \in S) = Pr(I_k = 1) = \sum_{s \ni k} p(s).$$

Here, $s \ni k$ denotes that the sum is over those samples $s$ that contain the given $k$.

The probability that both of the elements $k$ and $l$ will be included is denoted by $\pi_{kl}$ and is obtained from the given $p(.)$ as follows:

$$\pi_{kl} = Pr(k&l \in S) = Pr(I_k I_l = 1) = \sum_{s \ni k&l} p(s).$$

3

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Note that \( \pi_{kl} = \pi_{lk} \) for all \( k, l \). Also when \( k = l \) we get

\[
\pi_{kk} = Pr(I_k^2 = 1) = Pr(I_k = 1) = \pi_k.
\]

That is \( \pi_{kk} \) will be interpreted as identical to \( \pi_k \), for \( k = 1, \ldots, N \).

For a given design \( p(.) \) there are \( N \) associated quantities

\[
\pi_1, \ldots, \pi_k, \ldots, \pi_N.
\]

They constitute the set of first-order inclusion probabilities. Moreover, with \( p(.) \) there are \( N(N - 1)/2 \) associated quantities

\[
\pi_{12}, \pi_{13}, \ldots, \pi_{kl}, \ldots, \pi_{N-1,N}
\]

which are called the set of second-order inclusion probabilities.

A sampling design is often chosen to yield certain desired first- and second-order inclusion probabilities. Sometimes even though \( p(.) \) may be complicated, we can determine expected value and variance of certain estimators from the knowledge of the \( \pi_k \) and the \( \pi_{kl} \) alone.

A sampling design is called a probability sampling design if

\[
\pi_k > 0 \quad \text{for all} \quad k \in U.
\]

This requirement ensures that every element has a chance to appear in the sample. A sample \( s \) realized by such a design is called a probability sample. A sampling design is said to be measurable if

\[
\pi_{kl} > 0 \quad \text{for all} \quad k \neq l \in U.
\]
A measurable design allows the calculations of a valid variance estimate and a valid confidence interval based on the observed survey data.

1.1.3 Results Related to the Membership Indicators

The design based estimators are expressed as functions of the membership indicators which are defined previously. It is therefore important to describe the basic properties of the statistics $I_k = I_k(S)$, for $k = 1, \ldots, N$, as in the following lemma.

**Lemma 1.1** For an arbitrary sampling design $p(s)$, and for $k, l = 1, \ldots, N$,

\[
E(I_k) = \pi_k \quad \text{(1.1)}
\]

\[
Var(I_k) = \pi_k(1 - \pi_k) \quad \text{(1.2)}
\]

\[
Cov(I_k, I_l) = \pi_{kl} - \pi_k \pi_l. \quad \text{(1.3)}
\]

**Proof:** Lemma follows from the fact that $I_k$'s are correlated Bernoulli random variables with $E(I_k I_l) = \pi_{kl}$. □

Depending on the design, the $Cov(I_k, I_l)$ can be zero, positive, or negative.

The sample size $n$ can be expressed in terms of the indicators $I_k$ as

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

The first two moments of the statistic $n$ are

\[
E(n) = \sum_{k=1}^{N} \pi_k
\]
and

\[
Var(n) = \sum_{k=1}^{N} \pi_k (1 - \pi_k) + 2 \sum_{k=1}^{N} \sum_{l>k} (\pi_{kl} - \pi_k \pi_l) \\
= \sum_{k=1}^{N} \pi_k - (\sum_{k=1}^{N} \pi_k)^2 + 2 \sum_{k=1}^{N} \sum_{l>k} \pi_{kl}. \quad (1.4)
\]

When the sample size is fixed, we have the following

\[
\sum_{k=1}^{N} \pi_k = n \quad (1.5)
\]

\[
\sum_{l \neq k}^{N} \pi_{kl} = (n - 1) \pi_k \quad (1.6)
\]

\[
\sum_{k=1}^{N} \sum_{l > k}^{N} \pi_{kl} = \frac{1}{2} n(n - 1). \quad (1.7)
\]

### 1.1.4 Horvitz-Thompson Estimator and its Statistical Properties

If we have only one study variable, \(y\), we can think of the population parameter, \(\theta\), as a function of \(y_1, \ldots, y_N\), the \(N\) values of \(y\). Thus \(\theta = \theta(y_1, \ldots, y_N)\).

In sampling theory, frequently the population total is a parameter of interest, that is

\[
T = \sum_{k=1}^{N} y_k.
\]

The Horvitz-Thompson estimator of the population total \(T\) is given by

\[
\hat{T} = \sum_{k=1}^{N} \frac{y_k}{\pi_k} f_k. \quad (1.8)
\]
Since $E(I_k) = \pi_k$ and $\pi_k > 0$ for all $k \in U$, it follows that $\hat{T}$ is unbiased for $T$.

The random elements in (1.8) are expressed by the indicators $I_1, \ldots, I_N$ and the values $y_k$ and $\pi_k$ are fixed constants.

**Lemma 1.2** (i) The variance of $\hat{T}$ is given by

$$\text{Var}(\hat{T}) = \sum_{k}^N \frac{(1 - \pi_k)}{\pi_k} y_k^2 + 2 \sum_{k}^N \sum_{l>k} (\pi_{kl} - \pi_k \pi_l) \frac{y_k y_l}{\pi_k \pi_l}. \quad (1.9)$$

(ii) An unbiased estimator of $\text{Var}(\hat{T})$ is given by

$$\hat{\text{Var}}_{HT}(\hat{T}) = \sum_{k}^N I_k \frac{(1 - \pi_k)}{\pi_k^2} y_k^2 + 2 \sum_{k}^N \sum_{l>k} I_k I_l \frac{(\pi_{kl} - \pi_k \pi_l)}{\pi_{kl}} \frac{y_k y_l}{\pi_k \pi_l}. \quad (1.10)$$

*Proof:* Since $\hat{T} = \sum_{k=1}^N \frac{y_k}{\pi_k} I_k$ where $\frac{y_k}{\pi_k}$ are treated as constants and the $I_k$ are the random variables, the variance of $\hat{T}$ can be written as

$$\text{Var}(\hat{T}) = \sum_{k}^N \frac{y_k^2}{\pi_k^2} \text{Var}(I_k) + 2 \sum_{k}^N \sum_{l>k} \text{Cov}(I_k, I_l) \frac{y_k y_l}{\pi_k \pi_l}.$$  

We get (1.10), on using (1.2) and (1.3).

Now (ii) follows from the fact that $E(I_k) = \pi_k$ and $E(I_k I_l) = \pi_{kl}$.

The variance estimator (1.10) is due to Horvitz and Thompson (1952). □

**Lemma 1.3** If the design has fixed sample size, then the variance of $\hat{T}$ can be written as

$$\text{Var}(\hat{T}) = - \sum_{k=1}^N \sum_{l>k} (\pi_{kl} - \pi_k \pi_l) \left( \frac{y_k}{\pi_k} - \frac{y_l}{\pi_l} \right)^2. \quad (1.11)$$

An unbiased estimator of $\text{Var}(\hat{T})$ is given by
\[ \hat{V}_{YG}(\hat{T}) = - \sum_{k=1}^{N} \sum_{l> k}^{N} I_{k} I_{l} \left( \frac{\pi_{kl} - \pi_{k} \pi_{l}}{\pi_{kl}} \right) \left( \frac{y_{k} - y_{l}}{\pi_{k} - \pi_{l}} \right)^{2}, \]  

(1.12)

provided that \( \pi_{kl} > 0 \) for all \( k \neq l \in U \). The variance estimator formula (1.12) is due to Yates and Grundy (1953).

**Proof:** By using (1.5) and (1.6), we obtain

\[ \sum_{l \neq k}^{N} (\pi_{kl} - \pi_{k} \pi_{l}) = (n - 1) \pi_{k} - \pi_{k} (n - \pi_{k}) = -\pi_{k} (1 - \pi_{k}). \]

Substituting for \((1 - \pi_{k})\) in the first term of (1.9), gives

\[ \sum_{k}^{N} \frac{(1 - \pi_{k})}{\pi_{k}} y_{k}^{2} = \sum_{k \in U} \sum_{l \neq k}^{N} (\pi_{kl} - \pi_{k} \pi_{l}) \left( \frac{y_{k}}{\pi_{k}} \right)^{2} \]

\[ = \sum_{k}^{N} \sum_{l \neq k}^{N} (\pi_{kl} - \pi_{k} \pi_{l}) \left[ \left( \frac{y_{k}}{\pi_{k}} \right)^{2} + \left( \frac{y_{l}}{\pi_{l}} \right)^{2} \right]. \]

Hence

\[ \text{Var}(\hat{T}) = - \sum_{k=1}^{N} \sum_{l \neq k}^{N} (\pi_{kl} - \pi_{k} \pi_{l}) \left( \frac{y_{k}}{\pi_{k}} - \frac{y_{l}}{\pi_{l}} \right)^{2}. \]

Now, by using this form of the variance, it can be easily seen that \( \hat{V}_{YG}(\hat{T}) \) is unbiased estimator of the variance of \( \hat{T}. \) □

The Yates-Grundy variance estimator is usually preferred over the Horvitz-Thompson estimator. As described by Cassel, Sarndal, and Wretman (1977, pg. 166), one of the primary reasons for this is that for many sampling designs the Yates-Grundy estimator is always non-negative, while the Horvitz-Thompson variance estimator can take negative values. However, for some designs, neither of these estimators may be satisfactory because of their high variability.
1.2 Sampling from a Continuous Universe

In this section we present the results from Cordy (1993) concerning sampling from a continuous universe. This section is an extension of sampling from a finite universe. All the formulas shall be presented without derivations.

Suppose there is an $n$-dimensional region, $D$, of interest, and at each point $x \in D$ it is possible to measure a response $z(x)$. Assume $D$ is an open bounded subset of $\mathbb{R}^n$, and that $z$ is a bounded integrable function on $D$. The integral of $z$ over $D$ is the parameter of interest, that is

$$T = \int_D z(x) \, dx.$$

We consider estimating $T$ and deriving an estimate of the variance of the estimator of $T$ based on a sample selected from $D$. Many of the common methods of probability sampling produce samples of variable size. To overcome this difficulty, we can use some regular form which allows consideration of sampling design with fixed sample size. That is, consider the set $U$, referred to as the universe, which is a subset of $\mathbb{R}^n$ that contains $D$ and may extend beyond the boundaries of $D$ in all directions. An efficient selection of the universe $U$ is to be as small as possible such that $D$ is a subset of it. The requirement that $U$ must satisfy depends on the particular design under consideration. Extend the function $z$ to all of $U$ by defining $z(x) = 0$ for each $x \notin D$. Then the integral of $z$ over $U$ is the same as the integral of $z$ over $D$. The results which follows will be developed in terms of $U$.

The procedure for selecting a probability sample from $U$ will be considered as obtaining the realizations $x_1, x_2, \ldots, x_k$ of the random vectors $X_1, X_2, \ldots, X_k$
respectively, where the range of each of these random vectors is a subset of $U$. We restrict our attention to the case where the distribution of $(X_1, X_2, \ldots, X_k)$ has a density function $f$. This density function is determined by the sampling design.

For $i = 1, 2, \ldots, k$, let $f_i$ denote the marginal density of $X_i$. For $x \in U$ the inclusion density function is defined as

$$\pi(x) = \sum_{i=1}^{k} f_i(x).$$

Next, we define the pairwise inclusion density function on $U \times U$. For $i = 1, 2, \ldots, k$ and $j \neq i$, let $f_{ij}$ denote the joint marginal density of $X_i$ and $X_j$. For $x, x' \in U$, we define the pairwise inclusion density function by

$$\pi(x, x') = \sum_{i=1}^{k} \sum_{j \neq i} f_{ij}(x, x').$$

These inclusion density functions play a role similar to that played by the inclusion probability functions in the finite universe case. That is, they contain the information about the sampling design which is necessary to estimate means, totals, and variances.

In this section we define the estimators in terms of the sample response variable and the sample inclusion density functions without using indicator variables as in the previous section. The Horvitz-Thompson estimator of $T$ is defined by

$$\hat{T} = \sum_{i=1}^{k} \frac{z(x_i)}{\pi(x_i)}.$$  \hspace{1cm} (1.13)

This is an unbiased estimator of $T$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
**Lemma 1.4** Using the conceptual device which is introduced above, giving a fixed sample size design, the variance of $\hat{T}$ is given by

$$\text{Var}(\hat{T}) = \frac{1}{2} \int_U \int_U (\pi(x) \pi(x') - \pi(x, x')) \left( \frac{z(x)}{\pi(x)} - \frac{z(x')}{\pi(x')} \right)^2 \, dx \, dx'. \quad (1.14)$$

The Yates-Grundy variance estimator is given by

$$v_1 = \frac{1}{2} \sum_{i=1}^{k} \sum_{j \neq i} \left( \frac{\pi(x_i) \pi(x_j) - \pi(x_i, x_j)}{\pi(x_i, x_j)} \right) \left( \frac{z(x_i)}{\pi(x_i)} - \frac{z(x_j)}{\pi(x_j)} \right)^2. \quad \Box \quad (1.15)$$

The variance of the Horvitz-Thompson estimator can be re-expressed as

$$V(\hat{T}) = \int_U \frac{(z(x))^2}{\pi(x)} \, dx - \int_U \int_U \frac{\pi(x) \pi(x') - \pi(x, x')}{\pi(x) \pi(x')} \, dx \, dx' \, z(x) z(x') \, dx \, dx'$$

$$= \int_D \frac{(z(x))^2}{\pi(x)} \, dx - \int_D \int_D \frac{\pi(x) \pi(x') - \pi(x, x')}{\pi(x) \pi(x')} \, dx \, dx' \, z(x) z(x') \, dx \, dx'. \quad (1.16)$$

Note that since $z(x) = 0$ for $x \notin D$, this formula can be expressed in terms of integrals over $D$, rather than $U$, without changing its value. But this is not the case in equation (1.14).

**Lemma 1.5** The latest form of the variance gives rise to the Horvitz-Thompson variance estimator given by

$$v_2 = \sum_{i=1}^{k} \left( \frac{(z(x_i))^2}{\pi(x_i)} \right) - \sum_{i=1}^{k} \sum_{j \neq i} \left( \frac{\pi(x_i) \pi(x_j) - \pi(x_i, x_j)}{\pi(x_i) \pi(x_j) \pi(x_i, x_j)} \right) z(x_i) z(x_j). \quad \Box \quad (1.17)$$
The Yates-Grundy variance estimator is expected to be more stable than the Horvitz-Thompson estimator, because it uses extra information from the sampled data points provided that \( \pi(x, x') > 0 \) almost everywhere \( x, x' \in U \). Both of these variance estimators are design unbiased, that is, their expectation with respect to the sampling design is equal to the true variance.
Chapter 2

VARIANCE ESTIMATION IN STRATIFIED SPATIAL SAMPLING AND THE EDGE EFFECT PROBLEM

2.1 Introduction

In this chapter we consider the issue of variance estimation for probability sampling from a continuous and spatially distributed universe. We illustrate sampling from a continuous universe by the following examples. Suppose a study variable changes with respect to time over some time interval, then the problem that could be raised is to estimate the population total over the given interval and also to estimate the variance of the estimator. Such a problem is an example of sampling from a continuous universe in one dimension. Sampling in a plane is a well known problem in sampling surveys which is an example of sampling from a continuous universe in two dimensions. The procedure of sampling certain water points from a lake, for instance, or sampling in a plane area over time are examples of sampling
from a three-dimensional population. Also sampling from a four-dimensional population is exhibited by considering sampling water points from a lake over a fixed interval of time.

In some sampling designs, the relevant information for variance estimation can be expressed in terms of the relative positions of the sample locations. In such cases, it is possible to express the sampling variance in terms of the inclusion density functions which are functions of the relative positions of the sampled points.

Suppose there is a domain, \( D \), of interest and that at each point \( u \in D \), it is possible to measure a study variable \( z(u) \). The domain \( D \) can be a subset of \( \mathbb{R}^n \) where \( n \geq 1 \). We are interested in estimating the population parameter

\[
T = \int_D z(u) \, du
\]

and deriving an estimator of the variance of the estimator of \( T \) based on a probability sample from \( D \).

Before presenting any sampling method, we introduce a universe \( U \) which has a regular shape such as a rectangle in two dimensions and contains the domain \( D \) of the study variable of interest. The sampling is to be performed from \( U \) and further explanation about this universe is in the following section.

### 2.2 Sampling Universe

Frequently the population boundary is not known prior to sampling or the population does not have a simple regular shape such as a rectangle. Also, most
of the sampling methods applied on such a population produce a variable sample size. To overcome this difficulty, we consider a set $U$ such that $D$ is a subset of $U$. A point in $U$ may be represented by $x = (x_1, x_2, ..., x_n)$. Extend the function $z$ to include all of $U$ by defining $z(x) = 0$ if $x \notin D$ (refer to section 2.6 of this chapter for investigation of the effects of this extension). $U$ is referred to in this chapter as the sampling universe. However, if the boundary of the domain $D$ are known prior to sampling and a sampling method can lead to a fixed sample size, then a more efficient estimator of the population total might be obtained. In the remaining part of this section we explain the relationship between $U$ and $D$ and also explain how to lay a grid over $U$ which shall be used to sample from within its cells.

(i) When the domain $D$ has one dimension, then $U$ will be considered as an interval containing $D$, with length $M_1$. We position $U$ on the real number axis such that the left extreme of $U$ is at the origin of the axis. Superimpose a one-dimensional grid on $U$ with spacing $l_1 = \frac{M_1}{m_1}$, where $m_1$ is the number of strata in the universe. $M_1$ is selected such that each of the $m_1$ strata intersect the domain $D$. If $D$ is an interval with its boundary known before sampling, then $U$ should be

\[ \text{Figure 2.1: Example of } D \text{ with a known boundary and its corresponding } U \text{ in a one-dimensional space } (U = D). \]
Figure 2.2: Example of region $D$ in a two-dimensional space and its corresponding $U$ when the boundary of $D$ is known prior to sampling.

selected to be equal to $D$ (Figure 2.1), and $M_1$ becomes the length of the interval $D$.

(ii) When the domain $D$ is two-dimensional, then $U$ will take the shape of a rectangle with width $M_1$ and length $M_2$. We place on $U$ a two-dimensional cartesian space with axis denoted by $e_1$ and $e_2$. The lower side of $U$ is positioned over the $e_1$ axis and its left side is positioned over the $e_2$ axis. Divide this rectangle by a two-dimensional grid into $m_1 m_2$ rectangles each of area $l_1 l_2$ to yield $m_1$ columns and $m_2$ rows. Each rectangle has a width $l_1$ and a length $l_2$. So $M_1 = m_1 l_1$ and $M_2 = m_2 l_2$. $M_1$ and $M_2$ are selected such that each of the $m_1$ columns and $m_2$
Figure 2.3: Example of region $D$ in a two-dimensional space and its corresponding $U$ when the boundary of $D$ is not known prior to sampling.
rows intersect the domain $D$. If we select $l_1 = l_2$, then each grid cell will have the form of a square. This is a desired division of the universe $U$ because a probability sample selected from within the grid cells can provide a good spatial coverage of the domain $D$. Figure 2.2 and Figure 2.3 presents two different examples of $D$ and $U$ in a two-dimensional space.

(iii) When the domain $D$ has $n$ dimensions, then an $n$-dimensional space is superimposed on $U$ with its axis denoted by $e_i$, $i = 1, 2, \ldots, n$, such that $U$ can be defined as

$$U = \{x = (x_1, x_2, \ldots, x_n) : 0 < x_i < M_i; \ i = 1, 2, \ldots, n\}. \quad (2.2)$$

Divide this universe $U$ by an $n$-dimensional grid into $m_1 m_2 \ldots m_n$ cells. Each grid cell has Lebesgue measure $\mu = l_1 l_2 \ldots l_n$. In the case of two dimensions, it is the area of a single cell of $U$, and in the case of three dimensions it is the volume of a single cell of $U$. Each cell is of length $l_i$ with respect to $e_i$-axis. The $M_i$'s can be selected as small as possible such that $D$ is a subset of $U$. Also it could be practical to fix $l_i = l_j$ for $1 \leq i, j \leq n$.

Some attractive sampling methods, for sampling from $U$, having good spatial coverage of the populations, are considered. For each sampling method, the design-based variance estimators are investigated.

A paper by Cordy and Thompson (1994) which considers design-based variance estimation in sampling from a two-dimensional space, is relevant here.
2.3 Uniform Random Sampling

A uniform random sample (urs) of size $k$ from $U$ is a realization of $k$ independent continuous random variables which have a uniform distribution on $U$. This (urs) design is not practical because it frequently gives samples which do not provide satisfactory spatial coverage of the populations.

2.4 Stratified Random Sampling

From each cell of the grid over $U$ a (urs) of size $k$ is selected. The inclusion density functions are given by

$$\pi(x) = \frac{k}{\mu} \text{ for every } x \in U,$$  \hspace{1cm} (2.3)

and for $x \neq x' \in U$,

$$\pi(x,x') = \begin{cases} \frac{k(k-1)}{(\mu)^2} & \text{if } x \text{ and } x' \text{ are in the same stratum} \\ \frac{k^2}{\mu^2} & \text{if } x \text{ and } x' \text{ are in different strata}, \end{cases}$$ \hspace{1cm} (2.4)

where $\mu = l_1l_2\ldots l_n$. The design for which $k = 1$ provides the most even spatial coverage. But the variance is not unbiasedly estimable since $\pi(x,x') = 0$ when $x$ and $x'$ are in the same stratum. The following section provides a similar sampling design for which an unbiased variance estimation is possible for the case of $k = 1$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
2.5 Three-step Stratified Sampling

A three-step stratified sample is termed \(tss\), and is obtained by performing the following three steps independently:

(i) randomly translate the grid over \(U\), while keeping \(U\) in the same position;

(ii) independently select a point at random from each cell of the translated grid (Figure 2.4); and

(iii) if \(x = (x_1, x_2, ..., x_n)\) is a selected point in step (ii), then \(x\) is included in the sample with its \(x_i\) coordinate being suitably reduced by \(\text{mod}(M_i)\) for \(i = 1, 2, ..., n\) (Figure 2.5).

If \(x = (x_1, x_2, ..., x_n)\) and \(x' = (x'_1, x'_2, ..., x'_n)\) are two points in \(U\), then we view the distance between \(x_i\) and \(x'_i\) in a circular manner, for \(i = 1, 2, ..., n\). In other words the distance between \(x_i\) and \(x'_i\) is given by

\[
\delta(x_i, x'_i) = \min\{|x_i - x'_i|, M_i - |x_i - x'_i|\}.
\]

This new argument of using circular sampling in multidimensional domain has not been used before in sampling theory. For instance, Cordy and Thompson (1994) considered the same sampling design with \(D\) being in two dimensions. They suggested using a universe \(U\) that is larger than \(D\) by at least one grid cell from each side, but they did not consider circular sampling. Circular sampling argument can affect the variance of the population total estimator and the estimator of the
Figure 2.4: (tss) from a two-dimensional grid before the sampled points outside $U$ are reduced.

Figure 2.5: (tss) from a two-dimensional grid after the coordinates of the sample points 1-3, that fall outside $U$ in Figure 2.4, are reduced by modulus $M_1$ and $M_2$ respectively. The resulting points are denoted by 1-3 in this figure.

variance through the distance which is defined above. Also by using circular sampling, the universe $U$, defined in this chapter, can be as small as possible as long as it contains $D$. In section 2.6 of this chapter we compare the estimators by using Cordy's argument and the new argument presented in this chapter.

An analogous design for finite population sampling in one dimension was presented by Fuller (1970). We now describe the sampling method, introduced in

21

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
the three steps above, from the universe $U$ in one dimension, two dimensions, and $n$ dimensions.

When $U$ is in one dimension, select a uniform random number, $w_1$, from the interval $(0, l_1)$. Let $W_1$ be a random variable for which $w_1$ is its realized value. Translate the grid in such a way that the grid edge originally at zero is changed to $w_1$. Now, select a uniform random number, $x_{1i_1}$, from the interval $(w_1 + (i_1 - 1)l_1, w_1 + i_1 l_1)$, for $i_1 = 1, \ldots, m_1$. Let $X_{1i_1}$ be a random variable associated with the realized value $x_{1i_1}$, and reduce $x_{1i_1}$ by mod$(M_1)$. Now, the selected sample is given by the set of points

$$s = \{(x_{1i_1}) : i_1 = 1, 2, \ldots, m_1\}.$$  

Note that $w_1$ is used as a reference point without including it in the sample.

When $U$ is in two dimensions, select a uniform random number, $w_j$, from the interval $(0, l_j)$ with respect to the $e_j$ axis, for $j = 1, 2$. Let $W_j$ be the random variable for which $w_j$ is its realized value. Translate the grid in such a way that the grid point originally at $(0, 0)$ is changed to $(w_1, w_2)$. Now, select a uniform random number, $x_{ji_2}$, from the interval $(w_j + (i_j - 1)l_j, w_j + i_j l_j)$ for $j = 1, 2$ and $i_j = 1, \ldots, m_j$. Let $X=(X_{1i_1}, X_{2i_2})$ be the random vector associated with the realized point $x=(x_{1i_1}, x_{2i_2})$ (Figure 2.4). Now, reduce each of the coordinates, $x_{1i_1}$ and $x_{2i_2}$, of $x$ by mod$(M_1)$ and mod$(M_2)$ respectively (Figure 2.5). The selected sample is given by the set of elements

$$s = \{(x_{1i_1}, x_{2i_2}) : i_1 = 1, 2, \ldots, m_1; \ i_2 = 1, 2, \ldots, m_2\}.$$
In general, when \( U \) is in \( n \) dimensions, select a uniform random number, \( w_j \), from the interval \((0, l_j)\) with respect to the \( e_j \) axis, for \( j = 1, 2, ..., n \). Let \( W_j \) be the random variable for which \( w_j \) is its realized value. Translate the grid in such a way that the grid point originally at \((0,0, ..., 0)\) is changed to \((w_1, w_2, ..., w_n)\). Now, select a uniform random number, \( x_{ji} \), from the interval \((w_j + (i_j - 1)l_j, w_j + i_jl_j)\) for \( i_j = 1, ..., m_j \) and \( j = 1, 2, ..., n \). Let \( (X_{1i_1}, X_{2i_2}, ..., X_{ni_n}) \) be the random vector associated with the realized \((x_{1i_1}, x_{2i_2}, ..., x_{ni_n})\). Now, reduce each of the coordinates, \( x_{1i_1}, x_{2i_2}, ..., x_{ni_n} \) of \( x \) by \( \text{mod}(M_1) \), \( \text{mod}(M_2) \), ..., \( \text{mod}(M_n) \) respectively. The selected sample is given by the points in the set

\[
s = \{(x_{1i_1}, x_{2i_2}, ..., x_{ni_n}) : i_j = 1, 2, ..., m_j; j = 1, 2, ..., n\}. \tag{2.5}
\]

Before moving further into this section we introduce a lemma which can be helpful for the calculation of the inclusion density functions in many spatial sampling designs. Also in this chapter we shall use \( h = (h_1, h_2, ..., h_n) \), \( h' = (h'_1, h'_2, ..., h'_n) \) and \( h \to 0 \) shall mean \( h_i \to 0 \) for \( i = 1, 2, ..., n \).

**Lemma 2.1** Let \((x, y)\) be the realized values of \((X, Y)\) where each of \(X\) and \(Y\) is a vector of continuous random variables. Also, let \( W \) be a vector of continuous random variables and let \( A \) be a borel set such that \( \Pr(W \in A) > 0 \). If the conditional density of \((X, Y | W = w)\) at \((x, y)\), denoted by \( f(x, y | w) \), does not depend on \( w \) when \( w \in A \), then \( f(x, y | w) \) can be written as

\[
f(x, y | w) = \lim_{h, h' \to 0} \frac{Pr(x - h < X < x + h, y - h' < Y < y + h' | W = w)}{2^n h_1 h_2 ... h_n h'_1 h'_2 ... h'_n} \tag{2.6}
\]
where the limit is taken as $h \to 0$ and $h' \to 0$.

The proof of this Lemma is trivial by using the rules of probability and its application is in the later part of this chapter.

### 2.5.1 Calculation of the Inclusion Density Functions

Suppose the universe $U$ is of $n$ dimensions. Define $\Omega$ to be the sample space of $W$, that is

$$\Omega = \{w = (w_1, w_2, ..., w_n) : 0 < w_i < l_i, \ i = 1, ..., n\}.$$ 

Also, define $B_{xx'}$ to be the event that $x$ and $x'$ are in two different cells of the translated grid, and $A_{xx'}$ be the event that some cell of the randomly translated grid contains both $x$ and $x'$. Further, let $C$ be the set of the random variables which are associated with the realized sample points.

**First-order Inclusion Density.**

For any $x \in U$ there exist a vector of integers $k_x$ and a $w_x \in \Omega$ such that $w_x + k_x l = x$, where $l = (l_1, l_2, ..., l_n)$. Note that $l_i, i = 1, 2, ..., n$, is defined in section 2.1. Now the inclusion density function at $x$ may be given by

$$\pi(x) = \sum_{x \in C} f(x) = \sum_{x \in C} \lim_{h \to 0} \frac{Pr[x - h < X < x + h]}{(2^n)h_1h_2...h_n}$$

24

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\[
\begin{align*}
&= \lim_{h \to 0} \sum_{x \in C} \frac{Pr(x - h < X < x + h, W \in \Omega_x)}{(2^n)h_1 h_2 \ldots h_n} \\
&= \lim_{h \to 0} \sum_{x \in \mathcal{C}} \frac{Pr(x - h < X < x + h | W \in \Omega_x)}{(2^n)h_1 h_2 \ldots h_n} Pr[\Omega_x]
\end{align*}
\]

where \(\Omega_x = \Omega - \{w_x\}\). By using Lemma 2.1, we obtain

\[
\pi(x) = \sum_{x \in \mathcal{C}} f(x | w)(x).
\]

Given a realized point \(w\) of \(W\) in \(\Omega - \{w_x\}\), then there exists only one random variable in \(\mathcal{C}\) which can assume the value of \(x\). Suppose \(X\) is the r.v. that can assume \(x\) and have the grid cell containing \(x\) as its conditional domain. Then we can write

\[
\pi(x) = f(X | w)(x), \quad \text{(2.8)}
\]

that is

\[
\pi(x) = \frac{1}{l_1 l_2 \ldots l_n}. \quad \text{(2.9)}
\]

Second-order Inclusion Density.

The joint inclusion density \(\pi(x, x')\), \(x \neq x' \in U\), can be written as the limit, as \(h \to 0\) and \(h' \to 0\), of

\[
\sum_{x, x' \in \mathcal{C}} \frac{Pr(x - h < X < x + h, x' - h' < Y < x' + h')}{(2^{2n})h_1 h_2 \ldots h_n h_1' h_2' \ldots h_n'}. \quad \text{(2.10)}
\]

To calculate \(\pi(x, x')\), two possible cases can be considered.
(i) The first case is when $\delta(x_i, x'_i) \geq l_i$ for at least some $i$. Expression (2.10) can be written as

$$\sum_{X,Y \in \mathcal{C}} Pr(x - h < X < x + h, x' - h' < Y < x' + h' | W \in \Omega) \prod_{j=1}^{2n} h_j h_j' \cdots h_n h_n'$$  \hspace{1cm} (2.11)

In this case $B_{xx'} = \Omega$ a.e., and (2.11) can be written as

$$\sum_{X,Y \in \mathcal{C}} Pr(x - h < X < x + h, x' - h' < Y < x' + h' | W \in B_{xx'}) \prod_{j=1}^{2n} h_j h_j' \cdots h_n h_n'$$  \hspace{1cm} (2.12)

Further by using Lemma 2.1, the limit of expression (2.12), as $h \to 0$ and $h' \to 0$, can be written as

$$\sum_{X,Y \in \mathcal{C}} f(X,Y|W)(x,x')$$

where $(w \in B_{xx'})$. Given any point $w$ in $B_{xx'}$, then $x$ and $x'$ are in two different grid cells. Therefore, there are two r.v., say $X$ and $X'$, in $\mathcal{C}$ which can assume the values of $x$ and $x'$ respectively. Now, $X$ has the grid cell containing $x$ as its conditional domain and the other $X'$ has the grid cell containing $x'$ as its conditional domain. Then the limit of expression (2.12), as $h \to 0$ and $h' \to 0$, can be written as

$$f(X,X'|W)(x,x')$$

where $(w \in B_{xx'})$. As a result we obtain
\[ \pi(x, x') = \frac{1}{l_1 l_2 \cdots l_n}. \] (2.13)

Since in this case \( Pr(B_{xx'}) = 1 \), the pairwise inclusion density can be written as

\[ \pi(x, x') = \frac{Pr(B_{xx'})}{l_1 l_2 \cdots l_n}. \] (2.14)

(ii) The second case is when \( \delta(x_i, x'_i) < l_i \), for every \( i = 1, 2, \ldots, n \). The numerator of (2.10) can be written as

\[
\sum_{x, y \in c} Pr[\{x - h, x + h\}, \{y - h', y + h'\}] \text{Pr}(B_{xx'}) = \sum_{x, y \in c} Pr[\{x - h, x + h\}, \{y - h', y + h'\}] \in B_{xx'}] + \\
\sum_{x, y \in c} Pr[\{x - h, x + h\}, \{y - h', y + h'\}] \in A_{xx'}]. \quad (2.15)
\]

The second term on the right hand side of (2.15) is zero, because, for \( W \in A_{xx'} \), \( x \) and \( x' \) are in one grid cell and there is only one r.v. in \( C \) which can fall in that cell. The first term on the right hand side of (2.15) can be written as

\[
\sum_{x, y \in c} Pr[\{x - h, x + h\}, \{y - h', y + h'\}] \text{Pr}(B_{xx'}) \text{Pr}(B_{xx'}). \quad (2.16)
\]

Now, when (2.16) is divided by \( 2^n h_1 h_2 \ldots h_n h'_1 h'_2 \ldots h'_n \), and its limit is taken, as \( h \to 0 \) and \( h' \to 0 \), and by using Lemma 2.1, we obtain

\[
\sum_{x, y \in c} f(x, y | w)(x, x') \Pr(B_{xx'})
\]
where \( w \in B_{xx'} \). Given the point \( w \) in \( B_{xx'} \), \( x \) and \( x' \) are in two different grid cells. Therefore, there are two r.v., say \( X \) and \( X' \), in \( C \) which can assume \( x \), and \( x' \) respectively. Note that \( X \) and \( X' \) have the grid cell containing \( x \) and \( x' \) as their conditional domain respectively. As a result we obtain

\[
\pi(x, x') = f_{(X, X'|w)}(x, x') \cdot Pr(B_{xx'}),
\]

where \( w \in B_{xx'} \). Since

\[
f_{(X, X'|w)}(x, x') = \frac{1}{l_1^2 \ldots l_n^2},
\]

the pairwise inclusion density becomes

\[
\pi(x, x') = \frac{Pr(B_{xx'})}{l_1^2 \ldots l_n^2}.
\]

Finally, the results can be summarized by

\[
\pi(x) = \frac{1}{\mu}, \quad \text{for any } x \text{ in } U
\]

and

\[
\pi(x, x') = \frac{Pr(B_{xx'})}{\mu^2}, \quad \text{for any } x \neq x' \in U,
\]

where \( \mu = l_1 l_2 \ldots l_n \).

Now, we calculate the probability of \( B_{xx'} \). Let us define \( B_i \) as
\[ B_i = \{ w_i : \min\{x_i, x'_i\} < w_i + k_i l_i < \max\{x_i, x'_i\}, k_i \in \{1, \ldots, m_i\} \} \quad (2.19) \]

where \( i = 1, 2, \ldots, n \). It is clear that the events \( A_{XX'}, B_{XX'} \) and \( B_i \) are induced by the r.v. \( W \), where \( i = 1, 2, \ldots, n \). Some facts related to these events are given by

\[ \Omega = A_{XX'} \cup B_{XX'} \text{ almost everywhere,} \quad (2.20) \]

\[ B_{XX'} = \bigcup_{i=1}^{n} B_i, \quad (2.21) \]

and

\[ Pr(B_i) = \begin{cases} \frac{\delta(x_i, x'_i)}{l_i} & \text{if } 0 < \delta(x_i, x'_i) < l_i \\ 1 & \text{if } l_i \leq \delta(x_i, x'_i). \end{cases} \quad (2.22) \]

Using (2.22) in (2.21), we obtain

\[ Pr(B_{XX'}) = 1, \quad \text{if } l_i \leq \delta(x_i, x'_i) \text{ for some } i. \quad (2.23) \]

Using (2.21), the probability of \( B_{XX'} \) may be written as

\[ Pr(B_{XX'}) = S_1 - S_2 + S_3 - S_4 + \cdots + (-1)^{n+1} S_n \quad (2.24) \]

where

\[ S_1 = \sum Pr(B_i), \quad S_2 = \sum Pr(B_i B_i'), \quad S_3 = \sum Pr(B_i B_i' B_i'), \ldots \quad (2.25) \]
Here 1 \leq i_1 < i_2 < i_3 < \cdots \leq n, so that in the above sums, each combination appears once and only once; hence \( S_r \) has \( \binom{n}{r} \) terms. The last sum, \( S_n \), reduces to the single term \( \Pr(B_1 B_2 B_3 \cdots B_n) \) which is the probability of the simultaneous realization of all \( n \) events. Further, the events \( B_1, B_2, \ldots, B_n \) are mutually independent. Then,

\[
\Pr(B_{i_1} B_{i_2} \cdots B_{i_r}) = \frac{\prod_{j=1}^{r} \delta(x_{i_j}, x'_{i_j})}{l_{i_1} l_{i_2} \cdots l_{i_r}},
\]

if \( 0 < \delta(x_{i_j}, x'_{i_j}) < l_{i_j} \), for every \( j = 1, \ldots, r \), and \( 1 \leq r \leq n \).

Substituting (2.26) in (2.25) we obtain

\[
S_r = \sum \frac{\prod_{j=1}^{r} \delta(x_{i_j}, x'_{i_j})}{l_{i_1} l_{i_2} \cdots l_{i_r}},
\]

if \( 0 < \delta(x_{i_j}, x'_{i_j}) < l_{i_j} \) for every \( j = 1, \ldots, r \), and \( 1 \leq r \leq n \).

Finally, using (2.23) and (2.24), the probability of \( B_{xx'} \) may be given by

\[
\Pr(B_{xx'}) = \begin{cases} 
1 & \text{if } l_i \leq \delta(x_i, x'_i) \text{ for some } i \\
S_1 - S_2 + S_3 - S_4 + \cdots + (-1)^{n+1} S_n & \text{otherwise},
\end{cases}
\]

where \( S_r \) is given in (2.27). We present the first two particular cases of the pairwise inclusion density function:

For \( n = 1 \),

\[
\pi(x, x') = \begin{cases} 
\frac{1}{l_1^2} & \text{if } l_1 \leq \delta(x_1, x'_1) \\
\frac{\delta(x_1, x'_1)}{l_1^3} & \text{otherwise},
\end{cases}
\]

(2.29)
and for \( n = 2 \),

\[
\pi(x, x') = \begin{cases} 
\frac{1}{l_1^2} & \text{if } l_i \leq \delta(x_i, x_i') \text{ for } i = 1 \text{ or } 2 \\
\frac{1}{l_1^2} \left( \sum_{i=1}^{2} \frac{\delta(x_i, x_i')}{l_i} - \frac{\delta(x_1, x_1') \delta(x_2, x_2')}{l_1 l_2} \right) / (l_1^2 l_2^2) & \text{otherwise.}
\end{cases}
\] (2.30)

### 2.6 Edge Effect on the Estimators

In this section we assume the boundary of the domain \( D \) to be known prior to sampling. Earlier in this chapter, when the domain \( D \) was smaller than the universe \( U \), we extended the response function to all of \( U \) by defining

\[
z(x) = 0, \text{ when } x \in U \text{ and } x \notin D.
\]

Such extension is useful in order to obtain a sample with fixed size which allows the use of Yates-Grundy variance estimator. But, based on a sample selected using a given sampling design, this extension can affect the estimator of the total and the estimator of the variance. If a sample of fixed size can be selected from \( D \), then the estimator of the total and the estimator of the variance will be more stable than the ones based on a sample selected from \( U \). This is true given that the same sampling design and the same grid are used on both \( D \) and \( U \). This can be done when the sampling is from a one-dimensional domain.
In Example 1 below, we illustrate the edge effect on estimation by comparing sampling from $U$ using a noncircular method and sampling from $D$ using a circular method.

For many sampling designs, when the sampling domain has more than one dimension, we cannot select a sample from $D$ with a fixed size unless $D$ has some extreme regular shape. However, later as discussed in Example 2, when the sampling is from a two-dimensional domain, we compare the sampling method presented by Cordy and Thompson (1994) to the new method by using the circular sampling with respect to each dimension from the universe $U$, introduced in section 2.2 of this chapter. The new method which makes use of circular sampling in multidimensions reduces the edge effect which will be shown later by numerical simulation in Example 2.

### 2.6.1 Numerical Example 1

Suppose the response function in one-dimensional domain is given by

$$ z(x) = 20 + 48x^2(1 - x), \quad x \in D, $$

where the domain $D$ is given by $D = \{x : 0 < x < 1\}$. The parameter of interest is given by

$$ T = \int_0^1 z(x) \, dx. $$

Since the response function is known, the exact value of $T$ can be computed as $T = 24$. In this example, based on the $(tss)$ design and for a given spacing between
the grid cells, we compare two estimators of the total $T$, by comparing the variance of these estimators. Further, by simulation, we examine the variances of the Yates-Grundy type of variance estimators. The two methods of estimation are given below.

(i) Let us refer to the first method as the Linear Grid Method. We study this method by using (tss) design and the response function is extended by one grid cell on either side of $D$ in such a way that

$$z(x) = 0, \text{ if } x \in U \text{ and } x \notin D$$

where

$$U = \{x: -d < x < 1 + d\}$$

and $d$ is the grid spacing or the distance between any two consecutive grid edges. For a given $d$, if we select a sample $s_e$ from $U$ using (tss) design, then the first-order inclusion density is given by

$$\pi(x) = \frac{1}{d}, \quad -d < x < 1 + d$$

and the estimator of the total based on this method is given by

$$\hat{T}_1 = \sum_{x_i \in s_e} \frac{z(x_i)}{\pi(x_i)}.$$ 

The second-order inclusion density for $x$ and $x' \in U$ is given by

$$\pi(x, x') = \begin{cases} 1 & \text{if } d < |x - x'| \\ \frac{1}{d^2} & \text{otherwise.} \end{cases}$$

(2.31)
Based on the selected sample \( s_e \) and the above inclusion density functions, we can evaluate the variance of \( \hat{T}_1 \) and we can estimate this variance by using the Yates-Grundy estimator denoted by \( v_1 \). Further, by using simulations we obtain the empirical variance of \( v_1 \). For different values of \( d \), the variance of \( \hat{T}_1 \) and the empirical variance of \( v_1 \), that is based on 5000 iterations, are given in Table 2.1. Note that we did not use circular sampling.

(ii) Let us call the second method the Cyclic Grid Method. By viewing the domain \( D \) in a circular manner and using a one-dimensional cyclic grid, no extension of the response function is needed. Based on a given \( d_i \), if we select a sample \( s \) using \((tss)\) design from \( D \), then the first-order inclusion density is given by

\[ \pi(x) = \frac{1}{d}, \quad x \in D \]

and the estimator of the total based on this method is given by

\[ \hat{T}_2 = \sum_{x_i \in s} \frac{y_i}{\pi(x_i)}. \]

By using the circular sampling, the distance between \( x \) and \( x' \) is given by

\[ \delta(x, x') = \min \{|x - x'|, 1 - |x - x'|\}. \]

The pairwise inclusion density for \( x \) and \( x' \in D \), is given by

\[ \pi(x, x') = \begin{cases} 
\frac{1}{d^2} & \text{if } \delta(x, x') > d \\
\frac{\delta(x, x')}{d^3} & \text{if } \delta(x, x') < d.
\end{cases} \]  

(2.32)
Table 2.1: Comparison between the two methods of estimation explained in (i) and (ii) of Example 1.

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Linear Grid Method</th>
<th>Cyclic Grid Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Empirical Var($\hat{T}_1$)</td>
<td>Var($v_1$)</td>
</tr>
<tr>
<td>.5</td>
<td>43.39</td>
<td>6902</td>
</tr>
<tr>
<td>.25</td>
<td>9.7</td>
<td>557</td>
</tr>
<tr>
<td>.125</td>
<td>2.27</td>
<td>24.3</td>
</tr>
<tr>
<td>.1</td>
<td>1.43</td>
<td>15.2</td>
</tr>
</tbody>
</table>

Based on the selected sample $s$ and the above inclusion density functions, we can evaluate the variance of $\hat{T}_2$ and we can estimate this variance by using the Yates-Grundy estimator denoted by $v_2$. Further, by using simulations we find the empirical variance of $v_2$. For different values of $d$, the variance of $\hat{T}_2$ and the empirical variance of $v_2$, that is based on 5000 iterations, are given in Table 2.1.

2.6.2 Numerical Example 2

Suppose the response function is from a two-dimensional domain and is given by

$$z(x, y) = 4 + \sqrt{1 - (x - 2)^2 - (y - 2)^2}, (x, y) \in D$$

where the domain $D$ is given by

35
\[ D = \{(x, y) : (x - 2)^2 + (y - 2)^2 \leq 1 \}. \]

Since the response function is known in this example, the total is given by

\[
T = \int_2^4 \int_{L_1(y)}^{L_2(y)} z(x, y) \, dx \, dy = \frac{14}{3} \pi,
\]

where \( L_1(y) = 2 - \sqrt{-y^2 + 4y - 3} \) and \( L_2(y) = 2 + \sqrt{-y^2 + 4y - 3} \). Based on the \( (tss) \) design and for a given size of the grid cell we compare two estimators of the total \( T \) by comparing the variance of these estimators. Further, we study by simulation the variances of the variance estimators. The two methods of estimation are given below.

(i) Let us call the first method Cordy's Noncyclic Grid Method. Cordy and Thompson (1994) suggested using a rectangular shape universe, say \( U_C \) which expanded by at least one grid cell from all sides of \( D \). Also, they used the same \( (tss) \) design except that they did not make use of the two-dimensional cyclic in the formulas of the pairwise inclusion density. By considering their arguments, we study this method by using \( (tss) \) design from a universe \( U_C \) which we define to be a square given by

\[
U_C = \{(x, y) : 1 - d < x < 3 + d; \text{and } 1 - d < y < 3 + d \}.
\]

We use a grid on \( U_C \) such that each cell is a square with each side being of length \( d \) (Figure 2.6). The response function is extended over \( U_C \) in such a way that

\[ z(x, y) = 0, \quad \text{if } (x, y) \in U_C \text{ and } (x, y) \notin D. \]
Figure 2.6: The domain $D$ is the circle with center at $(2,2)$ and radius 1, and its corresponding $U_C$ as explained in (i) of Example 2.

For a given $d$, if we select a sample $s_e$ from $U_C$ using (tss) design, then the first-order inclusion density is given by

$$\pi(x, y) = \frac{1}{d^2}, \quad (x, y) \in U_C$$

and the estimator of total based on this method is given by

$$\hat{T}_1 = \sum_{(x,y) \in s_e} \frac{z(x,y)}{\pi(x,y)}.$$ 

The second-order inclusion density, for $x = (x_1, x_2)$, $x' = (x'_1, x'_2) \in U_C$, is given by

$$\pi(x, x') = \begin{cases} 
\frac{1}{d^4} & \text{if } d < |x_i - x'_i| \text{ for } i = 1 \text{ or } 2 \\
\sum_{i=1}^{2} \frac{|x_i - x'_i|}{d^2} - \frac{|x_1 - x'_1||x_2 - x'_2|}{d^6} & \text{otherwise.}
\end{cases} \quad (2.33)$$
Based on the selected sample $s_e$ and the above inclusion density functions, we can evaluate by simulations the variance of $\hat{T}_1$ and we can estimate this variance by using the Yates-Grundy estimator which we denote by $v_1$. For different values of $d$, the empirical variance of $\hat{T}_1$ and the empirical variance of $v_2$ are based on 3000 iterations and are given in Table 2.2.

(ii) Let us call the second method as the New Cyclic Grid Method. We study this method by using $(tss)$ design from a universe $U$ which we define to be a square given by

$$U = \{(x, y): 1 < x < 3; \text{and } 1 < y < 3\}.$$  

We use a grid on $U$ such that each cell is a square with either side being of length $d$ (Figure 2.7). The response function is extended over $U$ in such a way that

$$z(x, y) = 0, \quad \text{if } (x, y) \in U \text{ and } (x, y) \notin D.$$  

We view the universe $U$ in a circular manner with respect to each dimension. Based on a given $d$, if we select a sample $s$ using $(tss)$ design from $U$, then the first-order inclusion density is given by

$$\pi(z, y) = \frac{1}{d^2}, \quad (x, y) \in U,$$

and the estimator of total based on this method is given by

$$\hat{T}_2 = \sum_{(x, y) \in s} \frac{z(x, y)}{\pi(x, y)}.$$  

By using the circular sampling with respect to each dimension of $U$, the distance between $x_1$ and $x'_1$, where $x_1$ and $x'_1$ are the first coordinates of $(x_1, x_2)$
Figure 2.7: The domain $D$ is the circle with center at $(2,2)$ and radius 1, and its corresponding universe $U$ as explained in (ii) of Example 2.

and $(x'_1, x'_2)$, is given by

$$\delta(x_1, x'_1) = \min \{|x_1 - x'_1|, 1 - |x_1 - x'_1|\}.$$  

Similarly for $\delta(x_2, x'_2)$. The pairwise inclusion density of $x = (x_1, x_2)$ and $x' = (x'_1, x'_2) \in U$, is given by

$$\pi(x, x') = \begin{cases} 
\frac{1}{d^4} & \text{if } d \leq \delta(x_i, x'_i), \text{ for } i = 1 \text{ or } 2 \\
\sum_{i=1}^{2} \frac{\delta(x_i, x'_i) \delta(x_2, x'_2)}{d^5} & \text{otherwise.}
\end{cases} \tag{2.34}$$

Based on the selected sample $s$ and the above inclusion density functions, we can evaluate by simulations the empirical variance of $\hat{T}_2$ and this variance can be estimated by using the Yates-Grundy estimator which we denote by $v_2$. For
Table 2.2: Comparison between the two methods of estimation explained in (i) and (ii) of Example 2.

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Cordy’s Noncyclic Grid Method</th>
<th>New Cyclic Grid Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Empirical Var((\hat{T}))</td>
<td>Empirical Var((u_1))</td>
</tr>
<tr>
<td>(d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21.8</td>
<td>620</td>
</tr>
<tr>
<td>.5</td>
<td>2.52</td>
<td>2.17</td>
</tr>
<tr>
<td>.25</td>
<td>.306</td>
<td>.0168</td>
</tr>
</tbody>
</table>

different values of \(d\), the empirical variance of \(\hat{T}_1\) and the empirical variance of \(u_2\) are based on 3000 iterations and are given in Table 2.2.

We compare the estimators using the \((tss)\) design based on noncyclic argument combined with Cordy’s universe \(U_C\) and the cyclic argument combined with the universe \(U\) defined above.

2.7 Conclusion

In this chapter, when the sampling was from an \(n\)-dimensional space, a new mathematical argument for calculating the inclusion density functions for the \((tss)\) design is provided. It should be noted that the variance estimators based on the \((tss)\) designs take only nonnegative values. Also the method of viewing the sampling
from multidimensional space in a circular manner with respect to each dimension increases the precision of the estimators.

The universe $U$ is commonly used in practice and the edge effect is ignored by many statisticians. As Table 2.1 and Table 2.2 show, the edge effect can possibly have severe effects on the stability of the estimators especially when the response function is far from zero near the boundary from inside the domain $D$. This is a common situation in practice and its effect can be reduced by selecting $U$ to be as small as possible such that $D$ is a subset of $U$. Alternatively, it could be practical to use a regular shape $U$, which is a subset of $D$ and the problem becomes that of estimating the parameters of interest over $U$, only (Figure 2.8). Note that when the sampling is from $U$, there will be no edge effect by using the circular sampling argument.
Chapter 3

SAMPLING FROM AN ORDERED POPULATION WITH A GIVEN AUXILIARY VARIABLE

3.1 Introduction

Consider a population in which the units on which the measurements are made are ordered in some way. The ordering may be natural, such as contiguous strips of land, or the sampler may sometimes have enough advance knowledge to arrange the units at least roughly in an order for which a precision can be gained. Also, the units differ considerably in respect of a size measure $x$, positively correlated with the study variable $y$. Then one efficient sampling scheme is to divide the population into strata and sample proportionately one unit from each stratum. This sampling procedure shall be explained properly and illustrated by an example in the next section.
For a population with the characteristics defined above, a very common sampling procedure in the statistical literature combines systematic sampling with probability proportional to size (pps) sampling. Using this technique, a sample of size $n$ is drawn following the steps given below.

(i) Cumulate the sizes of the units, $\Pi_i = \Pi_{i-1} + x_i$ where $\Pi_0 = 0$;

(ii) let the sampling interval be, $d = \frac{\Pi_n}{n}$;

(iii) select a uniform random number $r$ from 0 to $d$;

(iv) select unit $i$ if $\Pi_{i-1} < r + jd \leq \Pi_i$, $j = 0, 1, 2, \ldots, (n-1)$.

This technique is termed fixed order pps systematic (pps-sys) sampling and was suggested by Madow (1949). Its advantages are that it provides efficient estimators of the population total and the sample is easy to draw for any $n$. Furthermore, the units can be arranged in an order for which the systematic sampling performs well. Goodman and Kish (1950) used this design in controlled selection. Also this technique is presently widely used in many sampling situations. Murthy and Rao (1988) presented some useful results concerning this design. We review a few common results for this sampling method. An unbiased estimator of the population total is given by

$$\hat{T} = d \sum_{i=1}^{N} t_i \frac{y_i}{x_i} \quad (3.1)$$

where $t_i$ is the number of times the $i^{th}$ unit is selected in the sample.

A drawback of the fixed order pps systematic method is, as usual, the absence of a valid variance estimator. To get simple variance estimator Murthy and Rao
(1988) suggested using pps systematic sampling with multiple random starts. It is known that the fixed order pps systematic with multiple random starts method is less efficient than the one with one random start. In the case when the units in the population are in random order, \( N \) is large, and \( x_i < d \) for \( i = 1, 2, \ldots, N \), Hartley and Rao (1962) derived approximate expressions for the variance and variance estimator. These are given by

\[
\text{Var}(\hat{T}) = \sum_{i=1}^{N} \pi_i \left[ 1 - \frac{(n - 1)}{n} \pi_i \right] \left( \frac{y_i - \hat{T}}{\pi_i} \right)^2, \quad (3.2)
\]

and

\[
u(\hat{T}) = \frac{1}{(n - 1)} \sum_{i=1}^{n} \sum_{j>i} \left[ 1 - (\pi_i + \pi_j) + \left( \frac{\sum_{k=1}^{N} \pi_k^2}{n} \right) \right] \left( \frac{y_i - y_j}{\pi_i \pi_j} \right)^2, \quad (3.3)
\]

where \( \pi_i = \frac{x_i}{d} \) for \( i = 1, 2, \ldots, N \). The technique which led to these formulas is termed random order pps systematic sampling. The variance estimator given in (3.3) is frequently used. But the formulas given in (3.2) and (3.3) are not applicable to a population in a given order. Also both assumptions that \( x_i < d \) for all \( i \), and \( N \) being large may not be satisfied in some populations. Stehman and Overton (1994) provided an estimator of the variance formula given in (3.2), for which the size variable \( x \) is only available for the sample units. Hidiriglou and Gray (1980) provided a FORTRAN program for computing the exact \( \pi_{ij} \)'s under random order pps systematic sampling. Further, this technique has been considered in detail, among others, by Grundy (1954), Des Raj (1965), Connor (1966), Hartley (1966), [44]
Murthy (1967), Hidiroglou and Gray (1975), and Isaki and Pinciaro (1977). The focus of this chapter is to investigate the sampling procedure which is presented in the following section.

3.2 Sampling Procedure

(i) Arrange the units in a certain order (perhaps according to their natural ordering) and denote the ordering by \( j = 1, 2, ..., N \) and let \( \Pi_j = \sum_{i=1}^{j} x_i \), \( \Pi_0 = 0 \), be the partial sum of the \( (x_i) \) in that order. Consider a circle with perimeter length \( \Pi_N \) and marked off with arcs of lengths \( x_1, x_2, ..., x_N \) in clockwise direction starting from the top (\( j^{th} \) unit corresponds to \( j^{th} \) arc of length \( x_j \) on the circle).

(ii) Superimpose a one-dimensional cyclic grid on the circle described above. Without loss of generality, we position the grid in such a way that one of its edges is on the top of the circle. Let \( d = \frac{\Pi_N}{n} \) where \( n \) is the desired sample size. Consider \( d \) to be the arc length between any two consecutive grid edges. Note that there will be \( n \) grid cells. This sampling design allows for replacement.

(iii) Select a uniform random point \( g \) with \( 0 \leq g < d \). Rotate the grid in a clockwise direction along the circle with length \( g \). Independently select a uniform random point from within each cell of the rotated grid. Denote these realized points by \( (c_1, c_2, ..., c_n) \). Then, the sampled units are those whose indices \( j \) satisfy

\[
\Pi_{j-1} \leq c_i < \Pi_j \tag{3.4}
\]
for at least some \( i \). This sampling technique is termed as fixed order stratified sampling.

The fixed order stratified sampling has apparent advantages over the random order pps systematic sampling. It does not require the assumption that \( d \) must be greater than the largest \( x \). Also, the population can retain any order of the units, natural ordering or artificial ordering. Further, three unbiased variance estimators will be provided, which can be calculated and are nonnegative.

### 3.3 Numerical Example

Consider a population of size \( N = 8 \) units (Table 3.1) arranged in a given order and with sizes \( x_i \) shown in the second column of Table 3.1. Four random points are to be selected. Then \( d = \frac{\Pi x}{4} = \frac{100}{4} = 25 \). Suppose \( g = 9.235 \) is a uniform random variate from uniform \((0,25)\). Then, starting from the top of the circle the grid is rotated clockwise along the circle with arc length 9.235. Therefore, the grid edges are at 9.235, 35.235, 59.235 and 84.235. Note that the shortest distance between the last and the first grid edge can be calculated as 109.235 - 84.235 = 25. The realizations of the uniform random points which are selected from within the cells of the rotated grid are \( c_1 = 25.35, c_2 = 57.32, c_3 = 79.34 \) and \( c_4 = 103.38 \). Because of the cyclic design and because \( c_4 > 100 \), then, we consider \( c_4 = 103.38 \mod (100) = 3.38 \). Since \( \Pi_2 \leq c_1 < \Pi_3, \Pi_5 \leq c_2, c_3 < \Pi_6 \) and \( \Pi_0 \leq c_4 < \Pi_1 \), the units \( j = 1, 3 \) and 6 are selected. Note that the design allows for multiplicity (e.g. the unit \( j = 6 \) is selected twice).
Table 3.1: Selection of $n = 4$ random arcs from a population of $N = 8$ units.

<table>
<thead>
<tr>
<th>Unit $j$</th>
<th>Size $x_j$</th>
<th>$\Pi_j = \sum_i^j x_i$</th>
<th>Assigned range</th>
<th>Selected units</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>7</td>
<td>0-7</td>
<td>$c_4 = 103.38 \mod (100) = 3.38 \ (j = 1)$</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>24</td>
<td>7-24</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>37</td>
<td>24-37</td>
<td>$c_1 = 25.35 \ (j = 3)$</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>48</td>
<td>37-48</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>57</td>
<td>48-57</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>81</td>
<td>57-81</td>
<td>$c_2 = 57.32 \ c_3 = 79.34 \ (j = 6, \ twice)$</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>87</td>
<td>81-87</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>100</td>
<td>87-100</td>
<td></td>
</tr>
</tbody>
</table>
3.4 Estimation

Let $t_j$ be the number of $c_i$ satisfying (3.4), for $j = 1, \ldots, N$. In other words, $t_j$ can be interpreted as the number of variates $c_i$ that fall in the range of $x_j$. Also, let $I_j$ be equal to one if at least one $c_i$ satisfy (3.4) and zero otherwise, for $j = 1, \ldots, N$.

The first estimator of total is

$$\hat{T}_1 = d \sum_{j=1}^{N} t_j \frac{y_j}{x_j}, \quad (3.5)$$

where $x_j$ and $y_j$ are constants and $t_j$ are the random variables.

The second estimator of total is

$$\hat{T}_2 = \sum_{j=1}^{N} I_j \frac{y_j}{\pi_j}, \quad (3.6)$$

where $\pi_j$ is the probability of the $j^{th}$ unit being selected in the sample.

The estimator given in (3.6) is the Horvitz-Thomson estimator of population total in sampling from finite population. Its variance is given in (1.9). Since the sample size is not fixed, the Yates-Grundy variance estimator can’t be used. So the Horvitz-Thompson variance estimator can be used and is given by

$$v_{HT}(\hat{T}_2) = \sum_{i=1}^{N} \frac{(1 - \pi_i)}{\pi_i^2} y_i^2 I_i + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{(\pi_{ij} - \pi_i \pi_j)}{\pi_{ij} \pi_i \pi_j} y_i \frac{y_j}{I_i I_j}, \quad (3.7)$$

where $\pi_{ij}$ is the pairwise inclusion probability.
We investigated the estimators $\hat{T}_1$ and $\hat{T}_2$ and we concluded that $\hat{T}_1$ is more efficient than $\hat{T}_2$. Further information about this fact is given in section 3.6. Therefore, in the remainder of this chapter, we investigate the properties of $\hat{T}_1$ only.

**Theorem 3.1** The estimator $\hat{T}_1$ is unbiased for $T$, with

$$Var(\hat{T}_1) = d^2 \sum_{i=1}^{N} Var(t_i) \left( \frac{y_i}{x_i} \right)^2 + 2 d^2 \sum_{i=1}^{N} \sum_{j>i}^{N} Cov(t_i, t_j) \frac{y_i y_j}{x_i x_j}. \quad (3.8)$$

**Proof:** Intuitively it can be seen that

$$E(t_i) = \frac{x_i}{d}.$$

Then

$$E(\hat{T}_1) = d \sum_{i=1}^{N} \frac{y_i}{x_i} E(t_i) = \sum_{i=1}^{N} y_i.$$

Therefore $\hat{T}_1$ is an unbiased estimator of $T$. The variance of $\hat{T}_1$ can be easily verified.

□

**Theorem 3.2** The variance of $\hat{T}_1$ given in (3.8) can also be expressed in the following form:

$$Var(\hat{T}_1) = -d^2 \sum_{i=1}^{N} \sum_{j>i}^{N} Cov(t_i, t_j) \left( \frac{y_i}{x_i} - \frac{y_j}{x_j} \right)^2. \quad (3.9)$$

**Proof:** Since $n = \sum_{j=1}^{N} t_j$ is fixed,

$$Cov(\sum_{j=1}^{N} t_j, t_i) = Cov(n, t_i) = 0, \text{ for } i = 1, \ldots, N.$$
Moreover,
\[ \text{Cov}(\sum_{j=1}^{N} t_j, t_i) = \text{Cov}(\sum_{j \neq i}^{N} t_j, t_i) + \text{Var}(t_i) = 0, \]
then
\[ \text{Var}(t_i) = -\sum_{j \neq i}^{N} \text{Cov}(t_j, t_i). \]

Substituting this in the first term in (3.8), we get
\[
d^2 \sum_{i=1}^{N} \left( \frac{y_i}{x_i} \right)^2 \text{Var}(t_i) = d^2 \sum_{i=1}^{N} \sum_{j \neq i}^{N} \text{Cov}(t_i, t_j) \left( \frac{y_i}{x_i} \right)^2 \\
= -d^2 \sum_{i=1}^{N} \sum_{j \neq i}^{N} \text{Cov}(t_i, t_j) \left[ \left( \frac{y_i}{x_i} \right)^2 + \left( \frac{y_j}{x_j} \right)^2 \right].
\]

Hence,
\[
\text{Var}(\hat{T}_1) = -d^2 \sum_{i=1}^{N} \sum_{j \neq i}^{N} \text{Cov}(t_i, t_j) \left[ \left( \frac{y_i}{x_i} \right)^2 + \left( \frac{y_j}{x_j} \right)^2 - 2\frac{y_i y_j}{x_i x_j} \right] \\
= -d^2 \sum_{i=1}^{N} \sum_{j > i}^{N} \text{Cov}(t_i, t_j) \left( \frac{y_i}{x_i} - \frac{y_j}{x_j} \right)^2. \]

**COROLLARY 3.3 (i).** Based on (3.8), unbiased estimators of the variance of \( \hat{T}_1 \) is
given by
\[
v_1(\hat{T}_1) = d^2 \sum_{i} t_i \frac{\text{Var}(t_i)}{E(t_i)} \left( \frac{y_i}{x_i} \right)^2 + 2d^2 \sum_{i} \sum_{j > i} \frac{t_i t_j}{E(t_i t_j)} \text{Cov}(t_i, t_j) \frac{y_i y_j}{x_i x_j} \quad (3.10)
\]
provided that \( E(t_i) > 0 \) and \( E(t_i t_j) > 0 \) for all units \( i \) and \( j \) and
\[
v_2(\hat{T}_1) = d^2 \sum_{i} \frac{I_i}{\pi_i} \text{Var}(t_i) \left( \frac{y_i}{x_i} \right)^2 + 2d^2 \sum_{i} \sum_{j > i} \frac{I_i I_j}{\pi_{ij}} \text{Cov}(t_i, t_j) \frac{y_i y_j}{x_i x_j} \quad (3.11)
\]
provided that \( \pi_i > 0 \) and \( \pi_{ij} > 0 \).

(ii). Based on (3.9), the unbiased estimators of the variance of \( \hat{T}_1 \) are given by

\[
v_3(\hat{T}_1) = -d^2 \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{t_i t_j}{E(t_i t_j)} \text{Cov}(t_i, t_j) \left( \frac{y_i}{x_i} - \frac{y_j}{x_j} \right)^2
\]

and

\[
v_4(\hat{T}_1) = -d^2 \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{I_i I_j}{\pi_{ij}} \text{Cov}(t_i, t_j) \left( \frac{y_i}{x_i} - \frac{y_j}{x_j} \right)^2.
\]

**Proof:** By taking the expected value of \( v_1(\hat{T}_1) \) and noting that \( t_i \) is the only random variable in (3.10), it can be seen that \( v_2(\hat{T}_1) \) is unbiased for the variance of \( \hat{T}_1 \). In (3.11), since \( E(I_i) = \Pr(t_i > 0) = \pi_i \) and \( E(I_i I_j) = \Pr(t_i t_j > 0) = \pi_{ij} \), it easily follows that \( v_2(\hat{T}_1) \) is an unbiased estimator for the variance of \( \hat{T}_1 \). By using similar arguments we can conclude that the expected values of \( v_3(\hat{T}_1) \) and of \( v_4(\hat{T}_1) \) is just the variance of \( \hat{T}_1 \) as presented in (3.9). □

Under the present sampling design, we have \( \pi_i > 0 \) and \( \pi_{ij} > 0 \) for any \( i \) and \( i, j \) in the population. These facts shall be confirmed in the results of next subsection which is to follow.

### 3.4.1 Calculations of the Inclusion Probability Functions

For a given vector of the \( x \) values and each grid cell of size \( d \), let \( a = x_i/d \),
\( b = x_j/d , i \neq j \), and let \( \tau \) be the shortest arc length between \( x_i \) and \( x_j \) divided by \( d \), that is
Also let $\tau'$ be the longest arc length between $x_i$ and $x_j$ divided by $d$. In other words

$$\tau' = n - a - b - \tau,$$

since

$$n = \frac{\Pi_N}{d}.$$

The $i^{th}$ and $j^{th}$ units shall be referred by the $a^{th}$ and $b^{th}$ units respectively. We assume it is known to the reader that $a$, $b$, $\tau$ and $\tau'$ are expressed in units of size $d$. Now, to calculate the inclusion probabilities, we can use this setup of $a$, $b$, $\tau$ and $\tau'$ on the grid with cells of size 1. Further, denote $\pi_i$ and $\pi_{ij}$ by $\pi_a$ and $\pi_{ab}$ respectively. We provide explicit formulas for these probabilities in terms of $a$ and $(a, b, \tau, \tau', n)$ respectively.

**First-order Inclusion Probability.**

Without loss of generality, we straighten the circle and the grid on the real number line in such a way that one of the grid edges is at point 0. Also the unit $a$ is placed between point 0 and the point $a$ on the real line. Now the grid is uniformly shifted such that the grid edge, originally at zero, assumes the variate $g$, where $g$ is a uniform variate from $(0, 1)$. Let $A$ be the event that unit $a$ is selected to be in the sample. For a given $g$, let $A_i$ be the event that the unit $a$ is selected to be in
the sample as a result of the random point that falls in \((g + i - 2, g + i - 1)\). Also let \(\tilde{A}_i\) be the complement of \(A_i\) in \((g + i - 2, g + i - 1)\) where \(i = 1, 2, 3, \ldots\) We now show how to calculate the first-order inclusion probability.

![Figure 3.1](image)

Starting with the case \(0 < a < 1\), if \(0 < g < a\) (Figure 3.1), then the conditional probability of unit \(a\) being selected to be in the sample, for the given \(g\), is given by

\[ p_1 = Pr(A | g) = Pr(A_1 \cup A_2) = Pr(A_1 \cap \tilde{A}_2) + Pr(A_2), \]

that is, \(p_1 = a(g + 1 - a) + (a - g)\). Also if \(a < g < 1\), then the conditional probability of unit \(a\) being selected to be in the sample is given by \(p_2 = Pr(A | g) = Pr(A_1) = a\). Therefore,

\[ Pr(A) = \int_0^a p_1 \, dg + \int_a^1 p_2 \, dg \]

and that leads to

\[ Pr(A) = a\{1 - \frac{a^2}{6}\}. \]

The second case is when \(1 < a < 2\). If \(0 < g < a - 1\), then \(p_1 = Pr(A | g) = Pr(A_1 \cup A_2) = Pr(A_2) = 1\). Also if \(a - 1 < g < 1\), then \(p_2 = Pr(A | g) = Pr(A_1 \cup A_2) = Pr(A_1 \cap \tilde{A}_2) + Pr(A_2)\), that is, \(p_2 = g(g + 1 - a) + (a - g)\). Therefore, the probability of \(A\) can be written as

\[ Pr(A) = \int_0^{a-1} p_1 \, dg + \int_{a-1}^1 p_2 \, dg \]
and that results in

\[ Pr(A) = 1 - \frac{(2 - a)^3}{6}. \]

Finally when the sample size \( n > 1 \), the first-order inclusion probability is given by

\[
\pi_a = \begin{cases} 
    a\{1 - \frac{a^2}{6}\} & \text{if } a \leq 1 \\
    1 - \frac{(2 - a)^3}{6} & \text{if } 1 < a \leq 2 \\
    1 & \text{if } 2 < a.
\end{cases}
\]  

(3.15)

**Pairwise Inclusion Probability.**

Note that \( \pi_{ab} \) takes many different forms. Since \( \pi_{ab} = \pi_{ba} \), it will be sufficient to provide the formulas of \( \pi_{ab} \) for which \( 0 < a < b \). Assume the unit \( a \) and the grid cell are placed on the real number line as explained above. Also let \( A_i, \tilde{A}_i \) and \( A \) be as defined above. Further the unit \( b \) is placed between point \( a + \tau \) and \( a + \tau + b \) on the real number line. Now the grid is uniformly translated such that the grid edge which originally at zero assumes the variate \( g \) where \( g \) is uniform \((0,1)\). From the translated grid a uniform random point is selected independently from each grid cell. Now, define \( B \) as the event that unit \( b \) is selected to be in the sample. For a given \( g \), let \( B_i \) be the event that the random point selected from \((g + i - 2, g + i - 1)\) falls in the range of unit \( b \). Also let \( \tilde{B}_i \) be the complement of \( B_i \) in \((g + i - 2, g + i - 1)\), where \( i = 1, 2, 3, \ldots \). For a given \( g \), let \( T_i \) be the event that the random point selected from \((g + i - 2, g + i - 1)\) falls between point \( a \) and \( a + \tau \) where \( i = 1, 2, \ldots \) (Figure 3.2-3.8). Now we provide all the possible
formulas of \( \pi_{ab} \) for which \( 0 < a < b < 2 \) and \( 0 \leq \tau < 2 \). At the end of this section we show that, for any two units in the population we can calculate their pairwise inclusion probability by using one of the forms presented below. Each of these formulas shall be associated with a number, say \( j \), and a set of restrictions on \( a, b \) and \( \tau \). Each combination denoted by a \( j^{th} \) number, a set of restrictions, and a \( \pi_{ab,j} \) formula shall be referred by the \( j^{th} \) case. From the first to sixteenth case, \( \tau \) is restricted to \( 0 < \tau < 1 \).

1. For given \( a, b \) and \( \tau \) such that \( 0 < a + \tau < 1 \) and \( n \geq 2 \), the calculation of the pairwise inclusion probability is as follow:

\[
\begin{align*}
A_1 & \quad A_2 & \quad T_2 & \quad B_2 \\
0 & \quad g & \quad a & \quad a + \tau & \quad a + \tau + b & \quad 1 & \quad g + 1 \\
\end{align*}
\]

Figure 3.2

If \( 0 < g < a \) (Figure 3.2), then the conditional probability of unit \( a \) and unit \( b \) being selected to be in the sample, for given \( g \), is given by

\[
p_1 = Pr(A \cap B | g) = Pr(A_1 \cap B_2) = gb.
\]

Figure 3.3

If \( a < g < a + \tau \) (Figure 3.3), then the probability can be given by

\[
p_2 = Pr(A \cap B | g) = Pr(A_1 \cap B_2) = ab.
\]
If $a + \tau < g < a + \tau + b$ (Figure 3.4), then the probability can be given by

$$p_3 = Pr(A \cap B \mid g) = Pr(A_1 \cap B_2) = a(a + b + \tau - g).$$

Also if $a + \tau + b < g < 1$, then $p_4 = Pr(A \cap B \mid g) = 0$.

Finally,

$$Pr(A \cap B) = \int_a^{a+\tau} p_1 \, dg + \int_{a+\tau}^{a+\tau+b} p_2 \, dg + \int_{a+\tau+b}^{1} p_3 \, dg$$

and that leads to the first formula of the pairwise inclusion probability which we denote by

$$\pi_{ab,1} = \frac{ab(a + 2\tau + b)}{2}.$$

2. By holding $a$ fixed as in case 1 and increasing $b$ until

$$0 \leq a + b + \tau - 1 \leq a \leq a + \tau \leq 1$$

is satisfied, we get, for $n = 2$,

$$\pi_{ab,2} = \pi_{ab,1} - \frac{(a + \tau + b - 1)^3}{3},$$

and if $n \geq 3$, then

$$\pi_{ab,2} = \pi_{ab,1} - \left(1 - \frac{\tau}{2}\right)\frac{(a + \tau + b - 1)^3}{3}.$$
3. Again by increasing $b$ until $0 \leq a \leq a + b + \tau - 1 \leq a + \tau \leq 1$ is satisfied, for $n = 2$, we have

$$\pi_{ab,3} = -a \left\{ \frac{a^2}{3} + \frac{b^2}{2} + \frac{1}{2}ab - b + (\tau - 1)(a + b + \tau - 1) \right\}$$

and if $n \geq 3$, then

$$\pi_{ab,3} = \frac{a}{6}(z_3 + 9b - b^3 - 5),$$

where $z_3 = a(-3b - 2a + 6) + \tau a(-9 + 3b + a + 3\tau) + \tau(-6(b - 2) + \tau(2\tau + 3b - 9))$.

For each of the cases $4 - 6$, $a$ is held fixed similar to case $3$, while $b$ is increased until a new restriction is satisfied. Under this new restriction a new form of the pairwise inclusion probability is calculated.

4. If $0 \leq a \leq a + \tau \leq a + b + \tau - 1 \leq 1$ and $n = 2$, then

$$\pi_{ab,4} = \frac{1}{2}a\left\{ \frac{1}{3}a^2 - (a + b + \tau - 1)^2 - (\tau - 1)^2 + b(a + 2) \right\}$$

and if $n \geq 3$, then

$$\pi_{ab,4} = \frac{a}{6}(z_3 - 3(2 - b)^2 + 6).$$

5. If $0 \leq a + b + \tau - 2 \leq a \leq a + \tau \leq 1$ and $n \geq 3$, then

$$\pi_{ab,5} = (1 - \tau)^2\left\{ \frac{(b + \tau)^3}{6} - (1 - \tau) \right\} + b\tau\{(b + \tau - 4)(1 - \frac{a}{2}) + \tau \}

- (b - 1)^2 - \frac{a}{6}\{(\tau - 2)^3 + 2a^2\} + \frac{(\tau + 2)^2}{3}. $$

57

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
6. If \( 0 \leq a \leq a + b + \tau - 2 \leq a + \tau \leq 1 \) and \( n \geq 3 \), then
\[
\pi_{ab,6} = \frac{a}{6} \{ 6 - a^2 + (b - 2)^3 \}.
\]

7. In this case, increase \( a \) and select \( b \) such that \( 0 \leq a + \tau - 1 \leq a \leq a + b + \tau - 1 \leq 1 \).

Then, for \( n = 2 \), we can show that
\[
\pi_{ab,7} = \frac{ab}{2} \{ 4 - a - b - 2\tau \} - \frac{(1 - \tau)^3}{3},
\]
and if \( n \geq 3 \), then
\[
\pi_{ab,7} = ab \{ 1 - \left( \frac{a^2 + b^3}{6} \right) \} - \frac{1}{6} \{ 2 - (a + b + \tau) \} (1 - \tau)^3.
\]

For the cases 8 – 11, \( a \) is held fixed as in case 7 while \( b \) is increased until a new restriction is satisfied. Under this new restriction a new form of the pairwise inclusion probability is calculated. Also for the 8th case we show how to calculate the formula of \( \pi_{ab,8} \) as a second example.

8. When \( 0 \leq a + b + \tau - 2 \leq a + \tau - 1 \leq a \leq 1 \), for \( n = 3 \), we get
\[
\pi_{ab,8} = -\frac{1}{3} (\tau - \frac{3}{2})^4 + \frac{1}{6} (a + b)^3 (1 - \tau) - \frac{1}{6} (a + b)^2 (3\tau^2 - 9\tau + 6 + ab)
\]
\[-\frac{2}{3} (a + b) \{ (\tau - \frac{3}{2})^3 + \frac{1}{8} \} + \frac{1}{3} ab (ab + 3) + \frac{1}{48}\]
and if \( n \geq 4 \), then the pairwise inclusion probability can be calculated as follows:
If $0 < g < a + b + \tau - 2$ (Figure 3.5), then the conditional probability of $a$ and $b$ being in the sample can be given by $p_1 = \Pr(A \cap B \mid g) = \Pr(A_1 \cap T_2 \cap B_3) + \Pr(A_1 \cap T_2 \cap T_3 \cap B_4) + \Pr(A_2 \cap T_3 \cap B_4) + \Pr(A_2 \cap B_3)$ and that is

$$p_1 = g(g + 1 - a)(g + 2 - a - \tau) + g(g + 1 - a)(a + \tau - g - 1)(a + b + \tau - g - 2) + (a - g)(a + \tau - g - 1)(a + b + \tau - g - 2) + (a - g)(g + 2 - a - \tau).$$

If $a + b + \tau - 2 < g < a + \tau - 1$ (Figure 3.6), then the conditional probability of $a$ and $b$ being in the sample can be given by $p_2 = \Pr(A \cap B \mid g) = \Pr(A_1 \cap T_2 \cap B_3) + \Pr(A_2 \cap B_3)$ and that is just

$$p_2 = g(g + 1 - a)b + (a - g)b.$$
If \( a + \tau - 1 < g < a \) (Figure 3.7), then the conditional probability of \( a \) and \( b \) being in the sample can be given by

\[
p_3 = Pr(A \cap B \mid g) = Pr(A_1 \cap B_1) + Pr(A_1 \cap T_2 \cap B_3) + Pr(A_2 \cap B_3) \quad \text{and that is just}
\]

\[
p_3 = g(g + 1 - a - \tau) + g\tau(a + b + \tau - g - 1) + (a - g)(a + b + \tau - g - 1).
\]

Figure 3.8

If \( a < g < d \) (Figure 3.8), then the conditional probability of \( a \) and \( b \) being in the sample can be given by

\[
p_4 = Pr(A \cap B \mid g) = Pr(A_1 \cap B_2) + Pr(A_1 \cap T_2 \cap B_3)
\]
and that simplifies to

\[
p_4 = a(g + 1 - a - \tau) + a(a + \tau - g)(a + b + \tau - g - 1).
\]

Finally,

\[
Pr(A \cap B) = \int_0^{a+\tau+\tau-2} p_1 \, dg + \int_{a+\tau-2}^{a+\tau-1} p_2 \, dg + \int_{a+\tau-1}^{a} p_3 \, dg + \int_{a}^{1} p_4 \, dg.
\]
Calculating these integrals and after lengthy algebraic manipulation, we obtain the form

\[
\pi_{ab,8} = \frac{-\tau}{6} \{ (a^2 + b^2)(ab + 6\tau + 19\tau + 21) + (a + b) \{ \frac{-\tau}{6} (9\tau + 25) - ab - \frac{31}{6} \} \}
\]

60

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\[+(\tau - 2)^3\{((\tau - 2)(\frac{\tau - 7}{30} + \frac{a + b}{12}) + \frac{(1 - a)(1 - b)}{6} - \frac{7}{3}\} + \frac{23}{3}ab\]
\[+(2 - a)^3\{(a - 2)(\frac{a}{20} + \frac{1}{15} + \frac{b}{12} + \frac{\tau}{6}) + \frac{\tau}{6}(\tau + 4) + \frac{b}{2}\} + \frac{7}{15}\]
\[+(2 - b)^3\{(b - 2)(\frac{b}{20} + \frac{1}{15} + \frac{a}{12} + \frac{\tau}{6}) + \frac{\tau}{6}(\tau + 4) + \frac{a}{2}\}.
\]

9. If \(0 < a + \tau - 1 < a + b + \tau - 2 < a \leq 1\) is satisfied, for \(n = 3\), then
\[
\pi_{ab,9} = -\frac{t}{6}\{(a - 3)^3 + (b - 3)^3 + 54\} - \frac{(2 - b)^3}{6} - \frac{(2 - a)^3}{6} + \frac{b}{6}(1 - a)^3
\]
\[-\frac{1}{3}(t - \frac{3}{2})^{4} + \frac{t^2}{6}(a + b)(18 - 4t) - \frac{ab}{2}(a + b - 6) - \frac{t^2}{2}(a + b)^2 + \frac{43}{16},
\]
and for \(n \geq 4\),
\[
\pi_{ab,9} = (2 - a)^3\{(a - 2)(\frac{a}{20} + \frac{1}{15} + \frac{b}{12} + \frac{\tau}{6}) + \frac{2\tau}{3} + \frac{b}{2} + \frac{(2 - b)^3(\tau - 1)}{6}\]
\[+(\tau - 2)^3\{((\tau - 2)(\frac{\tau - 7}{30} + \frac{a + b}{12}) + \frac{(1 - a)(1 - b)}{6} - \frac{2}{3}\}
\[-\frac{\tau}{6}\{3b[a(b - 1) + b(\tau - 1)] + a^3(b + \tau) + 4\tau + 11\}
\[-\frac{a}{2}\{b(2a - 9) + (\tau - 4)^2 - \frac{65}{6}\} + \frac{3}{5} + b(\frac{7\tau}{3} - \frac{21}{4}).
\]

10. When \(0 \leq a + \tau - 1 < a \leq a + b + \tau - 2 < 1\) is satisfied, for \(n = 3\), we get
\[
\pi_{ab,10} = \pi_{ab,9} + \frac{x_{10}^2}{6}(1 + a - b - x_{10}),
\]
where \(x_{10} = a + \tau - 1\), and for \(n \geq 4\), we have
\[
\pi_{ab,10} = \pi_{ab,9} + \frac{x_{10}^2}{6}\{a(2 - b) - .5(a + 2 - b)x_{10} + .2x_{10}^2\}.
\]

11. If \(0 \leq a + b + \tau - 3 < a + \tau - 1 < a \leq 1\) and \(n \geq 4\), then
\[
\pi_{ab,11} = (b - 2)^3\{\frac{(b - 2)^2}{20} - a\frac{(b - 4)}{12} + \frac{(a + \tau - 1)(b + \tau - 3)}{6}\} + a(1 - \frac{a^2}{6}).
\]
12. In this case, increase \( a \) and select \( b \) such that \( 0 \leq a - 1 \leq a + \tau - 1 \leq a + b + \tau - 2 \leq 1 \). For \( n = 3 \), we have

\[
\pi_{ab,12} = z_{12} + \frac{\tau}{6} \{(3-a-b-\tau)\tau^2 + (3-a-b-\tau)^3\},
\]

where \( z_{12} = 1 - \frac{(2-a)^3}{6} - \frac{(2-b)^3}{6} \), and for \( n \geq 4 \),

\[
\pi_{ab,12} = z_{12} + \frac{\tau}{6} \{(2-a)(2-b)\tau^2 + (3-a-\tau-b)^3\} - \frac{\tau^4}{12}(4-a-b) + \frac{\tau^5}{30}.
\]

Again for the cases 13 and 14, \( a \) is held fixed similar to case 12 while \( b \) is increased until a new restriction is satisfied. Also under this new restriction a new form of the pairwise inclusion probability is calculated.

13. If \( 0 < a + b + \tau - 3 < a - 1 < a + \tau - 1 < 1 \) and \( n \geq 4 \), then

\[
\pi_{ab,13} = z_{12} + \tau^3 \left\{ \frac{(a+b)(\tau-4)}{12} + \frac{(\tau-5)^2}{30} + \frac{(ab-1)}{6} \right\}.
\]

14. If \( 0 < a - 1 < a + b + \tau - 3 < a + \tau - 1 < 1 \) and \( n \geq 4 \), then

\[
\pi_{ab,14} = z_{12} + \frac{(b-2)^3}{6} \left\{ \tau(a+b+\tau-4) + \frac{(b-2)(5a+3b-16)}{10} \right\}.
\]

15. In this case, increase \( a \) and select \( b \) such that \( 0 \leq a + \tau - 2 \leq a - 1 \leq a + b + \tau - 3 \leq 1 \) is satisfied. For \( n \geq 4 \),

\[
\pi_{ab,15} = \frac{(b-2)^3}{6} \left\{ z_{10}(b+\tau-3) + \frac{a}{2}(4-b) \right\} + \frac{ab(z_{10}-1)^3}{6} + \frac{(b-2)^5}{20} + \frac{(a-2)^3}{6} + \frac{z_{10}}{6} \left\{ z_{10}(9a+3b-10) + (11-8a-4b) \right\} + z_{10}^2 \left\{ -\frac{z_{10}}{30} + \frac{z_{10}}{12}(a-b+2) + \frac{(1-2a)}{3} \right\} + \frac{(5a+3b)}{12} + \frac{11}{30}.
\]

62
16. Hold \(a\) fixed and increase \(b\) until \(0 \leq a + b + \tau - 4 \leq a + \tau - 2 \leq a - 1 \leq 1\) is satisfied. For \(n \geq 5\),

\[
\pi_{ab,16} = z_{12}.
\]

For the cases 17 – 32 we can obtain the restrictions by a similar way similar to cases 1 – 16 respectively but with \(1 < \tau < 2\). Under each of these restrictions we provide the formula of the pairwise inclusion probability and we show the calculation of one of them.

17. If \(0 < a \leq a + \tau - 1 \leq a + b + \tau - 1 \leq 1\) and \(n \geq 3\), then

\[
\pi_{ab,17} = ab\{1 - \frac{(a^2 + b^2)}{6}\}.
\]

18. If \(0 < a + b + \tau - 2 \leq a \leq a + \tau - 1 \leq 1\) and \(n \geq 4\), then

\[
\pi_{ab,18} = \tau - 2^3 \{ (\tau - 2)\left(\frac{\tau}{30} - \frac{1}{15} + \frac{a}{12} + \frac{b}{12}\right) + \frac{ab}{6}\} - \frac{\tau^2}{6}(a^3 + b^3)
\]

\[
+ (2-a)^3\{ (a-2)(\frac{a}{20} + \frac{1}{15} + \frac{b}{12} + \frac{\tau}{6}) + \frac{2}{3}\tau\}\n\]

\[
+ (2-b)^3\{ (b-2)(\frac{b}{20} + \frac{1}{15} + \frac{a}{12} + \frac{\tau}{6}) + \frac{2}{3}\tau\}\n\]

\[
- \frac{ab}{6}\{(a^2 + b^2)(\tau + 3) + 2\} + \frac{2}{3}(a + b - 2)(4\tau + 3ab) + \frac{32}{15}.
\]

19. If \(0 \leq a \leq a + b + \tau - 2 \leq a + \tau - 1 \leq 1\) and \(n \geq 4\), then

\[
\pi_{ab,19} = \frac{a^3}{12}\left(\frac{b}{20}(2-a) - \frac{\tau}{6}(a + b + \tau - 4) - \frac{a^2}{20} - \frac{a}{3} + \frac{2}{3}\right) + \frac{ab}{6}(6 - b^2).
\]

20. If \(0 \leq a \leq a + \tau - 1 \leq a + b + \tau - 2 \leq 1\) and \(n \geq 4\), then

\[
\pi_{ab,20} = \frac{a^3}{12}(2-a) - \frac{\tau}{6}(a + b + \tau - 4) - \frac{a^2}{20} - \frac{a}{3} + \frac{2}{3}\right) + \frac{a}{6}(b - 2)^3 + 6).
\]
21. If \( 0 \leq a + b + \tau - 3 \leq a \leq a + \tau - 1 \leq 1 \) and \( n \geq 5 \), then

\[
\pi_{ab,21} = \pi_{ab,10} \quad \text{but } x_{10} \text{ replaced by } x_{21} = 3 - b - \tau,
\]

where \( \pi_{ab,10} \) is the formula in case 10 associated with \( n \geq 4 \).

22. If \( 0 \leq a \leq a + b + \tau - 3 \leq a + \tau - 1 \leq 1 \) and \( n \geq 5 \), then

\[ \pi_{ab,22} = \pi_{ab,6} \]

23. If \( 0 \leq a + \tau - 2 \leq a \leq a + b + \tau - 2 \leq 1 \) and \( n \geq 4 \), then

\[
\pi_{ab,23} = \pi_{ab,17} + (2 - \tau)^3\left\{ (\tau - 2)\left( \frac{\tau}{30} - \frac{1}{15} + \frac{a + b}{12} \right) + \frac{ab}{6} \right\}
\]

24. If \( 0 \leq a + b + \tau - 3 \leq a + \tau - 2 \leq a \leq 1 \) and \( n \geq 5 \), then

\[
\pi_{ab,24} = (2 - a)^3\left\{ (a - 2)\left( \frac{a}{20} + \frac{b}{12} - \frac{1}{10} \right) + \frac{\tau(x_{24} + \tau)}{6} + \frac{b}{3} - \frac{1}{2} \right\}
\]

\[
+ (2 - b)^3\left\{ (b - 2)\left( \frac{b}{20} + \frac{a}{12} - \frac{1}{10} \right) + \frac{\tau(x_{24} + \tau)}{6} + \frac{a}{3} - \frac{1}{2} \right\}
\]

\[
+ \frac{(2 - \tau)^3}{6}\left\{ (\tau + 2(x_{24} - 2)) + x_{24}\left\{ (\tau - 1)^2\left( \frac{5}{2} - x_{24} \right) - \frac{37}{12} \right\}
\]

\[
+ \frac{ab}{2}\left\{ \tau(3\tau - 2x_{24} + 5) + 5 \right\} - 3\tau(\tau - \frac{1}{2}) + \frac{31}{10},
\]

where \( x_{24} = a + b \).

25. If \( 0 \leq a + \tau - 2 \leq a + b + \tau - 3 \leq a \leq 1 \) and \( n \geq 5 \), then

\[
\pi_{ab,25} = (2 - a)^3\left\{ (a - 2)\left( \frac{a}{20} + \frac{b}{6} + \frac{1}{10} \right) + \frac{1}{6}\left\{ (\tau + 1)^2 + b(\tau + 2) - 4 \right\} \right\}
\]

\[
- (2 - b)^3\left\{ (b - 2)\left( \frac{b}{20} + \frac{a}{6} + \frac{4}{15} \right) + \frac{a}{6} + \frac{(\tau - 1)^2}{6} \right\}
\]

\[
+ \frac{(2 - \tau)^3}{6}\left\{ \tau + 2(a + b - 2) \right\} + \frac{1}{6}(\tau(15a - 3b - 4) - 16a + 2b - 13)
\]

\[
- b\left\{ \frac{\tau}{2}(\tau + 2a - 7) - \frac{1}{6} \right\} + \frac{11}{4} - a(\tau - 1)^2 + \frac{1}{12} + \frac{21}{5}.
\]

64
26. If $0 < a + \tau - 2 \leq a \leq a + b + \tau - 3 \leq 1$ and $n \geq 5$, then

$$\pi_{ab,26} = \pi_{ab,10}, \text{ with } \tau \text{ replaced by } \tau - 1,$$

where $\pi_{ab,10}$ is the formula in case 10 associated with $n \geq 4$.

27. If $0 < a + b + \tau - 4 \leq a + \tau - 2 \leq a \leq 1$ and $n \geq 6$, then

$$\pi_{ab,27} = \pi_{ab,11}, \text{ with } \tau \text{ replaced by } \tau - 1.$$

28. If $0 < a - 1 \leq a + \tau - 2 \leq a + b + \tau - 3 \leq 1$ and $n \geq 5$, then

$$\pi_{ab,28} = (a - 2)^3 \{(a - 2)(\frac{a}{20} + \frac{b}{12} + \frac{1}{15}) + \frac{z_{28}}{6}(\tau - 3) + \frac{(1 + 2b)}{6}\}$$

$$+ (b - 2)^2 \{(b - 2)(\frac{b}{20} + \frac{a}{12} + \frac{1}{15}) + \frac{z_{28}}{6}(\tau - 3) + \frac{(1 + 2a)}{6}\}$$

$$+ (a + b)(\frac{z_{28}^2}{3} - \frac{z_{28}^2}{2} - \frac{2z_{28}}{3} + \frac{13}{12}) - \frac{ab}{6} \{3z_{28}(z_{28} - 3) + 7\}$$

$$- \frac{z_{28}}{6} \{z_{28}^2(2z_{28} + 2) + 5 - 10z_{28}\} + \frac{3}{10},$$

where $z_{28} = a + b + \tau - 3$.

29. If $0 < a + b + \tau - 4 \leq a - 1 \leq a + \tau - 2 \leq 1$ and $n \geq 6$, then

$$\pi_{ab,29} = \pi_{ab,13}, \text{ with } \tau \text{ replaced by } \tau - 1.$$

30. If $0 < a - 1 \leq a + b + \tau - 4 \leq a + \tau - 2 \leq 1$ and $n \geq 6$, then

$$\pi_{ab,30} = \pi_{ab,14}, \text{ with } \tau \text{ replaced by } \tau - 1.$$
31. If $0 \leq a + \tau - 3 \leq a - 1 \leq a + b + \tau - 4 \leq 1$ and $n \geq 6$, then

$$\pi_{ab,31} = \pi_{ab,15}, \text{ with } \tau \text{ replaced by } \tau - 1.$$ 

32. If $0 \leq a + b + \tau - 5 \leq a + \tau - 3 < a - 1 \leq 1$ and $n \geq 6$, then

$$\pi_{ab,32} = \pi_{ab,16}.$$ 

Finally, we present two trivial cases and explain how to calculate the pairwise inclusion probability when $2 \leq \tau$.

33. If $0 < a < 2$ and $2 \leq b$, then

$$\pi_{ab,33} = \pi_a.$$ 

34. If $2 \leq a$ and $2 \leq b$, then

$$\pi_{ab,34} = 1.$$ 

Note that the cases 33 and 34 have no restrictions on $\tau$. Further, from the arguments of generating the above cases we can state that, for any given $a$, $b$, $\tau$, and $n$, the $\pi_{ab}$ can be calculated by one of the formulas introduced in cases 1 — 34, except when $0 < a < b < 2$ and $2 \leq \tau$. We now show how to calculate the $\pi_{ab}$ under these restrictions.

**Theorem 3.4** For given $a$, $b$ and $\tau$ such that $0 < a < b < 2$ and $2 \leq \tau$, let 

$$\tau_r = \tau \mod (1) + 1.$$ 

Then, $a$, $b$ and $\tau_r$ will satisfy only one of the restrictions provided in cases 17 — 32 and the pairwise inclusion probability can be given by the associated formula with $\tau$ replaced by $\tau_r$. 

66
Proof: It is clear that $1 < \tau < 2$. Then, by using $\tau_r$ in place of $\tau$ in the restrictions provided in cases 17 – 32, one of these cases, say the $j^{th}$ case, is satisfied. For this case, we can calculate the pairwise inclusion probability using arguments similar to cases 1 and 8. Then, we obtain the explicit form of $\pi_{ab;j}$ with $\tau$ replaced by $\tau_r$. $\square$

Finally, we can state that for any given $a$, $b$, $\tau$ and $n \geq 2$ we can calculate the pairwise inclusion probability by using one of the cases 1 – 34 or by using the results presented in Theorem 3.4.

In the next subsection, we derive explicit formulas of the covariance of $(t_i, t_j)$ in terms of $a$, $b$, $\tau$, $\tau'$ and $n$.

### 3.4.2 Calculations of the Covariance Function

For any two units $i, j$ in the population, the value of $\text{Cov}(t_i, t_j)$ is needed for the calculations of the variance estimates of the total estimator. Again, suppose $a$, $b$, $\tau$, $\tau'$, and the grid are on the real number line as defined in section 3.4.1. Moreover, we denote $t_i$ by $t_a$ and $t_j$ by $t_b$. In this mode of $a$, $b$, $\tau$ and $\tau'$, the grid spacing is of length one and the inclusion density functions are just the ones presented in (2.17) and (2.29) and we can write them as

\[
\pi(c) = 1 \quad \text{where } 0 < c \leq \frac{\Pi d}{\lambda} = n \quad (3.16)
\]

and

\[
\pi(c, c') = \begin{cases} 
\delta(c, c') & \text{if } \delta(c, c') \leq 1 \\
1 & \text{if } \delta(c, c') > 1,
\end{cases} \quad (3.17)
\]
where $\delta(c, c') = \min\{|c - c'|, \frac{n_k}{\delta} - |c - c'|\}$. Additionally, because $\text{Cov}(t_a, t_b) = \text{Cov}(t_b, t_a)$, it is sufficient to consider the cases for which $0 < a < b$. The following theorem can be useful for calculating these covariances.

**Theorem 3.5** The covariance of $(t_a, t_b)$ and the variance of $t_a$ are given by

$$
\text{Cov}(t_a, t_b) = \int_0^a \int_{a+r}^u \left( \pi(c, c') - \pi(c)\pi(c') \right) dc dc'
$$

and

$$
\text{Var}(t_a) = \int_0^a \pi(c) dc + 2 \int_0^a \int_{c'}^u \left( \pi(c, c') \right) dc dc' - a^2.
$$

where $\pi(c)$ and $\pi(c, c')$ are given in (3.16) and (3.17) respectively and $u = a + r + b$.

**Proof:** Suppose the interval $[0, u]$ is partitioned into $M_3$ equal subintervals of length $\Delta = \frac{u}{M_3}$. Also assume, without loss of generality, that $\Delta$ divides $[0, a), [a, a + r), [a + r, u]$ into $M_1, M_2 - M_1$ and $M_3 - M_2$ intervals, respectively. For sufficiently small $\Delta$, none of these $M_3$ intervals of size $\Delta$ will contain more than one sample point and we define

$$
\delta(\Delta_k) = \begin{cases} 
1 & \text{if the } k\text{th interval in the partition contains one sample point} \\
0 & \text{otherwise},
\end{cases}
$$

where $k = 1, \ldots, M_1, M_1 + 1, \ldots, M_2, M_2 + 1, \ldots, M_3$.

Now,

$$
\text{Cov}(t_a, t_b) = \text{Cov} \left( \sum_{i=1}^{M_1} \delta(\Delta_i), \sum_{j=M_2+1}^{M_3} \delta(\Delta_j) \right)
$$

$$
= \sum_{i=1}^{M_1} \sum_{j=M_2+1}^{M_3} \text{Cov} (\delta(\Delta_i), \delta(\Delta_j)).
$$
By taking the limits, as $M_3 \to \infty$, of both sides of the above equation, we obtain

$$\text{Cov}(t_a, t_b) = \int_0^a \int_{a+\tau}^a (\pi(c,c') - \pi(c)\pi(c')) \, dc \, dc'. $$

The variance of $t_a$ can be written as

$$\text{Var}(t_a) = E(t_a^2) - E(t_a)^2 = E(t_a^2) - a^2.$$

Partitioning the interval $[0, a]$ into small subintervals and proceeding similarly as above, we obtain the form given in (3.19). □

Now by using (3.18) and (3.19), we can calculate explicit formulas for the covariance and variance which are as follows:

1. When $0 < a + b + \tau < 1$ is satisfied and $n \geq 2$, using (3.16) and (3.17) in (3.18) gives

$$\text{Cov}(t_a, t_b) = \int_0^a \int_{a+\tau}^{a+b+\tau} (c - c') \, dc \, dc' - ab.$$ 

This leads to

$$\text{Cov}(t_a, t_b) = -\frac{1}{6} \{ (1 - \tau)^3 - (1 - a - \tau)^3 - (1 - b - \tau)^3 + (1 - a - b - \tau)^3 \}.$$

Note that in this case $\tau' \geq 1$.

2. When $0 \leq a + b + \tau - 1 \leq a \leq a + \tau \leq 1$ is satisfied and $\tau' < 1$, then

$$\text{Cov}(t_a, t_b) = \int_{a+\tau}^{a+b+\tau-1} (\int_{a+\tau}^{c+1} (c - c') \, dc + \int_{c+1}^{a+b+\tau} (n - c' + c) \, dc) \, dc' \int_{a+\tau}^{a+b+\tau-1} (c - c') \, dc \, dc' - ab.$$
By calculating these integrals we obtain
\[
Cov(t_a, t_b) = -\frac{1}{6} \{(1 - \tau)^3 - (1 - a - \tau)^3 - (1 - b - \tau)^3 - (n - \tau' - 1)^3(2n + \tau' - 5)\}.
\]

In this case \( n = 2 \) because \( a + b + \tau + \tau' = n \) and \( n \geq 2 \).

3. When \( 0 \leq a + b + \tau - 1 \leq a \leq a + \tau \leq 1 \leq \tau' \) is satisfied, we get
\[
Cov(t_a, t_b) = \int_0^{a+b+\tau-1} \left( \int_{a+\tau}^{c'+1} (c - c') \, dc + \int_{c'+1}^{a+b+\tau} dc' \right) dc' + \int_{a+b+\tau-1}^2 \int_{a+\tau}^{a+b+\tau} (c - c') \, dcdc' - ab.
\]

On calculating these integrals we obtain
\[
Cov(t_a, t_b) = -\frac{1}{6} \{(1 - \tau)^3 - (1 - a - \tau)^3 - (1 - b - \tau)^3\}.
\]

4. When \( 0 \leq a + \tau \leq a + \tau' \leq 1 \leq b + \tau \) is satisfied and, using an argument similar to the one above, we obtain
\[
Cov(t_a, t_b) = -\frac{1}{6} \{(1 - \tau)^3 - (1 - a - \tau)^3 + (1 - \tau')^3 - (1 - a - \tau')^3\}.
\]

5. When \( 0 \leq a + \tau \leq 1 \leq b + \tau \) and \( 0 \leq \tau' \leq 1 \leq a + \tau' \) are satisfied, then
\[
Cov(t_a, t_b) = -\frac{1}{6} \{(1 - \tau)^3 - (1 - a - \tau)^3 + (1 - \tau')^3\}.
\]

6. When \( 0 \leq a + \tau \leq 1 \leq b + \tau \) and \( \tau' \geq 1 \) are satisfied, then using an argument similar to the one above, we obtain
\[
Cov(t_a, t_b) = -\frac{1}{6} \{(1 - \tau)^3 - (1 - a - \tau)^3\}.
\]
7. When \( 0 \leq \tau \leq \tau' < 1 \leq a + \tau \leq b + \tau \) is satisfied, then it can be shown that

\[
\text{Cov}(t_a, t_b) = -\frac{1}{6} \{ (1 - \tau)^3 + (1 - \tau')^3 \}.
\]

8. When \( 0 < \tau < 1 \leq a + \tau \leq b + \tau \) and \( \tau' \geq 1 \) are satisfied, then it can be shown that

\[
\text{Cov}(t_a, t_b) = -\frac{1}{6} \{ (1 - \tau)^3 \}.
\]

9. When \( \tau' \geq \tau \geq 1 \), then it can be shown that

\[
\text{Cov}(t_a, t_b) = 0.
\]

Actually, for any given \( a, b, \tau \) and \( \tau' \) such that \( a \leq b \) and \( \tau \leq \tau' \), one of the above nine restrictions will be satisfied and the associated covariance formula can be used to calculate \( \text{Cov}(t_a, t_b) \).

Also by using (3.19) we can calculate the variance of \( t_a \) in terms of \( a \) which is given by

\[
\text{Var}(t_a) = \begin{cases} 
    a(1 - a + \frac{a^2}{3}) & \text{if } a < 1 \\
    \frac{1}{3} & \text{if } a \geq 1.
\end{cases}
\]

In the next section we investigate the sampling design, presented in this chapter, using the approach of sampling from a continuous universe.
3.5 Estimation Using Sampling from a Continuous Universe

Given the sampling design which we are investigating in this chapter, the problem of estimation can be approached by thinking as though we are sampling the points from a continuous universe in one dimension. In this setup, for a given point \( c \) falling in the range of \( x_j \) we let \( y(c) = y_j \) and \( z(c) = x_j \) and we define the response function at \( c \) by

\[
z(c) = \frac{y(c)}{z(c)} = \frac{y_j}{x_j}.
\]

Therefore, in the presented Numerical Example in section 3.3 the sampled response variable at \( c_1, c_2, c_3, \) and \( c_4 \) is given by

\[
z(c_1) = \frac{y_3}{x_3}, \quad z(c_2) = \frac{y_6}{x_6}, \quad z(c_3) = \frac{y_6}{x_6} \quad \text{and} \quad z(c_4) = \frac{y_1}{x_1}.
\]

Actually the population total is just the integral of \( z \) over \( c \) where \( 0 < c < \Pi_N \), that is

\[
T = \int_0^{\Pi_N} z(c) \, dc = \sum_{i=1}^{N} \int_{\Pi_{i-1}}^{\Pi_i} \frac{y_i}{x_i} \, dc = \sum_{i=1}^{N} y_i.
\]

By using the theory of sampling from a continuous universe, we show that the estimator of population total \( \hat{T}_1 \) given in (3.5) is just the Horvitz-Thompson estimator of \( T \). Also we provide new variance estimators of the variance of \( \hat{T}_1 \).

Now, the grid spacing is of length \( d \) and the sampling design, that we are dealing with, is the same that is presented in the previous chapter with inclusion density
functions given in (3.20) and (3.21). These inclusion density functions can be written as

\[ \pi(c) = \frac{1}{d}, \quad 0 < c < \Pi_N \quad (3.20) \]

and

\[ \pi(c, c') = \begin{cases} \frac{\delta(c, c')}{d^2} & \text{if } \delta(c, c') \leq d \\ \frac{1}{d^2} & \text{if } \delta(c, c') > d. \end{cases} \quad (3.21) \]

where here \( \delta(c, c') = \min\{|c - c'|, \Pi_N - |c - c'|\} \).

The estimator of the population total can be calculated as

\[ \hat{T}_1 = \sum_{i=1}^{n} \frac{z(c_i)}{\pi(c_i)} = d \sum_{i=1}^{n} \frac{y(c_i)}{x(c_i)} = d \sum_{j=1}^{N} t_j \frac{y_j}{x_j}. \]

This shows that \( \hat{T}_1 \) is just the Horvitz-Thompson estimator under the sampling design from a continuous universe.

By substituting (3.20) and (3.21) in (1.14), we obtain a new expression for the variance of \( \hat{T}_1 \). Moreover (1.15) provides a new unbiased estimator of the variance of \( \hat{T}_1 \). This new estimator depends on the location of the sampled points \( (c_i) \) through the inclusion density functions rather than the sampled units \( (x_i) \) as in the previous estimators and is given by

\[ v_5(\hat{T}_1) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{\pi(c_i)\pi(c_j) - \pi(c_i, c_j)}{\pi(c_i, c_j)} \right) \left( \frac{y(c_i)}{x(c_i)} - \frac{y(c_j)}{x(c_j)} \right)^2. \quad (3.22) \]
This variance estimator shall be compared with the estimators given in (3.10), (3.11), (3.12) and (3.13).

3.6 Comparison of $\hat{T}_1$ and $\hat{T}_2$

We have conducted numerical simulations to investigate the behavior of both estimators $\hat{T}_1$ and $\hat{T}_2$ which are based on the fixed order stratified sampling scheme. For many populations which we have simulated and for different sample sizes, the variance of $\hat{T}_1$ was always smaller than the variance of $\hat{T}_2$. One of the reasons for this result could be that $\hat{T}_1$ is based on a fixed sample size while $\hat{T}_2$ is based on a random sample size. In light of this result, we investigate the properties of $\hat{T}_1$ only.

3.7 Comparison of the Variance Estimators of $\hat{T}_1$

We have presented five variance estimators of $\hat{T}_1$ and we have denoted them by $v_1$, $v_2$, $v_3$, $v_4$, and $v_5$. Now some properties related to the first four of these variance estimators can be concluded from the following remark.

REMARK 3.6 $Cov(t_i, t_j) \leq 0$ for any two units $i$ and $j$, $i \neq j$, in the population.

Proof: From Theorem 3.5 we have

$$Cov(t_a, t_b) = \int_0^a \int_{a+r}^u (\pi(c, c') - \pi(c)r(c')) \, dc \, dc',$$

(3.23)

where $u = a + \tau + b$. Also from the inclusion density functions which are given in (3.16) and (3.17), we conclude that

74
Finally, (3.23) and (3.24) imply that $\text{Cov}(t_i, t_j) \leq 0$. □

The estimators $\tilde{v}_1$ and $\tilde{v}_2$ are the Horvitz-Thompson type estimators of the variance of $\hat{T}_1$. From the form of these two estimators and Remark 3.6, it can be seen that both $\tilde{v}_1$ and $\tilde{v}_2$ can take negative values. Therefore, they are not practical estimators.

The estimators $\tilde{v}_3$ and $\tilde{v}_4$ are the Yates-Grundy type estimators of the variance of $\hat{T}_1$. From the form of these estimators and Remark 3.6, it can be seen that both $\tilde{v}_3$ and $\tilde{v}_4$ can take only nonnegative values. Therefore, these estimators are more stable than $\tilde{v}_1$ and $\tilde{v}_2$.

The estimator $\tilde{v}_5$ takes only nonnegative values. However, it is not difficult to see that it makes use of less information from the data than $\tilde{v}_3$ and $\tilde{v}_4$. This is true because it depends on the selected points rather than the selected units. Note that $\tilde{v}_5$ has much simpler form than $\tilde{v}_3$ and $\tilde{v}_4$.

We have conducted numerical simulations which confirmed that $\tilde{v}_3$ and $\tilde{v}_4$ are the most stable among the five estimators of the variance. Also our simulations confirmed that when the ratio of sample size to population size is small, say less than .4, then $\tilde{v}_3$ and $\tilde{v}_4$ almost equal each other. On the other hand, when this ratio is high, say more than .4, then $\tilde{v}_3$ performs better than $\tilde{v}_4$. 

\begin{equation}
(\pi(c, c') - \pi(c)\pi(c')) \leq 0. \tag{3.24}
\end{equation}
3.8 Comparison of Fixed Order Stratified Sampling Procedure to Other Sampling Procedures

An empirical study is conducted to compare the sampling variance and the efficiency of the current sampling procedure to two different pps sampling procedures which allow for unbiased variance estimation. These three procedures are compared using populations simulated from three different models which reflect different patterns of variability. The sampling procedures are the following:

1. The fixed order stratified sampling procedure is the one introduced and analyzed earlier in this chapter. From this procedure, we consider the estimator $\hat{T}_1$ and its sampling variance, given in (3.9). Note that this sampling variance depends on the order of the population units.

2. The random order pps systematic sampling procedure is presented earlier in this chapter. In relation to this chapter we consider the Hartley and Rao approximate variance formula given in (3.2). Note that this sampling procedure assumes that the population units are in a random order. Therefore the approximate variance formula is independent of the order of the population units.

3. The pps with replacement sampling procedure where the estimator of the population total and its variance are given by

$$\hat{T}_3 = \sum_{i=1}^{n} \frac{y_i}{\pi_i},$$  \hspace{1cm} (3.25)
and

\[ \text{Var}(\bar{T}_3) = \sum_{i=1}^{N} \pi_i \left( \frac{y_i}{\pi_i} - \frac{T_y}{n} \right)^2 = \frac{1}{n} \sum_{i=1}^{N} \sum_{j>i}^{N} \pi_i \pi_j \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2, \]

(3.26)

where \( \pi_i = n \frac{\pi_i}{X} \) and \( X = \sum_{i=1}^{N} x_i \). Also this sampling procedure and the variance formula are independent of the order of the population units.

Note that in all of the three procedures, if \( y \) is proportional to \( x \), then

\[ \left( \frac{y_i}{x_i} - \frac{y_j}{x_j} \right) = 0, \]

for all \( i \neq j \). Therefore, in such a case the three sampling variances assume zero values. Thus, we expect that the three sampling designs should perform very well when \( y \) is proportional to \( x \) plus a stochastic term.

**Peak Ordering.**

We now present an efficient arrangement of the population units in a certain order for which the fixed order stratified sampling method results in a small sampling error. The arrangement can be explained as follows:

Arrange the population units in an increasing order of \( y/x \). After arranging the \( N \) units of \( y/x \), if \( N \) is odd, say \( N = 2M + 1 \), then use the ordering

1, 3, 5,..., \( 2M - 1 \), \( 2M + 1 \), \( 2M \), \( 2M - 2 \),..., 4, 2,

and if \( N \) is even, say \( N = 2M \), then use the ordering

1, 3, 5,..., \( 2M - 1 \), \( 2M \), \( 2M - 2 \),..., 4, 2.
In this ordering the values of $y/x$ increase from the smallest $y/x$ to the peak value and then decrease back to the second smallest $y/x$. This order of the population units shall be referred to as peak ordering.

The populations were simulated from three different models. These models, associated with explanations regarding the simulation methods, are introduced in the following three groups of simulations.

3.8.1 Group I Simulations

We investigated populations generated by the first order autoregressive model with normally distributed errors,

$$z_i - \mu = \rho(z_{i-1} - \mu) + e_i,$$  \hspace{1cm} (3.27)

where $\rho$ is the first order autocorrelation, $\mu$ is the mean of the process, and

$$z_i = \frac{y_i}{x_i}.$$  

Actually many natural populations have properties close to the populations generated by the model given in (3.27). Each population of $(z_i, y_i)$ is simulated as follows. For each given value of $\rho$, we fixed $z_0 = 15$ and $\mu = 15$. Also $e_i$ is assumed to be a normal variate with mean 0 and variance 1. Then, we simulated $z_1, z_2, ..., z_{60}$ using model (3.27). Also we selected $z_i$ as a discrete uniform variate from (2, 3, ..., 9) and $y_i = z_i x_i$ for $i = 1, 2, ..., 60$. To remove the effects of the initial value $z_0$, the population units are the set of 40 values of $(x_i, y_i)$, $i = 21, 22, ..., 60$.  

78
Now, for $\rho = .9, .7, .5, .2, .05, 0$, we simulated 100 populations using the model given in (3.27), as explained above. Also, from each simulated population, two samples one of size 4 and the other of size 8 are selected using each of the three sampling procedures presented above. For each of the simulated population, sample size, and sampling procedure, we calculated the sampling variance of the population total estimator. The units were kept in their simulated order for the calculations of the variances of the estimators of the population totals for the fixed order stratified sampling method. Then, in Table 3.2 we presented the averages of these sampling variances calculated from the 100 simulated populations. Also, in Table 3.2 we bracketed the efficiencies, where each efficiency is the ratio of the average variances of the fixed order stratified sampling scheme to the average variances of the selected pps sampling scheme.

From Table 3.2 we observe the following concerning the performance of the three sampling procedures. Both sampling methods, the fixed order stratified sampling and the random order systematic sampling, always perform better than the pps with replacement sampling method. Now, we compare the first two sampling procedures by considering the following two cases:

**Case 1:** Applicable for values of $\rho$ approximately given by $0.1 \leq \rho < 1$.

The fixed order stratified sampling is more efficient than the random order systematic sampling. It can also be noted that this efficiency increases when $\rho$ increases towards 1 and it also increases when the sample size increases. In our simulations, we fixed $\mu = 15$ and the variance of $e_i$'s at one. However, we simulated populations
for different values of $\mu$ and different variances of $e_i$'s, and found that the efficiency level sometimes increased and sometimes decreased. However, we noted that when $p$ is more than .15, the fixed-order stratified sampling was always more efficient than the random-order systematic sampling.

**Case 2**: Applicable for small values of $p$ approximately given by $0 \leq p \leq .1$.

The random order systematic sampling procedure is more efficient than the fixed order stratified sampling procedure. Specifically, when $p = 0$, the units of any simulated population are in a random order. In such case, the random order systematic sampling procedure is expected to be more efficient than the fixed order stratified sampling procedure, and this has been confirmed by our simulations. We simulated many populations for which $p = 0$, for different values of $\mu$, and different variances of $e_i$'s, and we noted that the precision gained by using random order pps systematic sampling over fixed order stratified sampling does not increase beyond 10%.

Also recall that the new sampling procedure still has an advantage over the random order systematic sampling procedure because of the availability of its nonnegative variance estimators.
Table 3.2: Empirical average of the sampling variances of the estimators of the population totals and the efficiencies (in brackets) as explained in Group I Simulations.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Sampling Procedure</th>
<th>fixed order stratified sampling</th>
<th>random order ppsys sampling</th>
<th>pps with replacement sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>p = .9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>fixed order</td>
<td>20182</td>
<td>35028</td>
<td>38506</td>
</tr>
<tr>
<td></td>
<td>stratified sampling</td>
<td>(1)</td>
<td>(.57)</td>
<td>(.52)</td>
</tr>
<tr>
<td>8</td>
<td>random order</td>
<td>5725</td>
<td>14616</td>
<td>18483</td>
</tr>
<tr>
<td></td>
<td>ppsys sampling</td>
<td>(1)</td>
<td>(.39)</td>
<td>(.31)</td>
</tr>
<tr>
<td>p = .7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>fixed order</td>
<td>13808</td>
<td>18302</td>
<td>20139</td>
</tr>
<tr>
<td></td>
<td>stratified sampling</td>
<td>(1)</td>
<td>(.75)</td>
<td>(.68)</td>
</tr>
<tr>
<td>8</td>
<td>random order</td>
<td>4875</td>
<td>8330</td>
<td>10527</td>
</tr>
<tr>
<td></td>
<td>ppsys sampling</td>
<td>(1)</td>
<td>(.58)</td>
<td>(.46)</td>
</tr>
<tr>
<td>p = .5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>fixed order</td>
<td>11957</td>
<td>13645</td>
<td>15005</td>
</tr>
<tr>
<td></td>
<td>stratified sampling</td>
<td>(1)</td>
<td>(.88)</td>
<td>(.8)</td>
</tr>
<tr>
<td>8</td>
<td>random order</td>
<td>4310</td>
<td>5558</td>
<td>7039</td>
</tr>
<tr>
<td></td>
<td>ppsys sampling</td>
<td>(1)</td>
<td>(.78)</td>
<td>(.61)</td>
</tr>
</tbody>
</table>
Table 3.2 (continued): Empirical average of the sampling variances of the estimators of the population totals and the efficiencies (in brackets) as explained in Group I Simulations.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Sampling Procedure</th>
<th>fixed order stratified sampling</th>
<th>random order pps systematic sampling</th>
<th>pps with replacement sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \rho = .2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>11009 (1)</td>
<td>11432 (.96)</td>
<td>12570 (.88)</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>4668 (1)</td>
<td>4901 (.95)</td>
<td>6208 (.75)</td>
</tr>
<tr>
<td>( \rho = .05 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>10810 (1)</td>
<td>10901 (.99)</td>
<td>11986 (.90)</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>4742 (1)</td>
<td>4697 (1.01)</td>
<td>5958 (.79)</td>
</tr>
<tr>
<td>( \rho = 0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>10250 (1)</td>
<td>100267 (.100)</td>
<td>11288 (.91)</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>4711 (1)</td>
<td>4590 (1.03)</td>
<td>5806 (.81)</td>
</tr>
</tbody>
</table>
3.8.2 Group II Simulations

This group of simulations is obtained from the model given by

\[ z'_i = z_i + i\alpha, \]  

(3.28)

where \( z_i \) satisfy the stationary model given in (3.27), \( \alpha \) is the slope of the process \( z'_i \) with respect to \( i \), and

\[ z'_i = \frac{y_i}{x_i}. \]

A population of \((x_i, y_i)\) is simulated as follows.

For each given values of \( \rho \) and \( \alpha \) we assumed \( z_0 = 15 \) and \( \mu = 15 \). Also, \( e_i \) was assumed as a normal variate with mean 0 and variance 1. Then, we simulated \( z_1, z_2, \ldots, z_60 \), using the model given in (3.27), and we generated \( z'_1, z'_2, \ldots, z'_60 \) by using the model given in (3.28). Also we selected \( x_i \) as a discrete uniform variate from \((2, 3, \ldots, 9)\) and \( y_i = x_i z'_i \) for \( i = 1, 2, \ldots, 60 \). To reduce the effects of the initial value \( z_0 \), the population units are given by the set of 40 values of \((x_i, y_i)\) where \( i = 21, 22, \ldots, 60 \).

For \((\rho, \alpha) = (0.9, 3), (0.9, 6), (0.9, 1.2), (0.5, 3), (0.5, 5), (0.5, 1.2)\), we simulated 100 populations by using the model given in (3.28) as explained above. Also, from each simulated population, two samples one of size 4 and the other of size 8 are selected using each of the three sampling procedures presented above. For each of the simulated population, sample size, and sampling procedure, we calculated the sampling
variance of the population total estimator. The units were kept in their simulated order for the calculations of the variances for the fixed order stratified sampling method. Then, in Table 3.3 we presented the averages of these sampling variances of the estimators of the population totals calculated from the 100 simulated populations. Also in Table 3.3 we bracketed the efficiencies, where each efficiency is the ratio of the average variances of the fixed order stratified sampling scheme to the average variances of the selected pps sampling scheme.

Table 3.3 shows that the fixed order stratified sampling scheme always performs better than both the random order systematic sampling scheme and the pps with replacement sampling scheme.
Table 3.3: Empirical average of the sampling variances of the estimators of the population totals and the efficiencies (in brackets) as explained in Group II Simulations.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Sampling Procedure</th>
<th>fixed order stratified sampling</th>
<th>random order pps-sys sampling</th>
<th>pps with replacement sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho = .9, \alpha = .3$</td>
<td>4</td>
<td>85181 (1)</td>
<td>169749 (.50)</td>
<td>186567 (.46)</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>23539 (1)</td>
<td>75585 (.31)</td>
<td>95567 (.25)</td>
</tr>
<tr>
<td>$\rho = .9, \alpha = .6$</td>
<td>4</td>
<td>281817 (1)</td>
<td>593318 (.48)</td>
<td>652273 (.43)</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>70645 (1)</td>
<td>236903 (.30)</td>
<td>299734 (.24)</td>
</tr>
<tr>
<td>$\rho = .9, \alpha = 1.2$</td>
<td>4</td>
<td>1041744 (1)</td>
<td>2171979 (.48)</td>
<td>2385772 (.44)</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>280080 (1)</td>
<td>935172 (.30)</td>
<td>1184969 (.24)</td>
</tr>
</tbody>
</table>
Table 3.3 (continued): Empirical average of the sampling variances of the estimators of the population totals and the efficiencies (in brackets) as explained in Group II Simulations.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Sampling Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fixed order stratified sampling</td>
</tr>
<tr>
<td>$\rho = .5, \alpha = .3$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>74989 (1) (.51)</td>
</tr>
<tr>
<td>8</td>
<td>21790 (1) (.34)</td>
</tr>
<tr>
<td>$\rho = .5, \alpha = .6$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>261268 (1) (.49)</td>
</tr>
<tr>
<td>8</td>
<td>70580 (1) (.31)</td>
</tr>
<tr>
<td>$\rho = .5, \alpha = 1.2$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1025586 (1) (.49)</td>
</tr>
<tr>
<td>8</td>
<td>271462 (1) (.30)</td>
</tr>
</tbody>
</table>
3.8.3 Group III Simulations

Consider \( z_i \) as a linear function in \( i \) and this can be explained by the simple model,

\[
z_i = \mu + i\alpha,
\]

(3.29)

where \( \mu \) is the intercept, \( \alpha \) is the slope, and

\[
x_i = \frac{y_i}{x_i}.
\]

A population of \((x_i,y_i)\) is simulated as follows:

for a given value of \( \alpha \) we considered \( \mu = 15 \). Then, we simulated \( z_1, z_2, ..., z_{60} \) using the model given in (3.29). Also we selected \( x_i \) as a discrete uniform variate from \((2,3,...,9)\) and \( y_i = x_i z_i \) for \( i = 1, 2, ..., 60 \). The population units are given by the set of 40 values of \((x_i,y_i)\) where \( i = 21, 22, ..., 60 \).

For \( \alpha = .3, .6, 1.2, 12 \), we simulated 50 populations using the model given in (3.29) as explained above. Also, from each simulated population, two samples one of size 4 and the other of size 8 are selected using each of the three sampling procedures presented above. For each of the simulated population, sample size, and sampling procedure, we calculated the sampling variance of the population total estimator. The population units were arranged as follows:

(i) the first arrangement is such that the units were kept in their simulated order.

Using this arrangement, we calculated the variances (Table 3.4) of the estimators of the population totals for the fixed order stratified sampling method.
(ii) The second arrangement employed peak ordering (Table 3.5) as explained above. Then, in Table 3.4 and Table 3.5 we presented the averages of these sampling variances of the estimators of the population totals calculated from the 50 simulated populations. In addition, we presented the efficiencies, where a single efficiency is the ratio of the average variances of the fixed order stratified sampling scheme to the average variances of the selected pps sampling scheme.

Finally, Table 3.4 and Table 3.5 show that the fixed order stratified sampling scheme always performs better than both the random order systematic sampling scheme and the pps with replacement sampling scheme. When the peak ordering of the population units was employed, the fixed order stratified sampling scheme became vastly superior to the random order systematic sampling scheme.
Table 3.4: Empirical average of the sampling variances of the estimators of the population totals and the efficiencies (in brackets) as explained in Group III Simulations.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Sampling Procedure</th>
<th>fixed-order stratified sampling</th>
<th>random-order pps systematic sampling</th>
<th>pps with replacement sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>$\alpha = .3$</td>
<td></td>
<td>1351</td>
<td>2796</td>
<td>3064</td>
</tr>
<tr>
<td>$\alpha = .3$</td>
<td></td>
<td>(1)</td>
<td>(.48)</td>
<td>(.44)</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>8</td>
<td>464</td>
<td>1968</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>(1)</td>
<td>(.30)</td>
<td>(.24)</td>
</tr>
<tr>
<td>$\alpha = .6$</td>
<td></td>
<td>247595</td>
<td>516136</td>
<td>567446</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(.48)</td>
<td>(.44)</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>8</td>
<td>45285</td>
<td>202242</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>(1)</td>
<td>(.29)</td>
<td>(.22)</td>
</tr>
<tr>
<td>$\alpha = 1.2$</td>
<td></td>
<td>976113</td>
<td>2031848</td>
<td>2233367</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(.48)</td>
<td>(.44)</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>8</td>
<td>260166</td>
<td>1115547</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>(1)</td>
<td>(.30)</td>
<td>(.23)</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 3.5: Empirical average of the sampling variances of the estimators of the population totals and the efficiencies (in brackets) as explained in Group III Simulations. The units are arranged such that $z_i$ are in peak ordering.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>fixed-order stratified sampling</th>
<th>random-order pps systematic sampling</th>
<th>pps with replacement sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = .3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>29058</td>
<td>128000</td>
<td>140629</td>
</tr>
<tr>
<td></td>
<td>(1) (.23)</td>
<td>(.23)</td>
<td>(.21)</td>
</tr>
<tr>
<td>8</td>
<td>4246</td>
<td>54353</td>
<td>69130</td>
</tr>
<tr>
<td></td>
<td>(1) (.08)</td>
<td>(.08)</td>
<td>(.06)</td>
</tr>
<tr>
<td>$\alpha = .6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>118095</td>
<td>520355</td>
<td>571983</td>
</tr>
<tr>
<td></td>
<td>(1) (.23)</td>
<td>(.23)</td>
<td>(.21)</td>
</tr>
<tr>
<td>8</td>
<td>17179</td>
<td>219144</td>
<td>277813</td>
</tr>
<tr>
<td></td>
<td>(1) (.08)</td>
<td>(.08)</td>
<td>(.06)</td>
</tr>
<tr>
<td>$\alpha = 1.2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>475870</td>
<td>2096963</td>
<td>2303603</td>
</tr>
<tr>
<td></td>
<td>(1) (.23)</td>
<td>(.23)</td>
<td>(.21)</td>
</tr>
<tr>
<td>8</td>
<td>67234</td>
<td>858914</td>
<td>1087284</td>
</tr>
<tr>
<td></td>
<td>(1) (.08)</td>
<td>(.08)</td>
<td>(.06)</td>
</tr>
</tbody>
</table>
Chapter 4

VARIANCE ESTIMATION IN SYSTEMATIC SPATIAL SAMPLING

4.1 Introduction

The theory, presented by Cordy (1993), is not applicable to systematic sampling in a continuous universe as stated in his paper. In this chapter we introduce new concepts for which the spatial sampling theory becomes applicable to most of the systematic sampling problems. Suppose there is a domain, $D$, of interest, and that at each point $u \in D$ it is possible to measure a study variable $z(u)$. The domain $D$ can be a subset of $\mathbb{R}^n$ where $n \geq 1$. We are interested in estimating the population parameter

$$T = \int_D z(u) \, du$$

(4.1)

and deriving an estimator of the variance of the estimator of $T$ based on a probability sample from $D$. 

91
4.2 Systematic Sampling from a Finite Population

Before presenting some systematic sampling methods in spatial sampling, we review the systematic sampling methods from a finite universe. We limit our review to systematic sampling with one random start and to systematic sampling with multiple random starts. In addition, we discuss two approaches for analyzing the problem of systematic sampling. When the systematic sampling deals with multiple random starts, these two approaches lead to two different variance estimators.

4.2.1 Systematic Sampling with One Random Start from a Finite Population

Consider a finite population of size \( N \). Let the universe \( U \) be given by

\[
U = \{1, 2, \ldots, N\},
\]

and \( y_i \) be the response function at the \( i^{th} \) unit. Also, Suppose \( N = nk \). To select a sample of \( n \) units, draw with equal probability, \( 1/k \), a random integer, \( r \), from \( 1, 2, \ldots, k \). The selected sample consists of

\[
s_r = \{r + pk; \ p = 0, 1, \ldots, n - 1\}. \tag{4.2}
\]

For instance, if \( N = 100 \), \( n = 5 \), and \( k = 20 \) and if the first unit drawn is number 14, then the subsequent units are numbers 34, 54, 74 and 94. The selection
of the first unit determines the whole sample. This sampling method is called
simple systematic sampling or systematic sampling with one random start.

Traditional Sampling Approach.

This approach of analyzing the problem of estimation in systematic sampling
is based on treating each element \( i \in U \) as one sampling unit and proceeds as
follows:
because each element out of the \( N \) elements belongs to one and only one of the \( k \)
equally probable systematic samples, the inclusion probabilities are given by

\[
\pi_i = \frac{1}{k} \quad \text{for every } i \in U,
\]

and

\[
\pi_{ij} = \begin{cases} 
\frac{1}{k} & \text{if } i \text{ and } j \text{ belong to the same } s \\
0 & \text{otherwise.} 
\end{cases}
\]

The desirable property having all \( \pi_{ij} > 0 \) is not satisfied. As a result, an unbiased
variance estimator of the population total is not available.

The Horvitz-Thompson estimator of the population total

\[
T = \sum_{i=1}^{N} y_i,
\]

is given by

\[
\hat{T} = \sum_{p=0}^{n-1} \frac{y_{r+p}}{\pi_{r+p}},
\]
where $y$ is the study variable. Based on (4.6), the variance of $\hat{T}$ is given in the form

$$
Var(\hat{T}) = - \sum_{i=1}^{N} \sum_{j>i}^{N} (\pi_{ij} - \pi_i \pi_j) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2.
$$ (4.7)

The above formulas (4.3) and (4.4) are based on all units in $U$, and the associated formulas (4.6) and (4.7) use $y$ as the response function. This approach shall be referred to as the traditional sampling approach.

**Cluster Sampling Approach.**

Another way of looking at systematic sampling would be with $N = nk$ where the $k$ possible systematic samples are given by $s_1, s_2, \ldots, s_k$.

These samples represent a partition of $U$ into $k$ subpopulations. That is,

$$
U = \bigcup_{i=1}^{k} s_i.
$$

It is evident that the population has been divided into $k$ cluster units, each of which is just one systematic sample that contains $n$ of the original units. Let us, define a subset $U'$ of the set $U$ given by

$$
U' = \{1, 2, \ldots, k\}.
$$

$U'$ is the set of labels from which the first random start was selected and is considered as the universe of the cluster units. The population total given in (4.5) can be written as
The systematic sampling can thus be described as a uniform random selection of one cluster label, say $i$, from $U'$. The selected systematic sample is just the cluster $s_i$ and the units in $s_i$ are completely surveyed. The response function associated with $s_i$ is $t_s$ as given in (4.8). The first-order inclusion probability of the $i^{th}$ cluster is given by

$$
\pi_i^* = \frac{1}{k} \text{ for every } i \in U',
$$

and the second-order inclusion probability of $i^{th}$ and $j^{th}$ cluster units is given by

$$
\pi_{ij}^* = 0 \text{ for every } i \neq j \in U'.
$$

The Horvitz-Thompson estimator of (4.8) can be written as

$$
\hat{T} = \frac{t_s}{\pi_i^*},
$$

and the variance of $\hat{T}$ can be written in the form

$$
\text{Var}(\hat{T}) = -\sum_{i=1}^{k} \sum_{j \neq i} (\pi_{ij}^* - \pi_i^* \pi_j^*) \left( \frac{t_{si}}{\pi_i^*} - \frac{t_{sj}}{\pi_j^*} \right)^2.
$$

The above arguments are based on treating each element $i \in U'$ as one sampling unit and the response function associated with this unit is given by $t_s$. This approach shall be referred to as the cluster sampling approach.
Note that the resulted estimators from both approaches (4.6) and (4.10) for population total are the same. However, the difference between these two approaches will appear in the variance estimators. This cannot be shown in the systematic sampling with one random start because the variance is not unbiasedly estimable. The difference can be shown in the systematic sampling with multiple random starts presented below.

4.2.2 Systematic Sampling with Multiple Random Starts from a Finite Population

Let \( n = n_1 n_2 \) and \( N = N_1 n_1 \). Define

\[
U' = \{1, 2, \ldots, N_1\}.
\]

Select \( n_2 \) distinct random starts from \( U' \). Let \( s' \) be the set of these random starts which we denote by

\[
s' = \{r_1, r_2, \ldots, r_{n_2}\}.
\]

The associated sample \( s^* \) from \( U \) is given by

\[
s^* = \{r_i + j N_1 : i = 1, \ldots, n_2; j = 0, 1, \ldots, n_1 - 1\}.
\]

Traditional Sampling Approach.

The first-order inclusion probability is given by

\[
\pi_i = \frac{n_2}{N_1} \text{ for every } i \in U,
\]

and for every \( i \neq j \in U \), the second-order inclusion probability is given by
\[ \pi_{ij} = \begin{cases} \frac{n_2}{N_1} & \text{if } i \text{ & } j \text{ are covered by the same starting unit} \\ \frac{n_2(n_2 - 1)}{N_1(N_1 - 1)} & \text{otherwise.} \end{cases} \] (4.12)

The Horvitz-Thompson estimator of (4.5) is given by

\[ \hat{T} = \sum_{i \in s} \frac{y_i}{\pi_i}, \]

and the Yates-Grundy variance estimator of the variance of \( \hat{T} \) is given by

\[ v_1 = -\sum_{i \in s} \sum_{j > i} \left( \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \right) \left( \frac{y_i - y_j}{\pi_i} \right)^2. \] (4.13)

Cluster Sampling Approach.

The selected sample \( s^* \) can be written as a union of disjoint subsamples, that is

\[ s^* = \bigcup_{r_i \in s'} s_{r_i}, \]

where \( s_{r_i} \) is a systematic sample covered by the start \( r_i \) and is given by

\[ s_{r_i} = \{ r_i + jN_i : j = 0, 1, \ldots, n_1 - 1 \}, \quad i = 1, \ldots, n_2. \]

The first-order inclusion probability is given by

\[ \pi_{i}^* = \frac{n_2}{N_1} \text{ for every } i \in U', \]

and the second-order inclusion probability is given by

\[ \pi_{ij}^* = \frac{n_2(n_2 - 1)}{N_1(N_1 - 1)}, \quad i \neq j \in U'. \] (4.14)
The Horvitz-Thompson estimator of (4.8) is given by

$$\hat{T} = \sum_{i \in s'} \frac{t_{x_i}}{\pi_i^*} \text{ with } t_{x_i} = \sum_{j \in s_i} y_j. \quad (4.15)$$

Based on (4.15), we obtain a variance estimator of the variance of $\hat{T}$ as

$$v_2 = -\sum_{i \in s'} \sum_{j \in s_i} \frac{(\pi_{ij}^* - \pi_i^* \pi_j^*)}{\pi_i^*} \left( \frac{t_{x_i}}{\pi_i} - \frac{t_{x_j}}{\pi_j} \right)^2. \quad (4.16)$$

**Remark 4.1** It can be seen that in the traditional sampling approach as considered above, $\pi_{ij} < \pi_i \pi_j$ whenever $i$ and $j$ are generated by two different random starts. However, $\pi_{ij} > \pi_i \pi_j$ if $i$ and $j$ are generated by the same random start. As a result, the variance estimator $v_1$ given in (4.13) takes negative values with positive probability. While in the cluster sampling approach, $\pi_{ij}^* < \pi_i^* \pi_j^* \text{ for } i \neq j \in U'$. Therefore, the variance estimator $v_2$ given in (4.16) takes only nonnegative values. This attractive feature of $v_2$ inspires investigating cluster sampling approach whenever systematic sampling is used.

Cordy (1993) stated in his paper that his theory is not meant to cover systematic sampling. In this chapter we demonstrate how to manipulate the sampled points in such a manner that this theory can be used for most of these systematic sampling methods.

In the next section, we demonstrate how the known results in systematic sampling from a finite universe can be extended to systematic sampling from a continuous universe. Then, we introduce new systematic sampling methods which are useful in spatial sampling.
4.3 Systematic Sampling from a Continuous Universe

In this section we use all the notations that were defined in chapter 2. In spatial systematic sampling we consider the universe \( U \) as a union of \( m \) disjoint subsets of the same sizes and the domain of interest \( D \) included in \( U \). Without loss of generality, suppose each of these subsets is a line segment if \( D \) has one dimension and a rectangle if \( D \) has two dimensions, and so on. From any of these subsets, say \( U' \), we select a fixed-size sample of points, say \( s' \), according to some non-systematic sampling design. For the given sampling design, suppose the first- and the second-order inclusion density functions can be readily calculated and denoted by \( \pi(x) \) for \( x \in U' \) and \( \pi(x, x') \) for \( x \neq x' \in U' \). The selected sample \( s \) from \( U \) is obtained by extending \( s' \) systematically to all of the other \( m \) subsets which construct \( U \). Here \( s' \) is the set of the starting points of the sample \( s \). For each \( x'_i \in U' \), let \( x'_1, x'_2, \ldots, x'_m \) be the points in the other subsets which are covered systematically by the point \( x'_i \) and define \( t(x'_i) = \sum_{i=1}^{m} z(x'_i) \). This function \( t \) shall be referred to as the cluster response function. Note that \( x'_j \) is a deterministic function of \( x'_i \), for \( j = 1, 2, \ldots, m \).

**Theorem 4.2** The population total given in (4.1) can be written as

\[
T = \int_{U'} t(x'_i) \, dx'_i = \int_{U'} \sum_{i=1}^{m} z(x'_i) \, dx'_i. \tag{4.17}
\]

This theorem looks intuitively correct. To attempt to prove it in general is lengthy. Also, the notations get complicated. For this reason we avoid proving it.
However, we prove one particular case in the next subsection and the proof can be simply extended to all of the spatial systematic methods which we consider.

The above theorem promotes thinking of systematic sampling procedure same as non-systematic sampling from the subset $U'$ with response function $t$. The population parameters and their estimators using this technique are presented as follows:

For a given fixed-size sample $s'$ from $U'$, the estimator of (4.17) is given by

$$\hat{T} = \sum_{x_i \in s'} \frac{t(x_i)}{\pi(x_i)},$$

(4.18)

and the variance of $\hat{T}$ is given by

$$Var(\hat{T}) = \frac{1}{2} \int_{U'} \int_{U'} (\pi(x)\pi(x') - \pi(x,x')) \left( \frac{t(x)}{\pi(x)} - \frac{t(x')}{\pi(x')} \right)^2 dx dx'. \quad (4.19)$$

If the pairwise inclusion density $\pi(x,x') > 0$ almost everywhere in $U'$, then the variance estimator of (4.19) is given by

$$v = -\frac{1}{2} \sum_{x_i \in s'} \sum_{x_j \neq x_i} \frac{(\pi(x_i,x_j) - \pi(x_i)\pi(x_j))}{\pi(x_i,x_j)} \left( \frac{t(x_i)}{\pi(x_i)} - \frac{t(x_j)}{\pi(x_j)} \right)^2. \quad (4.20)$$

In what follows we illustrate how to use the above general argument for obtaining variance estimator for many systematic sampling methods.
4.3.1 Systematic Sampling with one Random Start from a Continuous Universe

Case of \( U \) in One Dimension.

Let \( U' \) be a subset of \( U \) defined as

\[
U' = \{ x' = (x'_1) : 0 < x'_1 < l_1 \},
\]

and

\[
U' + (i_1 l_1) = \{ x' = (x'_1 + i_1 l_1) : 0 < x'_1 < l_1 \}, i_1 = 0, 1, \ldots, m_1 - 1.
\]

Actually, we define \( U' = U' + (0) \). The subsets \( U' + (i_1 l_1) \) represent a partition of \( U \) into subpopulations, that is

\[
U = \bigcup_{i_1=0}^{m_1-1} [U' + (i_1 l_1)] \text{ almost everywhere (a.e.).} \tag{4.21}
\]

For any \( (x) \in U' \), define the cluster response function by

\[
t(x) = \sum_{i_1=0}^{m_1-1} z(x + i_1 l_1).
\]

The population total is given by

\[
T = \int_U z(x) \, dx. \tag{4.22}
\]

**Corollary 4.3** The population total can be written as

\[
T = \int_{U'} \sum_{i_1=0}^{m_1-1} z(x + i_1 l_1) \, dx = \int_{U'} t(x) \, dx.
\]
Proof: Recall that the population total is given in (4.22). Using the partition of 
$U$ given in (4.21) we obtain

$$T = \sum_{i_1=0}^{m_1-1} \int_{[U'+(i_1 l_1)]} z(x) \, dx. \quad (4.23)$$

Now,

$$\int_{[U'+(i_1 l_1)]} z(x) \, dx = \int_{i_1 l_1}^{(i_1+1) l_1} z(x) \, dx. \quad (4.24)$$

By using the change of variable $u = x - i_1 l_1$ in the right hand side of (4.24), we 
obtain

$$\int_{[U'+(i_1 l_1)]} z(x) \, dx = \int_0^{i_1 l_1} z(u + i_1 l_1) \, du = \int_{U'} z(x + i_1 l_1) \, dx. \quad (4.25)$$

Substituting the results of (4.25) in (4.23), we obtain

$$T = \int_{U'} \sum_{i_1=0}^{m_1-1} z(x + i_1 l_1) \, dx = \int_{U'} t(x) \, dx, \quad (4.26)$$

and the corollary is proved. \(\Box\)

A spatial systematic sample with one random start can be selected as follows.

Select a uniform random number, $x_1$, from the interval $(0, l_1)$. Let $X_1$ be a random
variable associated with the realized value $x_1$. The resulted sample from $U'$ contains the only point $x_1$, that is

$$s' = \{(x_1)\}.$$  

The selected sample $s$ from $U$, associated with $s'$ from $U'$, is given by the set of points

$$s = \{u = (x_1 + i_1 l_1) : (x_1) \in s'; \; i_1 = 0, 1, 2, \ldots, m_1 - 1\}.$$  

**Case of $U$ in Two Dimensions.**

The subset $U'$ of $U$ is defined as follows:

$$U' = \{x' = (x'_1, x'_2) : 0 < x'_1 < l_1; \; 0 < x'_2 < l_2\}.$$  

For each $x' = (x'_1, x'_2) \in U'$, define the cluster response function by

$$t(x') = \sum_{i_1=0}^{m_1-1} \sum_{i_2=0}^{m_2-1} z(x'_1 + i_1 l_1, x'_2 + i_2 l_2).$$  

Select two uniform random numbers $x_1$ and $x_2$ from the interval $(0, l_1)$ and $(0, l_2)$ respectively. Let $X=(X_1, X_2)$ be a random vector associated with $x=(x_1, x_2)$. The selected sample from $U'$ contains the only point $(x_1, x_2)$ that is

$$s' = \{(x_1, x_2)\}.$$  

A sample $s$ from $U$, associated with $s'$ from $U'$, is given by the set of points

$$s = \{u = (x_1 + i_1 l_1, x_2 + i_2 l_2) : (x_1, x_2) \in s'; \; i_j = 0, 1, \ldots, m_j - 1; \; j = 1, 2\}.$$
Case of $U$ in $n$ Dimensions.

Define

$$U' = \{x' = (x'_1, x'_2, \ldots, x'_n) : 0 < x'_i < l_i; \ i = 1, \ldots, n\}.$$  

Select, with respect to the $e_j$ axis, a uniform random number, $x_j$, from the interval $(0, l_j)$, for $j = 1, \ldots, n$. Let $X = (X_1, X_2, \ldots, X_n)$ be the random vector associated with $x = (x_1, x_2, \ldots, x_n)$. The selected sample $s'$ from $U'$ contains only the selected point

$$s' = \{x = (x_1, x_2, \ldots, x_n)\}.$$  

The sample $s$ from $U$, associated with $s'$ from $U'$, is given by the set of points

$$s = \{u = (x_1 + i_1 l_1, \ldots, x_n + i_n l_n) : (x_1, x_2, \ldots, x_n) \in s'; \ i_j = 0, 1, \ldots, m_j - 1; \ j = 1, 2, \ldots, n\}.$$  

Now, $t(x)$ will be observed if $X$ assumes $x$. The first-order inclusion density is given by

$$\pi(x) = f_X(x) = \frac{1}{\mu} \text{ for every } x \in U',$$

and the second-order inclusion density is given by

$$\pi(x, x') = 0 \text{ for every } x \neq x' \in U',$$

where $\mu = l_1 l_2 \ldots, l_n$.

By considering the cluster sampling approach as in systematic sampling from a finite population, the theory presented by Cordy (1993) can be used. Further, the
Horvitz-Thompson estimator of (4.26) is given in (4.18). An unbiased estimator of the sampling variance can not be obtained because the pairwise inclusion density is zero for any two different points in $U'$.

### 4.3.2 Systematic Spatial Sampling with Multiple Random Starts from a Spatial Universe

To allow for variance estimation, systematic sampling method with multiple random starts can be considered. Given $U$ in $n$ dimensions, define

$$U' = \{x' = (x'_1, x'_2, \ldots, x'_n) : 0 < x'_i < l_i; i = 1, \ldots, n\}.$$ 

For each point $x' = (x'_1, x'_2, \ldots, x'_n) \in U'$, define the cluster response function as

$$t(x') = \sum_{j_1=0}^{m_1-1} \cdots \sum_{j_n=0}^{m_n-1} z(x'_1 + j_1 l_1, \ldots, x'_n + j_n l_n).$$

Select $k$ random points from $U'$, denoted by $x_i$, for $i = 1, 2, \ldots, k$. Each of these random points can be written as $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$, where $x_{ij}$ is a uniform random variate from the interval $(0, l_j)$, for $j = 1, \ldots, n$. The selected sample $s'$ from $U'$ contains the $k$ initial random starts and is given by

$$s' = \{x_i = (x_{i1}, x_{i2}, \ldots, x_{in}) : i = 1, 2, \ldots, k\}.$$

Let $X_1, X_2, \ldots, X_k$ be the associated random vectors with the realized starting points $x_1, x_2, \ldots, x_k$ respectively. The associated sample $s$ from $U$ with $s'$ from $U'$ is given by the set of points
\[ s = \{ u = (x_{i1} + j_1 l_1, \ldots, x_{in} + j_n l_n) : (x_{i1}, x_{i2}, \ldots, x_{in}) \in s' ; i = 1, \ldots, k ; j_r = 0, 1, \ldots, m_r - 1 ; r = 1, \ldots, n \}. \]  

(4.28)

Note that, for any \( x \in U' \), \( t(x) \) is observed if for some \( j = 1, \ldots, k \), the random vector \( X_j \) assumes \( x \). Now the inclusion density functions are given by

\[ \pi(x) = \frac{k}{\mu}, \text{ for every } x \in U', \]

and

\[ \pi(x, x') = \frac{k(k-1)}{\mu^2}, \text{ for every } x \neq x' \in U'. \]

The Horvitz-Thompson estimator of \( T \) in (4.17) can be given by (4.18) and the variance estimator is given by (4.20). It is known that the systematic sampling with multiple random starts is less efficient than the one with one random start, given that the total sample size is the same for both sampling methods. This could be due to the sampling points in systematic sampling with multiple random starts having less spatial coverage than the one with one random start. In the next subsection we provide a more efficient systematic sampling method which provides more spatial coverage of the universe than the systematic sampling with multiple random starts method. This new systematic sampling method allows for unbiased variance estimation.
4.3.3 Systematic Sampling with Multiple Restricted Random Starts from a Spatial Universe

Case of $U$ in One Dimension.

Let $M_1$ be the length of $U$. Superimpose a one dimensional grid on $U$ with spacing $l_1 = \frac{M_1}{m_1}$, where $m_1$ is an integer and represents the number of strata in the universe $U$ and $l_1$ is the length of each stratum. Also assume $m_1 = k_1 r_1$, where $k_1$ and $r_1$ are positive integers. Consider $U'$ as a subset of $U$ and defined as follows:

$$U' = \{x' = (x'_1) : 0 < x'_1 < k_1 l_1\}.$$  

Note that $U'$ contains $k_1$ grid cells. Now select from $U'$ a sample of size $k_1$ using the sampling procedure denoted by $(tss)$ in chapter 2. Suppose the resulted sample from $U'$ is given by the set of points

$$s' = \{x = (x_{1i}) : i_1 = 1, 2, \ldots, k_1\}.$$  

A sample $s$ from $U$, associated with the sample $s'$, can be selected and is given by

$$s = \{u = (x_{1i} + \alpha_1 k_1 l_1) : (x_{1i}) \in s'; \alpha_1 = 0, 1, \ldots, r_1 - 1\}.$$  

This is a systematic sample with $k_1$ starts where these starts are selected with restricted randomization. This sampling method with the restricted randomization provides a more even spatial coverage of $U$ than the systematic sampling method with multiple random starts. Also a nonnegative unbiased variance estimator shall be derived.
Let \( X_1, X_2, \ldots, X_{1k} \) be the associated random variables with the realized starting points \( x_{11}, x_{12}, \ldots, x_{1k} \), respectively. For any \((x'_1) \in U'\), define the cluster response function by

\[
t(x'_1) = \sum_{\alpha_1=0}^{r_1-1} \pi(x'_1 + \alpha_1 k_1 l_1).
\]

Note that, for any \( x \in U' \), \( t(x) \) is observed if for some \( j = 1, \ldots, k_1 \), the random variable \( X_{1j} \) assumes \( x \). Here \( x \) is a vector of one component. Now the inclusion density functions can be given by

\[
\pi(x) = \frac{1}{\mu}, \quad \text{for every } x \in U',
\]

and

\[
\pi(x, x') = \frac{\Pr(B_{xx'})}{\mu^2}, \quad \text{for every } x \neq x' \in U',
\]

where \( \mu = l_1 \) and \( B_{xx'} \) is the event that \( x \) and \( x' \) are in two different cells of the translated portion of the grid over \( U' \).

The probability of \( B_{xx'} \), where \( x = (x_1) \in U' \) and \( x' = (x'_1) \in U' \), is given by

\[
\Pr(B_{xx'}) = \begin{cases} 
1 & \text{if } l_1 \leq \delta(x_1, x'_1) \\
\frac{\delta(x_1, x'_1)}{l_1} & \text{if } l_1 > \delta(x_1, x'_1),
\end{cases}
\]

(4.29)

where \( \delta(x_1, x'_1) = \min\{|x_1 - x'_1|, k_1 l_1 - |x_1 - x'_1|\} \). Note that we viewed the sampling in a circular manner from \( U' \). Using the above inclusion density functions in (4.18) and (4.20), we obtain the estimator of the total and the estimator of the variance.
Case of $U$ in Two Dimensions.

Let

$$U' = \{ x' = (x_1', x_2') : 0 < x_1' < k_1 l_1; 0 < x_2' < l_2 \}. $$

Note that $U'$ contains the first $k_1$ cells on the left of the first horizontal row of grid cells (Figure 4.1). Select, with respect to the $e_1$ axis, a uniform random number, $w_1$, from the interval $(0, l_1)$. Let $W_1$ be the random variable for which $w_1$ is its realized value. Translate the portion of the grid in $U'$ in such a way that the grid point originally at $(0,0)$ is transformed to $(w_1,0)$. Now, select a uniform random number, $x_{1i}$, from the interval $(w_1 + (i_1 - 1)l_1, w_1 + i_1 l_1)$ for $i_1 = 1, \ldots, k_1$. For each selected $x_{1i}$ select a uniform random number $x_{2i}$ from $(0, l_2)$. If $x = (x_1, x_2)$ is a selected point, then $x$ is included in the sample with its $x_1$ coordinate being suitably reduced by $\text{mod}(k_1 l_1)$. The selected sample from $U'$ is given by the set of points

$$s' = \{ x = (x_{1i}, x_{2i}) : i_1 = 1, 2, \ldots, k_1 \}. $$

A sample $s$, associated with the sample $s'$, can be selected from $U$ and is given by the set of points

$$s = \{ u = (x_{1i} + \alpha_1 k_1 l_1, x_{2i} + \alpha_2 l_2) : (x_{1i}, x_{2i}) \in s'; \quad \alpha_1 = 0, 1, \ldots, r_1 - 1; \quad \alpha_2 = 0, 1, \ldots, m_2 - 1 \}. \quad (4.30)$$

For each point $x = (x_1, x_2) \in U'$, define the cluster response function as

$$t(x) = \sum_{\alpha_1=0}^{r_1-1} \sum_{\alpha_2=0}^{m_2-1} z(x_1 + \alpha_1 k_1 l_1, x_2 + \alpha_2 l_2).$$

109
Let $X_{i_1}$ be the random vector associated with the realized point $(x_{1i_1}, x_{2i_1})$ in $U'$, for $i_1 = 1, ..., k_1$. Now, let $x = (x_1, x_2)$ and $x' = (x'_1, x'_2)$ be two points in $U'$. Note that $t(x)$ is observed if $X_{i_1}$ assumes $x$ for some $i_1 = 1, ..., k_1$. Using a similar argument to that presented in chapter 2, it can be proven that the inclusion density functions are given by

$$\pi(x) = \frac{1}{\mu} \text{ for any } x \text{ in } U',$$

and

$$\pi(x, x') = \frac{Pr(B_{XX'})}{\mu^2} \text{ for any } x \neq x' \in U',$$

where $\mu = l_1 l_2$ and $B_{XX'}$ is the event that $x$ and $x'$ are in two different cells of the translated portion of the grid over $U'$. $Pr(B_{XX'})$ is given by

110
\[ Pr(B_{XX'}) = \begin{cases} 
1 & \text{if } l_1 \leq \delta(x_1, x'_1) \\
\frac{\delta(x_1, x'_1)}{l_1} & \text{if } l_1 > \delta(x_1, x'_1). 
\end{cases} \] (4.31)

where \( \delta(x_1, x'_1) = \min\{|x_1 - x'_1|, k_l l_1 - |x_1 - x'_1|\} \). Note that we viewed the sampling in a circular manner from \( U' \) with respect to \( e_1 \) axis only.

Case of \( U \) in \( n \) Dimensions.

Let \( M_i \) be the length of \( U \) with respect to the \( e_i \) axis for \( i = 1, \ldots, n \). Superimpose an \( n \)-dimensional grid on \( U \) with spacing \( l_i = \frac{M_i}{m_i} \) where \( m_i \) is an integer and \( l_i \) is the length of each stratum with respect to the \( e_i \) axis. Also let \( n' \) be an integer satisfying \( 1 \leq n' \leq n \) and assume that \( m_i = k_i r_i \) where \( k_i \) and \( r_i \) are positive integers, for \( i = 1, \ldots, n' \). Consider \( U' \) as a subset of \( U \) and defined as

\[ U' = \{x' = (x'_1, x'_2, \ldots, x'_n) : 0 < x'_i < k_i l_i, \ i = 1, \ldots, n'; \\
0 < x'_j < l_j, \ j = n' + 1, \ldots, n\}, \] (4.32)

where \( 1 \leq n' \leq n \). The portion of the grid over \( U' \) shall be shifted with respect to the \( e_1, \ldots, e_n \) axis only. In other words, select with respect to the \( e_i \) axis a uniform random number \( w_i \) from the interval \((0, l_i)\) for \( i = 1, \ldots, n' \). Translate the portion of the grid in \( U' \) in such a manner that the grid point initially at the origin, where the origin is a vector of \( n \) zeros \((0,0, \ldots, 0)\), moves to \((w_1, w_2, \ldots, w_{n'}, 0, \ldots, 0)\). Then, the selected sample from \( U' \) is given by the set of points

\[ s' = \{x = (x_{1i_1}, x_{2i_2}, \ldots, x_{ni_{n'}}, x_{n'+1}, \ldots, x_n) : \ i_j = 1, 2, \ldots, k_j; \ j = 1, \ldots, n'\}. \]
Figure 4.2: Example of $U$ and $U'$ in a two-dimensional space where $U$ is the large rectangle and $U'$ is the rectangle in the lower left of $U$ with three of its vertices are at $2L_2$, $O$ and $3L_1$ ($n = n' = 2$).

where each of $x_{ji}$ is a uniform random variate from the interval $(w_j + (i_j - 1)l_j, w_j + i_jl_j)$ reduced by mod$(kjl_j)$ for $j = 1, \ldots, n'$, and where $x_j$ is a uniform random variate from the interval $(0, l_j)$ for $j = n' + 1, \ldots, n$.

A sample $s$, associated with the sample $s'$, can be selected from $U$ and is given by the set of points

$$s = \{(x_{i_1}, x_{n'i_n'}, x_{n'i_{n'}}, x_{n'i_{n'+1}}, \ldots, x_n + \alpha_n l_n):$$

$$(x_{i_1}, x_{2i_2}, \ldots, x_{n'i_{n'}}, x_{n'i_{n'+1}}, \ldots, x_n) \in s'; \quad \alpha_i = 0, 1, \ldots, r_i - 1, \quad i = 1, \ldots, n';$$

$$\alpha_j = 0, 1, \ldots, m_j - 1, \quad j = n' + 1, \ldots, n\}.$$  

(4.33)
For each $x = (x_1, x_2, \ldots, x_{n'}, x_{n'+1}, \ldots, x_n) \in U'$, define the cluster response function by

$$t(x) = \sum_{\alpha_1 = 0}^{m_1-1} \cdots \sum_{\alpha_{n'} = 0}^{m_{n'}-1} \sum_{\alpha_{n'} = 0}^{m_{n'}-1} \cdots \sum_{\alpha_n = 0}^{m_n-1} \{x(x_1 + \alpha_1 l_1, \ldots, x_{n'} + \alpha_{n'} l_{n'}, x_{n'+1} + \alpha_{n'+1} l_{n'+1}, \ldots, x_n + \alpha_n l_n)\}. \quad (4.34)$$

Similarly, it can be proven that the inclusion density functions are given by

$$\pi(x) = \frac{1}{\mu} \text{ for any } x \text{ in } U',$$

and

$$\pi(x, x') = \frac{Pr(B_{xx'})}{\mu^2} \text{ for any } x \neq x' \in U',$$

where $\mu = l_1 l_2 \cdots l_n$ and $B_{xx'}$ is the event that $x$ and $x'$ are in two different cells of the translated portion of the grid over $U'$. In this design $Pr(B_{xx'})$ is given by

$$Pr(B_{xx'}) = \begin{cases} 1 & \text{if } l_i \leq \delta(x_i, x'_i) \text{ for some } i \\ S_1 - S_2 + S_3 - S_4 + \cdots + (-1)^{n'+1} S_{n'} & \text{otherwise}, \end{cases} \quad (4.35)$$

where $\delta(x_i, x'_i) = \min\{|x_i - x'_i|, k l_i - |x_i - x'_i|\}$ and $S_i$ is given in (2.27), for $i = 1, 2, \ldots, n'$. Note that we viewed the sampling in a circular manner from $U'$ with respect to the $e_1, e_2, \ldots, e_n'$ axes. All of the variance estimators for the spatial sampling designs which we introduced in this chapter take only nonnegative values.
Chapter 5

VARIANCE ESTIMATION IN A MEAN BALANCED SPATIAL SAMPLING DESIGN

5.1 Introduction

Consider the continuum domain $D$, the universe $U$ and the grid superimposed on $U$ as defined in section 2.2. A mean balanced sample (mbs) can be obtained by performing the following three steps independently:

i) Randomly translate the grid over $U$ while keeping $U$ in the same position. In other words, select a uniform random number, $w_j$, from the interval $(0, l_j)$ with respect to the $e_j$ axis, for $j = 1, 2, ..., n$. Let $W_j$ be the random variable for which $w_j$ is its realized value. Translate the grid in such a way that the grid point originally at $(0, 0, ..., 0)$ is moved to $w = (w_1, w_2, ..., w_n)$.

ii) From each cell of the translated grid, select the first point at random and the second point symmetrical to the first with respect to the center of that cell (Figure 5.1).
iii) If \( x = (x_1, x_2, ..., x_n) \) is a selected point in step (ii), then \( x \) is included in the sample with its \( x_i \) coordinate being suitably reduced by \( \text{mod}(M_i) \) for \( i = 1, 2, ..., n \).

Note that \( M_i \) is defined to be the length of \( U \) with respect to the \( e_i \) axis.

Note that we viewed the sampling in a circular manner with respect to each axis.

Let \( C \) be the set of the random variables associated with the realized sample points.

Also let \( f_X(x) \) be the density function of \( X \) at \( x \).

This sampling design provides a good spatial coverage of the population.

Following the theory presented by Cordy (1993), we investigate this sampling design by calculating the inclusion density functions for which \( U \) is in one dimension and we also calculate these inclusion density functions in general, that is, \( U \) is in \( n \) dimensions.
5.2 The Mean Balanced Design from a One-dimensional Space

Suppose the domain of the response variable is of one dimension, that is, \( n = 1 \). Then, the domain of the study variable is a bounded subset of the set of real numbers. Place the left edge of the universe \( U \) at the origin of the real line. Also use the grid over \( U \) with \( m_1 \) cells each of length \( l_1 \) and represent each point by its single coordinate.

Let \( W \) be a random variable such that \( w \) is its realized value. When \( n = 1 \), the sample space of \( W \) is given by

\[
\Omega = \{ w : 0 < w < l_1 \}.
\]

For each \( x \neq x' \in U \), let \( B_{xx'} \) be the event that \( x \) and \( x' \) are in two different cells of the translated grid, and \( A_{xx'} \) be the event that some cell of the randomly translated grid contains both \( x \) and \( x' \). Also let \( \delta(x, x') = \min\{|x - x'|, M_1 - |x - x'|\} \).

For any \( x \in U \) there exist an integer \( k_x \) and a \( w_x \in \Omega \) such that \( w_x + k_x l_1 = x \).

Now, the first order inclusion density, for any \( x \) in \( U \), is given by

\[
\pi(x) = \sum_{x \in C} f_X(x)
= \sum_{x \in C} \lim_{h \to 0} \frac{Pr[x - h < X < x + h]}{2h}
= \lim_{h \to 0} \sum_{x \in C} \frac{Pr[x - h < X < x + h, W \in \Omega_x]}{2h}
= \lim_{h \to 0} \sum_{x \in C} \frac{Pr[x - h < X < x + h | W \in \Omega_x]Pr[\Omega_x]}{2h},
\]

where \( \Omega_x = \Omega - \{w_x\} \) and \( Pr[\Omega_x] = 1 \). Also given \( (W \in \Omega - \{w_x\}) \), \( x \) falls between two grid boundaries. Therefore, there are two possible random variables.
(r.v.) which can assume the value of $x$. Suppose $X$ and $X'$ are these r.v. which can assume $x$ and have the grid cell containing $x$ as their conditional domain. Then we can write

$$
\pi(x) = \lim_{h \to 0} \frac{Pr[x - h < X < x + h \mid W \in \Omega - \{w_x\}]}{2h} + \frac{Pr[x - h < X' < x + h \mid W \in \Omega - \{w_x\}]}{2h},
$$

(5.2)

and that leads to

$$
\pi(x) = f(x \mid w)(x) + f(x' \mid w)(x),
$$

(5.3)

where $w$ is a realized value of $W$ in $\Omega - \{w_x\}$. Since $X$ and $X'$ are conditionally distributed with uniform density $1/l_1$, we obtain

$$
\pi(x) = \frac{2}{l_1}, \quad x \in U.
$$

(5.4)

The pairwise inclusion density $\pi(x, x')$, $x \neq x' \in U$, can be written as the limit as $h \to 0$ and $h' \to 0$ of

$$
\sum_{x, x' \in \Omega} \frac{Pr(x - h < X < x + h, x' - h' < Y < x' + h')}{4hh'}. \quad (5.5)
$$

To calculate $\pi(x, x')$, two possible cases shall be considered:

Case 1: $\delta(x, x') \geq l_1$.

Expression (5.5) can be written as

117
\[
\sum_{x,y \in \mathcal{C}} \frac{Pr(x-h < X < x+h, x' - h' < Y < x' + h', W \in \Omega)}{4hh'}.
\]

(5.6)

In this case \(B_{xx'} = \Omega\) almost everywhere (a.e.), and hence (5.6) is the same as

\[
\sum_{x,y \in \mathcal{C}} Pr(x-h < X < x+h, x' - h' < Y < x' + h' \mid W \in B_{xx'})
\]

(5.7)

Given the event \(W \in B_{xx'}\), \(x\) and \(x'\) are in two different grid cells. Therefore, there are two pairs of r.v. in \(\mathcal{C}\) which can assume the values \(x\) and \(x'\) respectively. Suppose that the first pair \(X\) and \(X'\), have the grid cell containing \(x\) as their conditional domain and the other pair, say \(Y\) and \(Y'\), have the grid cell containing \(x'\) as their conditional domain. Then the numerator of (5.7) can be written as follows:

\[
Pr[X \in (x-h, x+h), Y \in (x'-h', x'+h') \mid W \in B_{xx'}] +
Pr[X \in (x-h, x+h), Y' \in (x'-h', x'+h') \mid W \in B_{xx'}] +
Pr[X' \in (x-h, x+h), Y \in (x'-h', x'+h') \mid W \in B_{xx'}] +
Pr[X' \in (x-h, x+h), Y' \in (x'-h', x'+h') \mid W \in B_{xx'}].
\]

(5.8)

Each of the above four terms, when divided by \(4hh'\) and its limit is taken, as \(h \to 0\) and \(h' \to 0\), equals to \(\frac{1}{l_1^2}\). Therefore we obtain

\[
\pi(x, x') = \frac{4}{l_1^2}, \quad \delta(x, x') > l_1.
\]

(5.9)

Case 2: \(\delta(x, x') < l_1\).

The numerator of (5.5) can be written as
The first term on the right hand side of (5.10) can be written as

\[
\sum_{X,Y \in C} \Pr[X \in (x - h, x + h), Y \in (x' - h', x' + h'), W \in \Omega] =
\]

\[
= \sum_{X,Y \in C} \Pr[X \in (x - h, x + h), Y \in (x' - h', x' + h'), W \in B_{x' \theta}'] +
\]

\[
= \sum_{X,Y \in C} \Pr[X \in (x - h, x + h), Y \in (x' - h', x' + h'), W \in A_{x' \theta}']. \quad (5.10)
\]

Given the event \((W \in B_{x' \theta})\), one of the grid boundaries falls between \(x\) and \(x'\).

Therefore, there are two possible pairs of r.v. in \(C\) which can assume the values of \(x\) and \(x'\) respectively. Suppose that the first pair \(X\) and \(X'\), have the grid cell containing \(x\) as their conditional domain and the second pair, say \(Y\) and \(Y'\), have the grid cell containing \(x'\) as their conditional domain. Then using (5.11), the first term on the right hand side of (5.10) can be expanded as

\[
\sum_{X,Y \in C} \Pr[X \in (x - h, x + h), Y \in (x' - h', x' + h'), W \in B_{x' \theta}] P[B_{x' \theta}]. \quad (5.11)
\]

\[
\sum_{X,Y \in C} \Pr[X \in (x - h, x + h), Y \in (x' - h', x' + h'), W \in B_{x' \theta}] P[B_{x' \theta}] +
\]

\[
\Pr[X \in (x - h, x + h), Y' \in (x' - h', x' + h'), W \in B_{x' \theta}] P[B_{x' \theta}] +
\]

\[
\Pr[X' \in (x - h, x + h), Y \in (x' - h', x' + h'), W \in B_{x' \theta}] P[B_{x' \theta}] +
\]

\[
\Pr[X' \in (x - h, x + h), Y' \in (x' - h', x' + h'), W \in B_{x' \theta}] P[B_{x' \theta}]. \quad (5.12)
\]
The above four terms on the right hand side are equal to each other. Therefore, it is sufficient to derive the value of the first term only. This term when divided by $4hh'$ and its limit taken as $h \to 0$ and $h' \to 0$, becomes

$$f(x,Y|w)(x,x')\frac{\delta(x,x')}{l_1}, \text{ where } w \in B_{xx'}. \quad (5.13)$$

Now, $f(x,Y|w)(x,x') = \frac{1}{l_1}$ because for any $w$ in $B_{xx'}$, $X$ and $Y$ are conditionally independently distributed with uniform density $1/l_1$. We conclude that, when we divide the first term on the right hand side of (5.10) by $4hh'$ and take its limit as $h \to 0$ and $h' \to 0$, it simplifies to

$$\left(\frac{\delta(x,x')}{l_1}\right)\left(\frac{4}{l_1^2}\right). \quad (5.14)$$

The second term on the right hand side of (5.10) can be written as

$$\sum_{X,Y \in C} Pr[X \in (x - h, x + h), Y \in (x' - h', x' + h') | W \in A_{xx'}]Pr[A_{xx'}].$$

Given that the event $(W \in A_{xx'})$, $x$ and $x'$ are within one grid cell. Then, there are two r.v., say $X$ and $X'$, in $C$ whose conditional domain is the grid cell containing $x$ and $x'$ simultaneously. Then using the second term on the right hand side of (5.10), we can write

$$\sum_{X,Y \in C} Pr[X \in (x - h, x + h), Y \in (x' - h', x' + h') | W \in A_{xx'}]Pr[A_{xx'}] = Pr[X \in (x - h, x + h), X' \in (x' - h', x' + h') | W \in A_{xx'}]Pr[A_{xx'}] +$$
Equation (5.15) is nothing but

\[ Pr[X' \in (x - h, x + h), X \in (x' - h', x' + h')] | W \in A_{xx'} Pr[A_{xx'}]. \]  

(5.15)

The two terms on the right hand side of (5.16) are equal to each other because of the symmetry between \( X \) and \( X' \). Also, for sufficiently small positive numbers \( h \) and \( h' \) we can prove the following equality

\[ [X \in (x - h, x + h), X' \in (x' - h', x' + h')] , W \in A_{xx'} ] = [X \in (x - h, x + h), X' \in (x' - h', x' + h')] . \]  

(5.16)

It is evident that the left hand side of (5.17) is included in the right hand side. The problem then is reduced to only showing that the right hand side is included in the left hand side and this can be proven by using the following theorem.

**Theorem 5.1** For any \( x \) and \( x' \) which satisfy \( 0 < |x - x'| < l_1 \), there exist sufficient small positive numbers \( h \) and \( h' \) such that if \( x - h < X < x + h \) and \( x' - h' < X' < x' + h' \), then \( W \in A_{xx'} \).

**Proof:** Suppose \( x' < x \). From the design we can see that \( W \) is related to \( X \) and \( X' \) by

\[ W + kl_1 = \frac{X + X'}{2} - \frac{l_1}{2}. \]  

(5.18)
for some integer \( k \). We can write

\[
\frac{X + X'}{2} - \frac{l_1}{2} < \frac{x + x'}{2} - \frac{l_1}{2} + \frac{h + h'}{2},
\]

(5.19)

and

\[
\frac{X + X'}{2} + \frac{l_1}{2} > \frac{x + x'}{2} + \frac{l_1}{2} - \frac{h + h'}{2}.
\]

(5.20)

It can be seen that \( \frac{1}{2} - x + \frac{x + x'}{2} > 0 \) because \( 0 < x - x' < l_1 \). Select \( h \) and \( h' \) to satisfy

\[
0 < h, \ h' \leq \frac{l_1}{2} - x + \frac{x + x'}{2}.
\]

(5.21)

By combining (5.19), (5.20) and (5.21), we obtain

\[
\frac{X + X'}{2} - \frac{l_1}{2} < x' \quad \text{and} \quad \frac{X + X'}{2} + \frac{l_1}{2} > x.
\]

(5.22)

Making use of (5.18) in (5.22) yields

\[
W + kl_1 < x' \quad \text{and} \quad W + (k + 1)l_1 > x.
\]

(5.23)

The two inequalities given in (5.23) prove that \( W \in A_{xx'} \). Similar arguments can be made for the case where \( x' > x \). \( \square \)

Using the above theorem, we can conclude that the right hand side of (5.17) is included in the left hand side. Finally, the proof of equation (5.17) is completed.

By making use of (5.17), equation (5.16) can be written as
\[
\sum_{x,y \in \mathcal{C}} \Pr[X \in (x-h, x+h), Y \in (x'-h', x'+h'), W \in \mathcal{A}_{xx'}] = \\
\Pr[X \in (x-h, x+h), X' \in (x'-h', x'+h')] + \\
\Pr[X' \in (x-h, x+h), X \in (x'-h', x'+h')].
\]

(5.24)

Dividing equation (5.24) by \(4hh'\) and taking its limit, we obtain

\[
\lim_{h \to 0, h' \to 0} \sum_{x,y \in \mathcal{C}} \frac{\Pr[X \in (x-h, x+h), Y \in (x'-h', x'+h'), W \in \mathcal{A}_{xx'}]}{4hh'} = f(x, x')(x', x') + f(x', x)(x, x').
\]

(5.25)

The second term on the right hand side of (5.10) depends on the unconditional joint density of \((X, X')\) which is the same as the density of \((X', X)\).

Calculation of the Unconditional Joint Density of \((X, X')\):

Suppose \(x < x'\), and let \(W'\) be the r.v. associated with the realized grid boundary such that \(x + \frac{x'}{2} - \frac{h}{2}\) falls in its domain (\(x + \frac{x'}{2} - \frac{h}{2}\) is in one cell of the translated grid with probability 1). We use \(w', v,\) and \(v'\) as the realized values of \(W'\), \(X\), and \(X'\) respectively. Without loss of generality, assume that \(W'\) is distributed as uniform \((0, l_1)\), \((X|w')\) distributed as uniform \((w', w' + l_1)\), and \((X'|w')\) distributed as uniform \((w', w' + l_1)\). Note that the r.v. \(W'\) is equal to the r.v. \(W\) plus a constant. By using \(f(x, w')(v, w') = f(x|w')(v) f_{W'}(w')\), the joint density of \((X, W')\) can be given as
\[ f_{X,W}(v, w') = \frac{1}{l_1^2}, \quad 0 < w' < l_1, \quad w' < v < w' + l_1. \]  

Consider the transformation from \((X, W')\) to \((Z, X')\), given by

\[
\begin{cases}
Z = X \\
X' = 2W' - X + l_1
\end{cases}
\]  

This transformation is one to one and the inverse transformation is given by

\[
\begin{cases}
X = Z \\
W' = \frac{X' + Z - l_1}{2}
\end{cases}
\]

The Jacobian of the transformation is given by

\[ J = \begin{vmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix} = \frac{1}{2}. \]

Because we have \(Z = X\), the joint density of \((Z, X')\) is the same as the joint density of \((X, X')\) and is given by

\[ f_{X,X'}(v, v') = \frac{1}{2l_1^2}, \quad -l_1 < v - v' < l_1 < v + v' < 3l_1. \]

Because \((x, x')\) in (5.25) is a realized point of \((X, X')\) and \((X', X)\), we have

\[ f_{X,X'}(x, x') = \frac{1}{2l_1^2}, \quad \text{and} \quad f_{X,X}(x, x') = \frac{1}{2l_1^2}. \]

As a result, the right hand side of (5.25) can be written as

\[ \frac{1}{l_1^2}. \]
Finally, when (5.10) is divided by $4h h'$ and its limit is taken, it results in the sum of (5.14) and (5.30). Therefore, we obtain

$$
\pi(x, x') = \left( \frac{1}{4} + \frac{\delta(x, x')}{l_1} \right) \left( \frac{4}{l_1^2} \right), \quad \delta(x, x') < l_1.
$$

(5.31)

The results of this section can be summarized by

$$
\pi(x) = \frac{2}{l_1}, \quad x \in U,
$$

and for any $x \neq x' \in U$,

$$
\pi(x, x') = \begin{cases} 
\frac{4}{l_1^2}, & \delta(x, x') > l_1 \\
\left( \frac{1}{4} + \frac{\delta(x, x')}{l_1} \right) \left( \frac{4}{l_1^2} \right), & \delta(x, x') < l_1.
\end{cases}
$$

(5.32)

5.3 The Mean Balanced Design from an $n$-dimensional Space

Suppose the domain of the response variable is of $n$ dimensions. Use the universe $U$ as defined in section (2.2) and the grid superimposed on $U$ whose grid length is $l_i$ with respect to (wrt) $e_i$ axis and $m_i = \frac{M_i}{l_i}$. For any $x \in U$ there exist a vector of integers $k_x$ and $w_x \in \Omega$ such that $w_x + k_x l = x$, where $l = (l_1, l_2, \ldots, l_n)$. Define $\Omega$ to be the sample space of $W$, that is

$$
\Omega = \{ w = (w_1, w_2, \ldots, w_n) : 0 < w_i < l_i, \ i = 1, \ldots, n \}.
$$
In this section we calculate the inclusion density functions for which $U$ is in $n$ dimensions.

The following notations are used:

$h = (h_1, h_2, ..., h_n)$, $h' = (h'_1, h'_2, ..., h'_n)$, and $h \to 0$ shall be used to represent $h_i \to 0$ for $i = 1, 2, ..., n$. The same applies to $h'$.

The first order inclusion density for any $x$ in $U$ is given by

$$
\pi(x) = \sum_{x \in C} f_x(x)
$$

$$
= \sum_{x \in C} \lim_{h \to 0} \frac{Pr[x - h < X < x + h]}{(2^n)h_1h_2...h_n}
$$

$$
= \lim_{h \to 0} \sum_{x \in C} \frac{Pr[x - h < X < x + h, W \in \Omega_x]}{(2^n)h_1h_2...h_n}
$$

$$
= \lim_{h \to 0} \sum_{x \in C} \frac{Pr[x - h < X < x + h | W \in \Omega_x] Pr[\Omega_x]}{(2^n)h_1h_2...h_n},
$$

where $\Omega_x = \Omega - \{w_x\}$.

Given $(W \in \Omega - \{w_x\})$, $x$ is inside one grid cell. Therefore, there are two possible random variables in $C$ which can assume the value of $x$. Suppose $X$ and $X'$ are these r.v. which can assume $x$ and have the grid cell containing $x$ as their conditional domain. Then, we can write

$$
\pi(x) = \lim_{h \to 0} \left\{ \frac{Pr[x - h < X < x + h | W \in \Omega - \{w_x\}]}{(2^n)h_1h_2...h_n} \right\} + \frac{Pr[x - h < X' < x + h | W \in \Omega - \{w_x\}]}{(2^n)h_1h_2...h_n},
$$

and that is nothing but

126
\[ \pi(x) = f(x|w)(x) + f(x'|w)(x), \quad (5.35) \]

where \( w \) is a realized value of \( W \) in \( \Omega - \{ w_x \} \). Given \( w \) in \( \Omega - \{ w_x \} \), then the conditional densities of \( X \) and \( X' \) are equal to \( \frac{1}{l_1 l_2 \ldots l_n} \). As a result, we obtain

\[
\pi(x) = \frac{1}{l_1 l_2 \ldots l_n} + \frac{1}{l_1 l_2 \ldots l_n} \\
= \frac{2}{l_1 l_2 \ldots l_n}. \quad (5.36)
\]

The joint inclusion density \( \pi(x, x') \), \( x \neq x' \in U \), can be written as the limit, as \( h \to 0 \) and \( h' \to 0 \), of

\[
\sum_{X, Y \in \mathcal{C}} Pr(x - h < X < x + h, x' - h' < Y < x' + h') \\
\div (2^{2n}) ! h_1 h_2 \ldots h_n h'_1 h'_2 \ldots h'_n. \quad (5.37)
\]

To calculate \( \pi(x, x') \), two possible cases can be considered:

Case 1: \( \delta(x_i, x'_i) \geq l_i \) for at least some \( i \).

Recall that \( B_{XX'} \) is defined to be the event that \( x \) and \( x' \) are in two different cells of the translated grid, and \( A_{XX'} \) is the event that some cell of the randomly translated grid contains both \( x \) and \( x' \).

Expression (5.37) can be written as

\[
\sum_{X, Y \in \mathcal{C}} Pr(x - h < X < x + h, x' - h' < Y < x' + h', W \in \Omega) \\
\div (2^{2n}) ! h_1 h_2 \ldots h_n h'_1 h'_2 \ldots h'_n. \quad (5.38)
\]
In this case \( B_{xx'} = \Omega \) a.e., then (5.38) is the same as

\[
\sum_{X, Y \in C} \frac{Pr(x - h < X < x + h, x' - h' < Y < x' + h' \mid W \in B_{xx'})}{(2^{2n})h_1 h_2 \cdots h_n h'_1 h'_2 \cdots h'_n}. \tag{5.39}
\]

Given \((W \in B_{xx'})\), \(x\) and \(x'\) are in two different grid cells. Therefore, there are two pairs of r.v. in \( C \), where the first one can assume the value \(x\) and the second can assume \(x'\). The first pair, suppose \(X\) and \(X'\), has the grid cell containing \(x\) as its conditional domain and the other pair, say \(Y\) and \(Y'\), has the grid cell containing \(x'\) as its conditional domain. Then, the numerator of (5.39) can be written as

\[
Pr[X \in (x - h, x + h) \cap Y \in (x' - h', x' + h') \mid W \in B_{xx'}] + \\
Pr[X \in (x - h, x + h) \cap Y' \in (x' - h', x' + h') \mid W \in B_{xx'}] + \\
Pr[X' \in (x - h, x + h) \cap Y \in (x' - h', x' + h') \mid W \in B_{xx'}] + \\
Pr[X' \in (x - h, x + h) \cap Y' \in (x' - h', x' + h') \mid W \in B_{xx'}]. \tag{5.40}
\]

Each one of the above four terms, when divided by \(2^{2n} \cdot h_1 h_2 \cdots h_n h'_1 h'_2 \cdots h'_n\) and its limit is taken as \(h \to 0\) and \(h' \to 0\), equals \(\frac{1}{l_1 l_2 \cdots l_n}\). Therefore, by making use of (5.37) and the above results, we obtain

\[
\pi(x, x') = \frac{4}{l_1 l_2 \cdots l_n}, \text{ if } \delta(x_i, x'_i) > l_i \text{ for some } i. \tag{5.41}
\]

Case 2: when \(\delta(x_i, x'_i) < l_i\) for every \(i = 1, 2, \ldots, n\).

Using numerator of (5.37), we can write
\[
\sum_{X,Y \in \mathcal{C}} Pr[X \in (x-h,x+h), Y \in (x'-h',x'+h'), W \in \Omega] = \\
\sum_{X,Y \in \mathcal{C}} Pr[X \in (x-h,x+h), Y \in (x',x'+h'), W \in B_{x'x}] + \\
\sum_{X,Y \in \mathcal{C}} Pr[X \in (x-h,x+h), Y \in (x'-h',x'+h'), W \in A_{xx'}]. \hspace{1cm} (5.42)
\]

The first term on the right hand side of (5.42) can be written as
\[
\sum_{X,Y \in \mathcal{C}} Pr[X \in (x-h,x+h), Y \in (x'-h',x'+h') | W \in B_{x'x}] Pr[B_{x'x}]. \hspace{1cm} (5.43)
\]

Given \((W \in B_{xx'})\), then one of the grid boundaries falls between \(x\) and \(x'\). There are two possible pairs of r.v. in \(C\) where the first one can assume the value of \(x\) and the second can assume \(x'\). The first pair, suppose \(X\) and \(X'\), has the grid cell containing \(x\) as its conditional domain and the other pair, say \(Y\) and \(Y'\), has the grid cell containing \(x'\) as its conditional domain. Then, (5.43) can be expanded as
\[
\sum_{X,Y \in \mathcal{C}} Pr[X \in (x-h,x+h), Y \in (x'-h',x'+h') | W \in B_{x'x}] Pr[B_{xx'}] = \\
Pr[X \in (x-h,x+h), Y \in (x'-h',x'+h') | W \in B_{x'x}] Pr[B_{xx'}] + \\
Pr[X \in (x-h,x+h), Y' \in (x'-h',x'+h') | W \in B_{x'x}] Pr[B_{xx'}] + \\
Pr[X' \in (x-h,x+h), Y \in (x'-h',x'+h') | W \in B_{x'x}] Pr[B_{xx'}] + \\
Pr[X' \in (x-h,x+h), Y' \in (x'-h',x'+h') | W \in B_{x'x}] Pr[B_{xx'}]. \hspace{1cm} (5.44)
\]

The above four terms on the right hand side are equal to each other. Therefore it is sufficient to derive the value of the first term only, which when divided by \(2^{2n}h_1h_2...h_nh_1'h_2'...h_n'\) and its limit taken as \(h \to 0\) and \(h' \to 0\) is given by

129

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\[ f(x, y | w)(x, x') \, Pr[B_{xx'}], \quad (5.45) \]

where \( w \in B_{xx'} \) and \( Pr[B_{xx'}] \) is given in (2.28).

Now, \( f(x, y | w)(x, x') = \frac{1}{h_1 \ldots h_n} \) because \( w \) is in \( B_{xx'} \), \( X \) and \( Y \) are conditionally independently distributed with uniform density \( \frac{1}{h_1 \ldots h_n} \). We conclude that, when we divide the first term on the right hand side of (5.42) by \( 2^{n} h_1 \ldots h_n h'_1 \ldots h'_n \) and we take its limit, as \( h \to 0 \) and \( h' \to 0 \), it becomes

\[ \left( \frac{4}{1^2 2^2 \ldots n^2} \right) \, Pr[B_{xx'}]. \quad (5.46) \]

The second term on the right hand side of (5.42) can be written as

\[ \sum_{x, y \in \mathcal{C}} Pr[X \in (x - h, x + h), Y \in (x' - h', x' + h') | W \in A_{xx'}] \, Pr[A_{xx'}]. \quad (5.47) \]

Given \((W \in A_{xx'})\), \( x \) and \( x' \) are inside one grid cell. Then, there are two r.v., suppose \( X \) and \( X' \), in \( \mathcal{C} \) whose conditional domain is the grid cell containing \( x \) and \( x' \). Using the second term on the right hand side of (5.42), we can write

\[ \sum_{x, y \in \mathcal{C}} Pr[X \in (x - h, x + h), Y \in (x' - h', x' + h') | W \in A_{xx'}] \, Pr[A_{xx'}] = \]

\[ Pr[X \in (x - h, x + h), X' \in (x' - h', x' + h') | W \in A_{xx'}] \, Pr[A_{xx'}] + Pr[X' \in (x - h, x + h), X \in (x' - h', x' + h') | W \in A_{xx'}] \, Pr[A_{xx'}]. \quad (5.48) \]

Equation (5.48) is nothing but

130
\[
\sum_{X,Y \in \mathbb{C}} Pr[X \in (x-h, x+h), Y \in (x'-h', x'+h')] \mid W \in A_{xx'}] Pr[A_{xx'}] = \\
Pr[X \in (x-h, x+h), X' \in (x'-h', x'+h'), W \in A_{xx'}] + \\
Pr[X' \in (x-h, x+h), X \in (x'-h', x'+h'), W \in A_{xx'}].
\] (5.49)

The two terms on the right hand side of (5.49) are equal to each other because of the symmetry between \(X\) and \(X'\). Also, given that \(h\) and \(h'\) with their coordinates are being sufficiently small positive numbers, we can prove the following equality:

\[
[X \in (x-h, x+h), X' \in (x'-h', x'+h'), W \in A_{xx'}]
= [X \in (x-h, x+h), X' \in (x'-h', x'+h')].
\] (5.50)

It is evident that the left hand side of (5.50) is included in the right hand side. The problem is then reduced to showing that the right hand side is included in the left hand side. This can be proven by using the following theorem.

**Theorem 5.2** For any \(x\) and \(x'\) which satisfy \(0 < |x_i - x'_i| < l_i\) for every \(i = 1, 2, \ldots, n\), we can find two vectors, \(h\) and \(h'\), such that their coordinates are sufficiently small positive numbers and if \(x_i - h_i < X_i < x_i + h_i\) and \(x'_i - h'_i < X'_i < x'_i + h'_i\) for every \(i\), then \(W \in A_{xx'}\).

**Proof:** From the sampling design we can see that \(W\) is related to \(X\) and \(X'\) by

\[
W_i + k_i l_i = \frac{X_i + X'_i}{2} - \frac{l_i}{2},
\]

131
for some integer $k_i$ and for every $i = 1, 2, \ldots, n$. Then, for each $i$ we have the same hypothesis as in Theorem 1. As a result, if $x_i' < x_i$, then

$$W_i + k_i l_i < x_i' \quad \text{and} \quad W_i + (k_i + 1) l_i > x_i,$$

(5.51)

and if $x_i < x_i'$, then

$$W_i + k_i l_i < x_i \quad \text{and} \quad W_i + (k_i + 1) l_i > x_i'.$$

(5.52)

The inequalities given in (5.51) and (5.52) prove that $W$ is in one cell of the translated grid. Therefore, $W \in A_{xx'}$. □

Using the above theorem, we can conclude that the right hand side of (5.50) is included in the left hand side. This completes the proof of (5.50).

By making use of (5.50), equation (5.49) can be written as

$$\sum_{X,Y \in c} Pr[X \in (x - h, x + h), Y \in (x' - h', x' + h'), W \in A_{xx'}] =$$

$$Pr[X \in (x - h, x + h), X' \in (x' - h', x' + h')]+$$

$$Pr[X' \in (x - h, x + h), X \in (x' - h', x' + h')].$$

(5.53)

Dividing equation (5.53) by $2^{2n} h_1 h_2 \ldots h_n h_1' h_2' \ldots h_n'$ and taking its limit, we obtain

$$\lim_{h \to 0, h' \to 0} \sum_{X,Y \in c} \frac{Pr[X \in (x - h, x + h), Y \in (x' - h', x' + h'), W \in A_{xx'}]}{2^{2n} h_1 h_2 \ldots h_n h_1' h_2' \ldots h_n'}$$

$$= f(x,x')(x,x') + f(x',x)(x,x').$$

(5.54)
Therefore, the second term on the right hand side of (5.48) depends on the unconditional joint density of \((X, X')\) which is the same as the density of \((X', X)\).

**Calculation of the Unconditional Joint Density of \((X, X')\):**

Let \(W'\) be the r.v. associated with the intersection of the grid boundaries such that the point

\[
\left(\frac{x_1 + x_1'}{2}, \frac{x_2 + x_2'}{2}, \ldots, \frac{x_n + x_n'}{2} - \frac{l_n}{2}\right)
\]

falls in its domain. We use \(w' = (w'_1, w'_2, \ldots, w'_n)\), \(v = (v_1, v_2, \ldots, v_n)\), and \(v' = (v'_1, v'_2, \ldots, v'_n)\) as the realizations of the r.v. \(W', X\) and \(X'\) respectively. Note that the r.v. \(W'\) is equal to the r.v. \(W\) plus a vector of constants. Without loss of generality, we assume \(W'_i\) is distributed as uniform \((0, l_i)\), \((X_i|w')\) is distributed as uniform \((w'_i, w'_i + l_i)\), and \((X_i'|w')\) is distributed as uniform \((w'_i, w'_i + l_i)\), for \(i = 1, 2, \ldots, n\). By using \(f_{(X,W)'}(v, w') = f_{(X|w')}(v) f_{W'}(w')\), we obtain the joint density of \((X, W')\), that is

\[
f_{(X,W)'}(v, w') = \frac{1}{l_1^2 l_2^2 \cdots l_n^2}, \quad 0 < w'_i < l_i, \quad w'_i < v_i < w'_i + l_i, \quad i = 1, \ldots, n.
\]

Consider the transformation from \((X, W')\) to \((Z, X')\),

\[
\begin{align*}
\{ Z_i &= X_i \\
X'_i &= 2W'_i - X_i + l_i \}
\end{align*}
\]

\(i = 1, 2, \ldots, n. \) (5.55)

This transformation is one to one, and the inverse transformation is given by
\[
\begin{align*}
\begin{cases}
X_i = Z_i \\
W'_i = \frac{X_i + Z_i - l_i}{2}
\end{cases}, \quad i = 1, 2, \ldots, n.
\end{align*}
\tag{5.56}
\]

The Jacobian of the transformation is given by the determinant of a \(2n\) by \(2n\) matrix as

\[
J = \begin{vmatrix}
1 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
\vdots & & & \ddots & & & & & \vdots \\
0 & 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0 & \frac{1}{2} & 0 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0 & 0 & \frac{1}{2} & \ldots & 0 \\
\vdots & & & \ddots & & & & & \vdots \\
0 & 0 & 0 & \ldots & \frac{1}{2} & 0 & 0 & \ldots & \frac{1}{2}
\end{vmatrix} = \frac{1}{2^n}. \tag{5.57}
\]

The joint density of \((X, X')\) is the same as the joint density of \((Z, X')\) because \(Z = X\) in the transformation. Then, the joint density of \((X, X')\) is given by

\[
f(x, x')(v, v') = \frac{1}{2^n l_1^2 l_2^2 \ldots l_n^2}, \quad -l_i < v_i - v'_i < l_i < v_i + v'_i < 3l_i, \quad i = 1, \ldots, n.
\]

Since \((x, x')\) in (5.54) is a realized point of \((X, X')\) and \((X', X)\), we have

\[
f(x, x')(x, x') = \frac{1}{2^n l_1^2 l_2^2 \ldots l_n^2}, \quad \text{and} \quad f(x', x)(x, x') = \frac{1}{2^n l_1^2 l_2^2 \ldots l_n^2}.
\]

As a result, the right hand side of (5.54) can be written as

\[
\frac{1}{2^{(n-1)} l_1^2 l_2^2 \ldots l_n^2}. \tag{5.58}
\]

134

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Finally, when (5.42) is divided by $2^{2n}h_1h_2...h_nh_1'h_2'...h_n'$ and on taking its limit, it results in the sum of (5.46) and (5.58). Therefore, we obtain

$$\pi(x, x') = \left( \frac{1}{2^{(n-1)}l_1l_2...l_n^2} \right) + \frac{4}{l_1^2l_2...l_n^2} Pr[B_{xx'}], \quad (5.59)$$

where $Pr[B_{xx'}]$ is given in (2.28).

From (5.36), (5.41), and (5.59), we summarize the results by

$$\pi(x) = \frac{2}{l_1l_2...l_n}, \quad x \in U,$$

and for any $x \neq x' \in U$,

$$\pi(x, x') = \begin{cases} 
\frac{4}{l_1^2l_2...l_n^2}, & \delta(x_i, x'_i) > l_i \text{ for some } i \\
\left( \frac{1}{2^{(n+1)}} + Pr[B_{xx'}] \right) \frac{4}{l_1^2l_2...l_n^2}, & \delta(x_i, x'_i) < l_i \text{ for every } i.
\end{cases} \quad (5.60)$$

By substituting the above inclusion density functions in the Yates-Grundy variance estimator given in (1.15), we can obtain an unbiased estimator of the variance of the estimator of the population total. Using simulation, we noticed that this variance estimator can take negative values, and it is highly unstable to be practical. Therefore, finding a nonnegative variance estimator for this sampling design is still an open problem.
Bibliography


136

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Biographical Statement

Place of Birth

Ras Baalbeck, Lebanon

Date of Birth

February 08, 1961

Education

B.S. in Statistics, 1986, Lebanese University, Beirut, Lebanon

M.S. in Applied Mathematics (Statistics Option), 1990, Old Dominion University, Norfolk, Virginia

Employment

Applied Marine Research Laboratory, Old Dominion University, 1991-1992, Research Assistant.


Center for Pediatric Research, Eastern Virginia Medical School, 1993, Research Associate.