LARGE SAMPLE THEORY FOR U-STATISTICS
IN UNEQUAL PROBABILITY SAMPLES

by

Rick L. Williams

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1860T

February 1969
LARGE SAMPLE THEORY FOR U-STATISTICS
IN UNEQUAL PROBABILITY SAMPLES

by

Rick L. Williams

A Dissertation submitted to the faculty of
The University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the
Department of Biostatistics

Chapel Hill

1988

Approved by:

[Signatures]

Advisor

Reader

Reader
RICK L. WILLIAMS. Large Sample Theory for U-Statistics in Unequal Probability Samples (Under the direction of Parnab K. Sen and Babubhai V. Shah.)

ABSTRACT

Since their first description by Hoeffding in 1948, U-statistics have been an important area of research and application. Their extension to simple random sampling without replacement was explored by Nandi and Sen in 1963. In 1984, Folsom developed the framework for U-statistics under a general unequal probability sample from a finite population.

This research builds upon the work cited above to develop a large sample theory for U-statistics in an unequal probability sample. A projection approach is used to determine the limiting distribution. The projection in the case of a general unequal probability sample from a finite population is determined. It is shown that the U-statistic and its projection are asymptotically equivalent in quadratic mean. The central limit theorem for the projection then extends to the U-statistic.

The developments are applied to Sampford's method of unequal probability sampling without replacement. An asymptotic approximation for the rth degree joint inclusion probabilities for Sampford's method is derived.

Finally, a numerical simulation is presented to demonstrate the applicability of the research.
ACKNOWLEDGEMENTS

I thank my committee members -- Parnab K. Sen, Babubhai Shah, William Kalsbeek, Ralph Folsom and Carl Shy -- for their help and patience during this research. I especially thank Dr. Sen for convincing me that I could still finish when I thought it was hopeless.

Further thanks are due the Research Triangle Institute and my many friends at RTI. Without their support and understanding this effort would not have come to fruition. Special acknowledgements are due Pat Parker, for her fine preparation of the manuscript, and Vincent Iannacchione and Jenny Milne, for preparing the figures.

Finally, I dedicate this dissertation to the memory of my best friend -- Douglas J. Drummond.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES...</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES...</td>
<td>viii</td>
</tr>
<tr>
<td>I. INTRODUCTION AND LITERATURE REVIEW</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Definitions and Preliminaries</td>
<td>2</td>
</tr>
<tr>
<td>1.2.1 The Spaces C[0,1] and D[0,1]</td>
<td>3</td>
</tr>
<tr>
<td>1.2.2 Wiener Measure and Brownian Motion</td>
<td>5</td>
</tr>
<tr>
<td>1.2.3 Selected Properties of Martingales</td>
<td>6</td>
</tr>
<tr>
<td>1.3 U-Statistics and Von Mises' Differentiable Statistical Functions</td>
<td>8</td>
</tr>
<tr>
<td>1.3.1 Estimators and Variances</td>
<td>8</td>
</tr>
<tr>
<td>1.3.2 Weak Convergence of ( {U_n} ) and ( {V_n} )</td>
<td>12</td>
</tr>
<tr>
<td>1.3.3 Projection Method</td>
<td>15</td>
</tr>
<tr>
<td>1.4 Some Central Limit Theorems for Without Replacement Sampling</td>
<td>19</td>
</tr>
<tr>
<td>1.5 Overview of Results</td>
<td>21</td>
</tr>
<tr>
<td>II. LARGE SAMPLE THEORY</td>
<td>24</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>24</td>
</tr>
<tr>
<td>2.2 A Sequence of Populations and Samples</td>
<td>24</td>
</tr>
<tr>
<td>2.3 Sample Inclusion Indicators</td>
<td>25</td>
</tr>
<tr>
<td>2.4 Degree m Population Total and Estimator</td>
<td>30</td>
</tr>
<tr>
<td>2.5 Components of Variance and Covariance</td>
<td>32</td>
</tr>
<tr>
<td>2.6 The Projection of a U-Statistic</td>
<td>39</td>
</tr>
<tr>
<td>2.7 Verification of ( \hat{U} )</td>
<td>44</td>
</tr>
<tr>
<td>2.7.1 Expected Value of ( \hat{U} )</td>
<td>44</td>
</tr>
<tr>
<td>2.7.2 With Replacement Sampling</td>
<td>46</td>
</tr>
<tr>
<td>2.7.3 Simple Random Sampling Without Replacement</td>
<td>47</td>
</tr>
<tr>
<td>2.8 Equivalence of ( U ) and ( \hat{U} ) in Quadratic Mean for Without Replacement Sampling</td>
<td>48</td>
</tr>
<tr>
<td>III. SAMPFORD'S METHOD OF UNEQUAL PROBABILITY SAMPLING WITHOUT REPLACEMENT</td>
<td>68</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>68</td>
</tr>
<tr>
<td>3.2 Sampford's Method of Sample Selection</td>
<td>68</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS
(continued)

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3</td>
<td>70</td>
</tr>
<tr>
<td>Approximate Multiple Inclusion Probabilities for Sampford's Method</td>
<td></td>
</tr>
<tr>
<td>3.4</td>
<td>78</td>
</tr>
<tr>
<td>Numerical Illustration</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>86</td>
</tr>
<tr>
<td>Large Sample Normality for Sampford's Method</td>
<td></td>
</tr>
<tr>
<td>IV.</td>
<td>90</td>
</tr>
<tr>
<td>NUMERICAL SIMULATION</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>91</td>
</tr>
<tr>
<td>The Data</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>93</td>
</tr>
<tr>
<td>Design of the Simulation</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>96</td>
</tr>
<tr>
<td>Findings</td>
<td></td>
</tr>
<tr>
<td>4.4</td>
<td>107</td>
</tr>
<tr>
<td>Conclusions</td>
<td></td>
</tr>
<tr>
<td>V.</td>
<td>110</td>
</tr>
<tr>
<td>RECOMMENDATIONS FOR FUTURE RESEARCH</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>110</td>
</tr>
<tr>
<td>Weak Convergence to a Gaussian Process</td>
<td></td>
</tr>
<tr>
<td>5.2</td>
<td>111</td>
</tr>
<tr>
<td>Variance Estimation</td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>113</td>
</tr>
<tr>
<td>Extension to Other Sample Designs</td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td>115</td>
</tr>
<tr>
<td>APPENDIX: Figures for the Numerical Simulation</td>
<td>121</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

Table 3.1: Recorded Acreage of Crops and Grass for 1947 for 35 Farms in Orkney........................................... 80

Table 3.2: Exact and Approximate Joint Inclusion Probabilities for Sampford's Method with n=5 for the Data in Table 3.1................................................................. 81

Table 3.3: Exact and Approximate Triple Inclusion Probabilities for Sampford's Method with n=5 for the Data in Table 3.1................................................................. 82

Table 3.4: Exact and Approximate Quadruple Inclusion Probabilities for Sampford's Method with n=5 for the Data in Table 3.1................................................................. 84

Table 3.5: Mean Absolute Percent Relative Difference Between the Exact and Approximate Inclusion Probabilities with n=5 for the Cases in Tables 3.2, 3.3 and 3.4.................................................. 87

Table 4.1: Descriptive Statistics on the Study Population of 2,000 U.S. Counties........................................... 92

Table 4.2: Pearson Correlation Matrix Among the Study Variables for 2,000 U.S. Counties........................................... 94

Table 4.3: Mean and Standard Error for Kendall's Rank Correlation from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure........................................ 97

Table 4.4: Mean and Standard Error for Relative With Replacement Variance Component from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure.. 98

Table 4.5: Skewness and Kurtosis for Kendall's Rank Correlation from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure........................................ 100

Table 4.6: Skewness and Kurtosis for Relative With Replacement Variance Component from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure........ 101

Table 4.7: Kolmogorov-Smirnov (K-S) Test of Normality for Kendall's Rank Correlation from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure.. 103

Table 4.8: Kolmogorov-Smirnov (K-S) Test of Normality for Relative With Replacement Variance Component from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure........................................ 105
LIST OF TABLES
(continued)

Table 4.9: Empirical Tail Probabilities for Kendall's Rank Correlation from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure..................106

Table 4.10: Empirical Tail Probabilities for Relative With Replacement Variance Component from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure..108
LIST OF FIGURES

Figure A.1: Frequency Distribution Kendall's Rank Correlation 1,000 Replicated Samples of Size 50 PPS to Population Size.................................122

Figure A.2: Frequency Distribution Kendall's Rank Correlation 1,000 Replicated Samples of Size 100 PPS to Population Size.................................123

Figure A.3: Frequency Distribution Kendall's Rank Correlation 1,000 Replicated Samples of Size 50 PPS to Root Population Size.................................124

Figure A.4: Frequency Distribution Kendall's Rank Correlation 1,000 Replicated Samples of Size 100 PPS to Root Population Size.................................125

Figure A.5: Frequency Distribution Kendall's Rank Correlation 1,000 Replicated Samples of Size 50 PPS to Log Population Size.................................126

Figure A.6: Frequency Distribution Kendall's Rank Correlation 1,000 Replicated Samples of Size 100 PPS to Log Population Size.................................127

Figure A.7: Frequency Distribution With Replacement Variance Component 1,000 Replicated Samples of Size 50 PPS to Population Size.................................128

Figure A.8: Frequency Distribution With Replacement Variance Component 1,000 Replicated Samples of Size 100 PPS to Population Size.................................129

Figure A.9: Frequency Distribution With Replacement Variance Component 1,000 Replicated Samples of Size 50 PPS to Population Size.................................130

Figure A.10: Frequency Distribution With Replacement Variance Component 1,000 Replicated Samples of Size 100 PPS to Root Population Size.................................131

Figure A.11: Frequency Distribution With Replacement Variance Component 1,000 Replicated Samples of Size 50 PPS to Log Population Size.................................132

Figure A.12: Frequency Distribution With Replacement Variance Component 1,000 Replicated Samples of Size 100 PPS to Log Population Size.................................133
LIST OF FIGURES
(continued)

Figure A.13: Normal Probability Plot Kendall's Rank Correlation
1,000 Replicated Samples of Size 50 PPS to Population Size.........................134

Figure A.14: Normal Probability Plot Kendall's Rank Correlation
1,000 Replicated Samples of Size 100 PPS to Population Size........................135

Figure A.15: Normal Probability Plot Kendall's Rank Correlation
1,000 Replicated Samples of Size 50 PPS to Population Size........................136

Figure A.16: Normal Probability Plot Kendall's Rank Correlation
1,000 Replicated Samples of Size 100 PPS to Root Population Size........................137

Figure A.17: Normal Probability Plot Kendall's Rank Correlation
1,000 Replicated Samples of Size 50 PPS to Log Population Size........................138

Figure A.18: Normal Probability Plot Kendall's Rank Correlation
1,000 Replicated Samples of Size 100 PPS to Log Population Size........................139

Figure A.19: Normal Probability Plot With Replacement Variance Component
1,000 Replicated Samples to Size 50 PPS to Population Size.............................140

Figure A.20: Normal Probability Plot With Replacement Variance Component
1,000 Replicated Samples to Size 100 PPS to Population Size.............................141

Figure A.21: Normal Probability Plot With Replacement Variance Component
1,000 Replicated Samples to Size 50 PPS to Root Population Size........................142

Figure A.22: Normal Probability Plot With Replacement Variance Component
1,000 Replicated Samples to Size 100 PPS to Root Population Size........................143

Figure A.23: Normal Probability Plot With Replacement Variance Component
1,000 Replicated Samples to Size 50 PPS to Log Population Size........................144

Figure A.24: Normal Probability Plot With Replacement Variance Component
1,000 Replicated Samples to Size 100 PPS to Log Population Size........................145
I. INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Most of the available analytic methods for finite population studies center around the estimation of simple descriptive statistics -- mainly population totals, means or functions of totals and means. This has resulted from the view of surveys as mainly being a device for the description of a population. In recent years more emphasis has been placed on the analysis of finite population surveys to determine underlying relationships. This has led to the development of a theory of regression inference from a finite population sample. Examples of this research are Fuller (1975), Kish and Frankel (1974), Shah, Holt and Folsom (1977), Nathan and Holt (1980) and Pfeffermann and Nathan (1981). Another thrust of research has been to use the categorical data analysis methods of Grizzle, Stramer and Koch (1969) with survey data. Koch, Freeman and Freeman (1975) provide a thorough description of the approach. This has been an active area of research as demonstrated in the papers of Holt, Scott, and Ewings (1980), Fellige (1980), Rao and Scott (1981, 1984 and 1987), Williams, Folsom and LaVange (1983) and Fay (1985). A third area of activity, due to Binder (1983), provides a method of accommodating logistic regression and log-linear contingency table models. Logistic Regression is also addressed by Roberts, Rao and Kumer (1987).
All of the above methods of analysis are usually placed under the general heading of parametric methods. Very little work has been done concerning nonparametric and robust methods. The original impetus for the current work was the need to analyze data under a complex sample from a highly skewed finite population. This leads to the consideration of nonparametric and robust methods. Some examples of recent research for medians in complex samples is Francisco and Fuller (1986), Williams and Perritt (1986), and Wheeless, Shah and LaVange (1988). The class of U-statistics includes many of the nonparametric and robust methods of interest. Folsom (1984 and 1986) extended Hoeffding's (1948) U-statistics to complex samples.

The lacking ingredient was an asymptotic distribution theory for U-statistics in complex samples to form the basis for statistical inference. This research is the first step in the development of such a theory. The case considered is a single-stage unequal probability sample with less than complete replacement. The main goal of the research is to provide the theoretical framework upon which applied inference can be based.

The remainder of this first chapter presents a review of related literature. This provides a background for the current research and suggests areas of future research.

1.2 Definitions and Preliminaries

In order to thoroughly review the past work on large sample theory for U-statistics, several concepts need to be developed. This section presents a brief overview of the spaces of continuous and discontinuous real-valued functions on the unit interval -- C[0,1] and D[0,1]. These spaces are necessary to the discussion of Wiener
processes which follows. Much of the current research on U-statistics has dealt with the weak convergence (convergence in distribution) of U-statistics to a Wiener process. Finally, the martingale class of dependent random variables is presented. U-statistics can be shown to form a martingale sequence.

1.2.1 The Spaces C[0,1] and D[0,1]

Let \( f:[0,1] \to \mathbb{R} \) be a real-valued function on the unit interval \([0,1]\). Then, \( C[0,1] \) is the space of all continuous real-valued functions \( f \) on \([0,1]\). Associate with \( C[0,1] \) the uniform topology by defining the metric

\[
\rho(x,y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)| \quad \text{where } x,y \in C[0,1]. \tag{1.1}
\]

Then, under the metric \( \rho \), \( C[0,1] \) is a complete and separable metric space.

An extension of \( C[0,1] \) is the space \( D[0,1] \) of real valued functions on \([0,1]\) that are right continuous and have a left-hand limit. That is, the function \( f:[0,1] \to \mathbb{R} \) is in \( D[0,1] \) if for \( 0 \leq t < 1 \)

\[
f(t+) = \lim_{s \uparrow t} f(s) \quad \text{exists and } f(t+) = f(t)
\]

and for \( 0 < t \leq 1 \)

\[
f(t-) = \lim_{s \uparrow t} f(s) \quad \text{exists.}
\]

Functions in \( D[0,1] \) are said to have only discontinuities of the first kind. A function \( f \) in \( D[0,1] \) has a jump at \( t \) if \( f(t) \neq f(t-) \).
Clearly, $C[0,1]$ is a subspace of $D[0,1]$ for which $f(t) = f(t^-)$ for all $0 \leq t \leq 1$.

As described by Billingsley (1968), two functions $x$ and $y$ are near one another in the uniform topology (1.1) of $C[0,1]$ if the graph of $x(t)$ differs from the graph of $y(t)$ by a uniformly (i.e., does not vary with $t$) small perturbation of the ordinates with the common abscissa held fixed but arbitrary. It will be useful in $D[0,1]$ to also allow a uniformly small change in the abscissa or time scale. Skorokhod (1956) devised a topology that allows this possibility.

Let $\Gamma$ denote the class of all strictly increasing continuous mappings of $[0,1]$ onto itself. For $x$ and $y$ in $D[0,1]$, define $d(x,y)$ to be the infimum over those positive $\varepsilon$ for which there exists a $\gamma$ in $\Gamma$ such that for $0 \leq t \leq 1$

$$\sup_{t} |\gamma(t) - t| \leq \varepsilon,$$

and

$$\sup_{t} |x(t) - y(\gamma(t))| \leq \varepsilon.$$

The function $d$ is a metric on $D[0,1]$ called the Skorokhod metric. This metric generates the Skorokhod topology on $D[0,1]$ which is termed the $J_1$ topology.

Billingsley (1968) notes that a necessary and sufficient condition for a sequence $\{f_n\}$ in $D[0,1]$ to converge to a limit $f$ in the $J_1$ topology is that there exist a sequence $\{\gamma_n\}$ in $\Gamma$ such that
\[
\lim_{n \to \infty} f_n(\gamma_n(t)) = f(t) \quad \text{uniformly in } t \quad \text{and} \quad \lim_{n \to \infty} \gamma_n(t) = t
\]

uniformly in \(t\). Also, by taking \(\gamma_n(t)=t\), it can be seen that if \(\{f_n\}\) converges uniformly to \(f\), then there is convergence in the \(J_1\) topology. Conversely, it is possible to have \(J_1\) convergence but for \(f_n \neq f\). However, convergence in the \(J_1\) topology does imply that \(f_n(t) \to f(t)\) for all points at which \(f\) is continuous.

1.2.2 Wiener Measure and Brownian Motion

Let \(W\) be a probability measure on \((C[0,1],C)\) where \(C\) is the sigma field generated by the open sets of \(C[0,1]\). Also, let \(X:[0,1] \to \mathbb{R}\) be a real valued random function in \(C[0,1]\). \(W\) is a Wiener measure if two properties hold. First, for each \(t\), the random variable \(X(t)\) is normally distributed under \(W\) with mean \(0\) and variance \(\sigma^2 t\). That is,

\[
W[X(t) \leq x] = (2\pi \sigma^2 t)^{-1/2} \int_{-\infty}^{x} \exp(-z^2/2\sigma^2 t) \, dz . \quad (1.2)
\]

For \(t = 0\), this is interpreted as \(W[X(0)=0] = 1\). Second, if

\(0 \leq t_0 \leq t_1 \leq \ldots \leq t_r \leq 1\),

then the random variables

\[
X(t_1) - X(t_0), X(t_2) - X(t_1), \ldots, X(t_r) - X(t_{r-1}) \quad (1.3)
\]
are independent under \( W \). The second property states that the stochastic process \( \{X(t); \ 0 \leq t \leq 1\} \) has independent increments under \( W \). This definition implies that the random variables given by the difference in (1.3) are all normal with zero means and variances

\[
\sigma^2(t_1-t_0), \sigma^2(t_2-t_1), \ldots, \sigma^2(t_r-t_{r-1}),
\]

respectively. The stochastic process \( \{X(t); \ 0 \leq t \leq 1\} \) is termed a **Wiener process** or a **Brownian motion**.

Another useful random function is \( X^* \) in \( C[0,1] \) given by

\[
X^*(t) = X(t) - tX(1), \quad 0 \leq t \leq 1,
\]

where \( \{X(t); \ 0 \leq t \leq 1\} \) is a Wiener process. The random function \( X^* \) is called a **Brownian bridge** (or a tied down Brownian motion). Note that \( X^*(0) = X^*(1) = 0 \) with probability 1. The distribution of \( X^* \) is also specified by the requirements

\[
E[X^*(t)] = 0 \text{ and } E[X^*(s)X^*(t)] = \sigma^2 s(1-t) \text{ if } s \leq t.
\]

### 1.2.3 Selected Properties of Martingales

Martingales form an important class of dependent random variables with a large volume of developed theory. Selected results for martingales are presented below.

First, some definitions are needed. Consider a probability space \( (\Omega, \mathcal{B}, P) \), a sequence of random variable \( \{X_n\} \) and a sequence of subsigma fields \( \{G_n\} \) contained in \( \mathcal{B} \). Further, assume that \( X_n \) is \( G_n \) measurable
and that $E|X_n| < \infty$. Then the sequence \{X_n, G_n\} is a forward martingale, or just a martingale, if

(i) \quad G_1 \subseteq G_2 \subseteq \ldots

(ii) \quad E[X_{n+1}|G_n] = X_n \text{ almost everywhere, } n \geq 1,

and a reverse martingale if

(i') \quad G_1 \supset G_2 \supset \ldots

(ii') \quad E[X_n|G_{n+1}] = X_{n+1} \text{ almost everywhere, } n \geq 1.

If the equality sign in (ii) or (ii') is replaced with a $\geq$ then \{X_n\} is termed a submartingale (either forward or reverse, respectively). Likewise, when replaced with a $\leq$, \{X_n\} is termed either a forward or a reverse supermartingale, as appropriate.

When \{X_n, G_n\} is a submartingale such that $\sup_n E|X_n| < \infty$, then there exists a random variable $X$ such that

$$X_n \rightarrow X \text{ almost surely as } n \rightarrow \infty \text{ and } E|X| < \sup_n E|X_n|.$$

To obtain additional results the added condition of uniform integrability is required. A sequence of random variables \{X_n\} is uniformly integrable if

$$\lim_{c \rightarrow \infty} E[I|X_n| I(|X_n| > c)] = 0 \text{ uniformly in } n.$$
where I is the indicator function. Then, when \( \{X_n, G_n\} \) is a (forward or reverse) submartingale for which the sequence \( \{X_n\} \) is uniformly integrable, there exists a random variable \( X \) such that

\[
X_n + X \text{ almost surely and } E|X_n - X| \to 0 \text{ as } n \to \infty. \quad (1.5)
\]

Finally, if \( \{X_n, G_n\} \) is a reverse martingale, then the \( X_n \) are uniformly integrable and (1.5) holds.

1.3 U-Statistics and Von Mises' Differentiable Statistical Functions

1.3.1 Estimators and Variances

This section introduces two related and important classes of estimators. The large sample invariance properties of these two classes are well understood for independent and identically distributed sequences of observations. The line of research which follows will extend some of these results to unequal probability without replacement sampling from a finite population.

The original work on these two estimators was done by von Mises (1947) and Hoeffding (1948). The development given below closely follows the third Chapter of Sen (1981). Additional material is drawn from Serfling (1980) and Fraser (1957).

Consider a sequence \( \{X_i; 1 \leq i\} \) of independent and identically distributed (i.i.d.) random variables each having a distribution function (d.f.) \( F \). Let \( F \) be the space of all d.f.'s belonging to some specified class. A homogeneous statistical function of degree \( m \) (21) is given by
\[ \theta = \int \cdots \int g(x_1, \ldots, x_m) dF(x_1) \cdots dF(x_m) \]

\[ = E_F[g(X_1, \ldots, X_m)] \quad \text{for all } F \in \mathcal{F}, \quad (1.6) \]

where \( \mathcal{F} = \{ F : I \theta(F) I < \infty \} \) and \( g \) is a Borel measurable function called a kernel. As usual, assume without loss of generality that \( g \) is a symmetric function of its \( m \) arguments. If there exists a symmetric kernel \( g \) of degree \( m \) for which (1.6) holds, then \( \theta \) is termed an estimatable parameter.

A natural estimator of \( \theta \) is obtained by replacing \( F \) with \( F_n \), the empirical distribution function. This estimator, known as von Mises' differentiable statistical function, was studied by von Mises in 1947 and is given by

\[ V_n = \int \cdots \int g(x_1, \ldots, x_m) dF_n(x_1) \cdots dF_n(x_m) \]

\[ = n^{-m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} g(X_{i_1}, \ldots, X_{i_m}) \quad (1.7) \]

Note that \( V_n \) is not necessarily an unbiased estimator of \( \theta \). However, for large \( n \), the bias in \( V_n \) is usually negligible.

Hoeffding (1948) introduced an unbiased estimator, called a U-statistic, of \( \theta \). Suppose that \( n \geq m \), then \( U_n \) is an unbiased estimator of \( \theta \) given by
\[ U_n = n^{-[m]} \sum_{1 \leq i_1 \neq \ldots \neq i_m \leq n} g(x_{i_1}, \ldots, x_{i_m}) \]

\[ = \left( \binom{n}{m} \right)^{-1} \sum_{1 \leq i_1 < \ldots < i_m \leq n} g(x_{i_1}, \ldots, x_{i_m}) \]  \hspace{1cm} (1.8)

where

\[ n^{-[m]} = (n^{[m]})^{-1} = (n \ldots (n-m+1))^{-1} \]

and

\[ \binom{n}{m} = n^{[m]}/m! \]

If \( m = 1 \), \( V_n = U_n \), but if \( m \geq 2 \), these two statistics are not generally the same.

Next, let \( G_n \) be the sigma field generated by the unordered collection \( \{x_1, \ldots, x_n\} \) and by \( x_{n+1}, x_{n+2}, \ldots \) for \( n \geq 1 \). It follows that for every estimatable \( \theta \), the sequence \( \{U_n, G_n; n \geq m\} \) is a reverse martingale and that \( U_n + \theta \) almost surely as \( n \to \infty \). The second proposition is a consequence of the reverse martingale convergence theorem. Sen (1981), as well as others, shows that under mild conditions

\[ |V_n - U_n| = O(n^{-1}) \quad \text{almost surely as } n \to \infty . \]  \hspace{1cm} (1.9)

Hence, it follows that

\[ V_n + \theta \quad \text{almost surely as } n \to \infty . \]

Now assume that \( \mathbb{E}g^2 < \infty \) and for \( r=0,\ldots,m \) define
\[ g_r(x_1, \ldots, x_r) = E g(x_1, \ldots, x_r, x_{r+1}, \ldots, x_m) \]

with \( g_0 = \theta \). Also define \( S_0 = 0 \) and, for \( r \geq 1 \),

\[ S_r = \text{Var}[g_r(X_1, \ldots, X_r)] \]
\[ = E g_r^2(X_1, \ldots, X_r) - \theta. \]

By Jensen's inequality it follows that

\[ 0 = S_0 \leq S_1 \leq \ldots \leq S_m = \text{Var}(g) < \infty. \]

\( \theta \) is said to be \textbf{stationary of order} \( d \) if

\[ S_d = 0 < S_{d+1}. \]

The most important case being when \( S_1 > 0 \) and \( \theta \) is said to be stationary of order 0.

Other important results follow from the fact that

\[ S_r = \text{Cov}[g(X_{i_1}, \ldots, X_{i_m}), g(X_{j_1}, \ldots, X_{j_m})] \tag{1.10} \]

whenever \( (i_1, \ldots, i_m) \) and \( (j_1, \ldots, j_m) \) share \( r \) indices in common, \( r=0, \ldots, m \). If follows that

\[ \text{Var}(U_n) = \binom{n}{m}^{-1} \sum_{r=1}^{m} \binom{m}{r} \binom{n-m}{m-r} S_r. \tag{1.11} \]
Then, if $\theta$ is stationary of order 0,

$$\text{Var}(U_n) = m^2 \xi_1/n + O(n^{-2}) \quad (1.12)$$

The variance expression for the von Mises statistics, $\text{Var}(V_n)$, is more complicated. However, under mild conditions it can be shown that

$$\text{Var}(V_n) = m^2 \xi_1/n + O(n^{-2})$$

if $\theta$ is stationary of order 0.

1.3.2 Weak Convergence of $\{U_n\}$ and $\{V_n\}$

First, consider the U-statistics case. For every $n \geq 2m$ define a stochastic process

$$Y_n = \{Y_n(t); 0 \leq t \leq 1\} \quad (1.13)$$

by letting

$$Y_n(t) = 0 \quad \text{for } 0 \leq t < (m-1)/n$$

$$Y_n(t) = Y_n(k/n) \quad \text{for } k/n \leq t < (k+1)/n, \ k = m-1, \ldots, n-1$$

and

$$Y_n(k/n) = \frac{k[U_k - \theta]}{m[n\xi_1]^{1/2}} \quad \text{for } k = m, \ldots, n$$

Miller and Sen (1972) show that if $\theta$ is stationary of order 0, then as $n \to \infty$

$$Y_n \overset{D}{\to} W \quad \text{in the } J_1 \text{ topology on } D[0,1] \quad (1.14)$$
where $W$ is a standard Brownian motion on $[0,1]$.

An analogous result holds for the von Mises statistic by defining

$$Y_n^* = \{Y_n^*(t); 0 \leq t \leq 1\}$$  \hspace{1cm} (1.15)

with

$$Y_n^*(t) = 0 \text{ for } 0 \leq t < n^{-1}$$

$$Y_n^*(t) = Y_n^*(k/n) \text{ for } k/n \leq t < (k+1)/n, \ k = 1, \ldots, n-1$$

and

$$Y_n^*(k/n) = \frac{k[V_k - \theta]}{m[n\xi_1]^{1/2}} \text{ for } k = 1, \ldots, n.$$  \hspace{1cm} (1.16)

Under a mild condition it follows that if $\theta$ is stationary of order $0$, then as $n \to \infty$

$$\rho(Y_n, Y_n^*) \to 0$$

and

$$Y_n^* \overset{D}{\to} W \text{ in the } L_1 \text{ topology on } D[0,1]$$

where $\rho$ is the uniform topology given in (1.1).

From (1.14) and (1.16) it follows that as $n \to \infty$ both

$$\frac{n^{1/2}[V_n - \theta]}{m[\xi_1]^{1/2}} \text{ and } \frac{n^{1/2}[V_n - \theta]}{m[\xi_1]^{1/2}}$$

coverage in distribution to $W(1)$, a standard normal random variable.
The above results summarize the past history of \( \{U_n\} \) and \( \{V_n\} \).

Loynes (1979) characterizes the future of these two sequences using the reverse martingale property of \( \{U_n\} \) and that, by (1.9), \( \{V_n\} \) behaves almost surely like a reverse martingale.

To see this, define the stochastic process

\[ Z_n = \{Z_n(t); \ 0 \leq t \leq 1\}, \text{ with } n \geq m, \text{ by} \]

\[ Z_n(t) = \frac{n^{1/2}[U_n(t) - \theta]}{m[\varsigma_1]^{1/2}} \quad (1.17) \]

where

\[ n(t) = \min\{k : n \leq tk\} \]

Also define the process \( Z^*_n = \{Z^*_n(t); \ 0 \leq t \leq 1\} \) by

\[ Z^*_n(t) = \frac{n^{1/2}[V_n(t) - \theta]}{m[\varsigma_1]^{1/2}} \quad (1.18) \]

where \( n(t) \) is as above. Under this framework, if \( \theta \) is stationary of order 0, then

\[ Z_n \overset{D}{\rightarrow} W \quad \text{in the } J_1 \text{ topology on } D[0,1] \quad (1.19) \]

Also, under a mild additional condition,
\[ p(Z_n, Z_n^*) \xrightarrow{p} 0 \]

and

\[ Z_n^* \xrightarrow{D} W \text{ in the } J_1 \text{ topology on } D[0,1] \quad (1.20) \]

1.3.3 **Projection Method**

Another approach for establishing the large sample distribution of \( U_n \) is the projection method first used by Hoeffding in his 1948 paper. This is the basic approach adopted in Chapter 2 for use with unequal probability sampling.

The projection of a U-statistic (see Serfling (1980), Section 5.3.1 for a good description) is

\[ \hat{U}_n = \sum_{i=1}^{n} E[U_n | X_i] - (n-1)\theta. \]

Hoeffding demonstrates that the random variables

\[ Y_n = \sqrt{n} (\hat{U}_n - \theta) \]

and

\[ Z_n = \sqrt{n} (U_n - \theta) \]

have the same limiting distribution by showing that \( Y_n \) and \( Z_n \) are equivalent in quadratic mean as \( n \) trends to infinite.

The advantage of considering \( \hat{U}_n \) is that it is a simple degree one summation rather than a degree \( m \) statistic. When the original observations are independent and identically distributed, \( \hat{U}_n \) is just
the sum of n i.i.d. random variables. Thus, the well developed theory for the mean or total of i.i.d. random variables extends to the degree m statistic $U_n$.

For the case considered in Chapter 2, the observations are dependent because of less than complete replacement sampling. While this complicates the situation, it is shown that under certain assumptions $U_n$ and $\hat{U}_n$ share the same limiting distribution. Then, if a central limit theorem has been proved for a total estimator from the design in question, it can be extended to the statistic $U_n$.

The case of equal probability sampling without replacement from a finite population has been well covered by Nandi and Sen (1963), Sen (1970) and Sen (1972). A compilation of these results is available in Sen (1981). It is shown that, for simple random sampling without replacement from a finite population, there is convergence to a Brownian bridge process.

Following Sen (1981), define $N = \{N: N \geq 1\}$, the set of all positive integers, and let $A_N = (a_{N1}, \ldots, a_{NN})$ be a sequence of real numbers. Let the random vector $X_N = (X_{N1}, \ldots, X_{NN})$ take on each permutation of $A_N$ with common probability $(N!)^{-1}$. Associate with $X_N$ the sequence of samples

$$x_{N}^{(n)} = (X_{N1}, \ldots, X_{Nn}) \text{ for } n=1, \ldots, N$$

consisting of the first $n$ elements of $X_N$.

For a symmetric kernel $g$ of degree $m$, define the parameter
\[ \theta_N = N^{-[m]} \sum_{P_{N,m}} g(a_{Ni_1}, \ldots, a_{Ni_m}) \]  

(1.21)

with

\[ N^{-[m]} = (N \ldots (N-m+1))^{-1} \]

and

\[ P_{N,m} = \{(i_1, \ldots, i_m) : 1 \leq i_1 \neq \ldots \neq i_m \leq N\} . \]

Based on the sample \( X_N^{(n)} \), a symmetric, unbiased estimator of \( \theta \) is

\[ U_{N,n} = n^{-[m]} \sum_{P_{n,m}} g(X_{Ni_1}, \ldots, X_{Ni_m}) . \]

(1.22)

The corresponding von Mises functional is

\[ V_{N,n} = n^{-m} \sum_{i_1=1}^{n} \ldots \sum_{i_m=1}^{n} g(X_{Ni_1}, \ldots, X_{Ni_m}) . \]

(1.23)

Next, define for every \( r = 0, \ldots, m \)

\[ g_r^{(N)}(X_{Ni_1}, \ldots, X_{Ni_r}) = (N-r)^{-[m-r]} \sum_{r}^* g(X_{Ni_1}, \ldots, X_{Ni_m}) . \]

(1.24)

where the summation extends over the \((N-r)^{[m-r]}\) possible choices of distinct \((i_{r+1}, \ldots, i_m)\) from the \(N-r\) units in \( A_N \) excluding those identified as \(X_{Ni_1}, \ldots, X_{Ni_r}\). By convention, take
\( g_0^{(N)} = \theta \) and \( g_m^{(N)} = g \). From this let

\[
S_{r,N} = N^{-r} \sum_{p_{N,r}} [g^{(N)}(a_{N1}^{(1)}, \ldots, a_{N1}^{(r)})]^2 - \theta_N^2 \quad r = 0, \ldots, m
\]  

(1.25)

with \( S_{0,N} = 0 \). Three basis assumptions are made concerning the sequence of populations:

\[
\inf_{N} S_{1,N} > 0 \quad , \quad \sup_{N} S_{m,N} < \infty
\]  

(1.26)

and

\[
\max_{1 \leq i \leq N} N^{-1}[g^{(N)}(a_{N1}^{(i)}) - \theta_N]^2 / S_{1,N} + 0 \quad \text{as} \quad N \to \infty
\]  

(1.27)

For a suitably defined sequence of sigma fields, \( \{U_{N,n}; n=1, \ldots, N\} \) is a reverse martingale. This suggests considering the stochastic process \( Y_N = \{Y_N(t); 0 \leq t \leq 1\}, N \geq 1 \), with \( Y_N(t) = 0 \) for \( n_t \leq m - 1 \) and

\[
Y_N(t) = \frac{n_t[U_{N,n_t} - \theta_N]}{m[N_{1,N}]^{1/2}}
\]  

(1.28)

where \( n_t \) is the largest integer contained in \( tN, 0 \leq t \leq 1 \). Then, under (1.26) and (1.27),

\[
Y_N \overset{D}{\to} W^* \quad \text{in the} \ J_1 \ \text{topology on} \ D[0,1]
\]  

(1.29)

where \( W^* \) is a standard Brownian bridge on \( [0,1] \).

For this sampling situation, the essential part of the projection is

\[
18
\]
\[ \hat{U}_{n, n} = \sum_{i=1}^{n} g_1^{(N)}(a_{N_i}) - (n-1) \theta_{N_i} \]  

This projection is used as shown before to determine the large sample distribution of \( U_{n, n} \). Also, note condition (1.27). A condition similar to this will be essential to the results for general unequal probability sampling.

1.4 Some Central Limit Theorems for Without Replacement Sampling

I now review some central limit theorems (CLTs) for without replacement sampling. The review will not be exhaustive, but will provide a little history and show the availability of CLTs for some common designs.

The earliest paper I have found is F. N. David (1938). This paper considers equal probability sampling without replacement from a "multinomial" universe. That is, the basic population consists of a fixed set of subgroups which are replicated over and over again to form the increasing sequence of populations. The asymptototic normality of linear functions from such a population is established.

Another early paper which dealt with equal probability sampling without replacement was Madow (1948). Madow used a generalization of the Wald and Wolfowitz (1944) result for permutations to prove the asymptotic normality of a linear function. Hajek (1960) also showed the large sample normality of a sample total under equal probability sampling without replacement under a condition similar to that of (1.27).

The asymptotic theory for successive sampling with unequal probabilities, and without replacement was studied by Rosen (1973).
He shows the large sample normality of a Horvitz-Thompson type estimator. Hajek (1964) developed the asymptotic theory of rejective sampling with unequal probabilities. Building upon this work, Visek (1979) shows that a total estimated under Sampford's method is asymptotically normal. Sampford's method will be used extensively in this research. Visek first norms the population so that

$$\sum_{i=1}^{N} Y_i(1 - \pi_i) = 0 \quad (1.31)$$

and

$$\sum_{i=1}^{N} Y_i^2(1 - \pi_i^{-1}) = 1 \quad (1.32)$$

where $Y_i$ is the population value for unit $i$ and $\pi_i$ is its inclusion probability. He then defines the sets, for a population, with $\epsilon > 0$

$$A_\epsilon = \{i : |Y_i| > \epsilon \pi_i\} \quad (1.33)$$

$$b(\epsilon) = \sum_{i \in A_\epsilon} Y_i^2(\pi_i^{-1} - 1) \quad (1.34)$$

and

$$e = \inf\{\epsilon : b(\epsilon) < \epsilon\}. \quad (1.35)$$

Next let

$$\hat{Y} = \sum_{i=1}^{N} (Y_i/\pi_i) I_i \quad (1.36)$$

and
\[ Y = \sum_{i=1}^{N} Y_i \]  

(1.37)

where \( I_i \) is the sample inclusion indicator. He then concludes that if \( e \rightarrow 0 \), then \( \hat{Y} - Y \) is asymptotically normal with mean zero and variance one.

Next, Fuller and Isaki (1981) demonstrate that the Horvitz-Thompson estimator of the mean converges in distribution to a normal law. This is done for Fuller's sampling method with random stratum boundaries (Fuller, 1970).

As a final example, I note Ohlsson's (1986) result for the Rao, Hartley and Cochran (1962) random group unequal probability without replacement sampling method. Ohlsson uses an application of the martingale CLT to obtain his result.

### 1.5 Overview of Results

The weak convergence of U-statistics to a Brownian motion process has been a fruitful area of research, as shown in Section 1.2.2. One advantage of such an approach is that the problem of a random sample size can be easily accommodated. The development of a weak convergence theorem is facilitated by first establishing a martingale property which simplifies the development. My original plan of research was to explore the weak convergence of U-statistics in unequal probability sampling to a Gaussian process as was done by Sen (1972) for equal probability sampling without replacement. In this vein, I first studied the potential of a proving a reverse martingale property. This quickly pointed out a basic difference in the two situations. In equal probability sampling, the data are exchangeable
in that all permutations of the data values have the same joint
distribution. This is not necessarily the case in unequal probability
sampling because each value is associated with its own probability of
being observed. This simple fact greatly complicates the situation.
As a result, the research presented here follows along the lines of
Hoeffding's 1948, paper but for unequal probability sampling. This
establishes the large sample normality of U-statistics from unequal
probability sampling, which is the first step in developing a weak
convergence theory.

In Chapter 2, I first define the basic sequences of populations
and samples I will consider. In general the sampling will be with
unequal probabilities and without replacement. However, some of the
first results will allow for some multiple selections of "large"
units. Next, I define a degree m population total as was done by
Folsom (1984). Then, using slightly different notation, define the U-
statistic estimator of the total, as was done by Folsom. The
development then proceeds to derive a components representation for
the covariance between two U-statistics of possibly different degrees.
The next major section develops the projection of a U-statistic for
unequal probability sampling with less then complete replacement. The
variance of the projection and the covariance of the projection with
the original U-statistic are displayed using the components
representation found earlier. The large sample equivalence in
quadratic mean of the two statistics is then shown.

Chapter 3 presents some specific results for Sampford's method of
sample selection. An asymptotic expression for the rth degree
inclusion probabilities is found. A numerical illustration is
presented to demonstrate the adequacy of the approximation. It is also shown that Sampford's methods satisfies the conditions of Chapter 2. It is then concluded that a U-statistic estimator under Sampford's sample selection method is asymptotically normal, using the central limit theorem proved by Visek (1979).

Finally, a numerical simulation is provided. This proceeds by comparing the estimates from six sets of 1,000 independent samples with a normal distribution.
II. LARGE SAMPLE THEORY

2.1 Introduction

This Chapter presents the main theoretical results of this dissertation. The sequence of populations and samples that will be considered is first defined. The U-statistic population parameter and estimator are defined following the work of Folsom (1984). Then, building upon Folsom's results, the projection of a U-statistic for unequal probability sampling is developed parallel to Hoeffding's (1948) work for i.i.d. observations. Finally, the U-statistic and its projection are shown to be equivalent in quadratic mean.

2.2 A Sequence of Populations and Samples

This section defines a sequence of finite populations and associated samples. Let $N_1, N_2, \ldots$ be a sequence of positive integers such that $N_1 < N_2 < \ldots$. Next define a sequence of finite populations $\{\Omega_t\}$ where

$$\Omega_t = \{u_a : a = 1, \ldots, N_t\} \quad (2.1)$$

consists of $N_t$ distinct units. Associated with $u_a$ is a vector of characteristics $(u_{Ya}, u_{Pa})$ where $u_{Ya}$ is a vector of values and, $u_{Pa} > 0$ for all $a$ with
\[ \sum_{a=1}^{N_t} t^p_a = 1. \]  \hspace{1cm} (2.2)

Next, let a sequence of samples \( \{ S_t^* \} \) of sizes \( \{ n_t \} \) be taken from the sequence of populations. The samples will be selected by a sequence of designs such that \( \{ S_t^* \} \) is composed of \( n_t \) selections from \( Q_t \) with \( 0 < n_1 < n_2 \ldots \) and

\[ S_t^* = \{ t^a_i : 1 \leq a_1, \ldots, n_t \leq N_t \} \]  \hspace{1cm} (2.3)

Notice that it is not necessarily required that the \( n_t \) selection in the sample \( S_t^* \) correspond to \( n_t \) distinct elements from \( Q_t \). The possibility of multiple selections of "large" units is allowed.

Finally, following the approach of Folsom (1984), define the randomly relabeled sample

\[ S_t = \{ t^s_1, \ldots, t^s_{n_t} \} \]  \hspace{1cm} (2.4)

where \( S_t \) consists of a random permutation of the elements of \( S_t^* \) labeled sequentially from left to right.

The population subscript \( t \) will often be dropped when working within a particular population and sample.

2.3 Sample Inclusion Indicators

This section presents the notation used to indicate which members of the population are included in the sample and the expected value of these indicators.
First, let $B = \{b_1, \ldots, b_k\}$ be a set of $k$ distinct ordered indices ($b_1 < \ldots < b_k$). The set of all $\binom{k}{m}$ unordered without replacement subsets of size $m$ taken from $B$ is denoted by

$$[B,m] = \{a : b_1 \leq a_1 < \ldots < a_m \leq b_k, a_i \in B\}$$  \hspace{1cm} (2.5)

Similarly, the set of all unordered with replacement subsets of size $m$ is given by

$$[B,m]' = \{a : b_1 \leq a_1 \leq \ldots \leq a_m \leq b_k, a_i \in B\}.$$  \hspace{1cm} (2.6)

We will be most interested in the sets $[Q,m]'$ and $[S,m]$. By definition, $[Q,m]'$ is the set of all size $m$ unordered with replacement subsets from the population, while $[S,m]$ is the set of all size $m$ unordered without replacement subsets from the relabeled sample.

By convention denote elements of the sample, $[S,m]$, by Greek letters (e.g., $a, \beta$); and elements of the population, $[Q,m]'$, by Roman letters (e.g., $a,b$). For example, $a=(a_1, \ldots, a_m) \in [Q,m]'$ will denote the set of population units $(u_{a_1}, \ldots, u_{a_m})$. Similarly, $a = (a_1, \ldots, a_m) \in [S,m]$ will denote the set of sample units $(s_{a_1}, \ldots, s_{a_m})$.

This is a slight abuse of the definition of a set. Technically, a set only contains distinct elements. In reality, the elements of $[Q,m]'$ are sequences since they can potentially contain duplicate elements. This point is well developed by Cassel, et. al (1977). In the following it will be understood that a "set" may contain duplicated elements.
In the following, we will need to count the number of distinct orderings of a set. Let \( a \) be an element of \([0,m]\)' and define \( \nu(a) \) to be the set of distinct elements in \( a \) and \( m_a(a) \) to be the number of times that \( a \) appears in \( a \). Then, the number of distinct reorderings of \( a \) is

\[
\tau(a) = \frac{m!}{\prod_{a \in \nu(a)} m_a(a)!}
\]

(2.7)

When \( a \) consists of \( m \) distinct elements, \( \tau(a) = m! \).

Following Folsom (1984), define the indicator random variable

\[
\lambda_\alpha(a) = \begin{cases} 
1 & \text{if } s_\alpha \text{ corresponds to } u_a \\
0 & \text{otherwise.}
\end{cases}
\]

(2.8)

Thus, \( \lambda_\alpha(a) \) equals one when population unit \( a \) is selected into the sample and randomly labeled \( \alpha \). Also, let \( E_{RIS} \) be the expectation operator over repeated random labelings conditional on the sample and \( E_S \) over repeated samplings of the population, Then,

\[
E[\lambda_\alpha(a)] = E_S[E_{RIS}(\lambda_\alpha(a))]
\]

\[
= E_S[n(a)/n]
\]

\[
= \pi(a)/n
\]

\[
= \phi(a)
\]

(2.9)

where \( n(a) \) is the number of times that population unit \( a \) appears in
the sample $S^*$, and $\pi(a)$ is the expected value of $n(a)$ over repeated sample selections.

To extend this notation, let $a$ be an element of $[S, m]$, an unordered degree $m$ subset of the relabeled sample, and $a$ be an element of $[\mathcal{U}, m]'$, an unordered degree $m$ subset of the population. Define the unordered degree $m$ sample inclusion indicator by

$$
\lambda_a(a) = \begin{cases} 
1 & \text{if } a \text{ corresponds to some reordering of } a \\
0 & \text{otherwise.}
\end{cases} \tag{2.10}
$$

Further, define the ordered degree $m$ sample inclusion indicator by

$$
\mu_a(a) = \begin{cases} 
1 & \text{if } a \text{ corresponds to } a \text{ in the same order} \\
0 & \text{otherwise}
\end{cases} \tag{2.11}
$$

Notice that

$$
\mu_a(a) = \prod_{i=1}^{m} \lambda_{a_i}(a_i) \tag{2.12}
$$

Also, the unordered indicator can be obtained from the ordered indicators via

$$
\lambda_a(a) = \sum_a^R \mu_a(a') \tag{2.13}
$$

where the summation is over the set of $\tau(a)$ distinct reorderings of $a$.

Turning to the expectation of the sample inclusion indicators, denote the expected ordered indicator by

$$
\eta(a) = E[\mu_a(a)] \tag{2.14}
$$
Also, define and note the relationship for the expected unordered indicator

\[ \phi(a) = E[\lambda_a(a)] \]

\[ = E[\sum_a^R \mu_a(a')] \]

\[ = \sum_a^R E[\mu_a(a')] \]

\[ = \tau(a) \eta(a) \]  

(2.15)

since the expectation of \( \mu_a(a') \) is constant over all \( \tau(a) \) reorderings of \( a \). Folsom (1984) gives a more detailed development of the expectation of the sample inclusion indicators. He defines the degree \( m \) sample selection frequency by

\[ n(a) = \sum_{a \in [S,m]} \lambda_a(a) \]  

(2.16)

which is the number of times that \( a \) is included in the sample. With \( \pi(a) = E[n(a)] \) we see that

\[ \pi(a) = \sum_{a \in [S,m]} E[\lambda_a(a)] \]

\[ = \binom{n}{m} \phi(a) \]  

(2.17)

or

\[ \phi(a) = \binom{n}{m}^{-1} \pi(a). \]  

(2.18)
2.4 Degree m Population Total and Estimator

In this section, the notation developed above is used to define a degree m population total and its U-statistic estimators as was done by Folsom (1984).

Let $G$ be a real-valued symmetric function mapping $\mathbb{R}^m$ into $\mathbb{R}$. The associated degree m population total is

$$\theta = \sum_{a \in [0,m]} G(a),$$

$$= \sum_{a_1=1}^{N} \sum_{a_2=a_1}^{N} \ldots \sum_{a_m=a_{m-1}}^{N} G(a_1, \ldots, a_m) \tag{2.19}$$

where $G(a) = G(Y_{a_1}, \ldots, Y_{a_m})$. The associated kernel defined on the sample is

$$G^*(a) = \sum_{a \in [0,m]} \lambda_{a}(a) G(a)/\phi(a) \tag{2.20}$$

with $a \in [S,m]$. Note that,

$$E[G^*(a)] = \sum_{a \in [0,m]} E[\lambda_{a}(a)] G(a)/\phi(a)$$

$$= \sum_{a} G(a)$$

$$= \theta \quad . \tag{2.21}$$

Thus, $G^*(a)$ is an unbiased estimator of $\theta$. Averaging over the $\binom{n}{m}$ possibilities yields the unbiased U-statistic estimator of $\theta$ given by
\[ U = \left( \frac{n}{m} \right)^{-1} \sum_{a \in [S,m]} G^*(a) \]

\[ = \left( \frac{n}{m} \right)^{-1} \sum_{1 \leq a_1 < \ldots < a_m \leq n} \sum_{1 \leq a_1 < \ldots < a_m \leq N} \lambda(a)G(a)/\phi(a). \tag{2.22} \]

This is the most general case. Actually, we will only include duplicate indices in the definition of \( \theta \) when the sample design permits the associated population unit to be selected multiple times. By construction, \( U \) is an unbiased estimator of \( \theta \) if \( \phi(a) \) is positive whenever \( G(a) \) is nonzero for \( a \in [\Omega,m]' \). In most of the following, the development will allow for the maximum level of replacement in the sampling. The results continue to hold if the definition of \( \theta \) is modified to be

\[ \theta = \sum_{a \in [\Omega,m]', \phi(a) > 0} G(a). \tag{2.23} \]

In addition, the convention is adopted that a kernel represented by a lower-case letter is centered so that its degree \( m \) population total is zero. This is accomplished by defining

\[ g(a) = G(a) - \phi(a)\theta. \tag{2.24} \]

With this convention, we have that
\[ g^*(a) = \sum_{a \in [0, m]} \lambda_a(a) g(a)/\phi(a) \]

\[ = \sum_{a \in [0, m]} \lambda_a(a) [G(a) - \phi(a)\theta]/\phi(a) \]

\[ = \sum_{a \in [0, m]} \lambda_a(a) G(a)/\phi(a) - \theta \sum_{a \in [0, m]} \lambda_a(a) \]

\[ = G^*(a) - \theta. \]  

(2.25)

Hence, the expected value of \( g^*(a) \) is zero. Finally, without loss of generality, it will often be assumed that \( \theta = 0 \) and the estimator defined in terms of the centered kernel by

\[ U = \binom{n}{m}^{-1} \sum_{a \in [S, m]} g^*(a). \]  

(2.26)

2.5 Components of Variance and Covariance

Similar to the discussion in Section 1.3.1, this section considers the covariance between two finite population sample U-statistics and its expansion into components. This components representation will be central to later developments.

First, let \( G \) and \( H \) be two symmetric kernels of degrees \( m_g \) and \( m_h \) (\( m_h \leq m_g \)) with population totals \( \theta_g \) and \( \theta_h \), respectively. The covariance between the two sample kernels when they share \( r \) (\( =0, \ldots, m_h \)) indices in common is denoted by

\[ \Sigma_r(gh) = \text{Cov}[G^*(a), H^*(\beta) \mid \#(a\wedge\beta) = r] \]  

(2.27)
where $a \in [S, m_g]$, $b \in [S, m_h]$ and $#(a \wedge b)$ is the number of elements in common between $a$ and $b$. The following theorem and corollary will be used extensively.

**THEOREM 2.1:**

Under the above definitions, for $r=0, \ldots, m_h$,

$$
S_r(gh) = \sum_{a \in [1, r]} \phi(a) \left[ \frac{\theta}{\theta - \theta_g} \left[ \frac{H_r(a)}{\phi(a) - \theta_h} \right] + \sum_{b \in [1, m_g - r]} \frac{\phi(b \wedge c)}{\phi(c) \phi(b)} \right]
$$

$$
\times \left[ \frac{G_r(a)}{\phi(ab) - \phi(a)} \right] \left[ \frac{H_r(a)}{\phi(ac) - \phi(a)} \right]
$$

(2.28)

where

$$
\phi(b \wedge c) = \frac{\tau(a) \tau(b) \tau(c)}{\tau(abc)} \frac{\phi(abc)}{\phi(a)} ,
$$

(2.29)

$$
G_r(a) = \sum_{d \in [1, m_g - r]} \frac{\tau(a) \tau(d)}{\tau(ad)} G(ad)
$$

(2.30)

and

$$
H_r(a) = \sum_{d \in [1, m_h - r]} \frac{\tau(a) \tau(d)}{\tau(ad)} H(ad)
$$

(2.31)
When $G=H$, the following corollary, which was proved directly by Folsom (1984), results.

**Corollary 2.1:**

With $G=H$ we have, for $r=0,1,\ldots,m$,

$$
\zeta_r = \zeta_r(G) \\
= \sum_{a \in \Omega, r} \phi(a) \left[ \frac{G_r(a)}{\phi(a)} - \theta \right]^2 \\
+ \sum_{a \in \Omega, r} \phi(a) \sum_{b \in \Omega, m-r} \sum_{c \in \Omega, m-r} \phi(b) \phi(c) \\
\cdot \left[ \frac{G(ab)}{\phi(ab)} - \frac{G_r(a)}{\phi(a)} \right] \left[ \frac{G(ac)}{\phi(ac)} - \frac{G_r(a)}{\phi(a)} \right]
$$

(2.32)

**Proof** Theorem 2.1:

First, define

$$
\lambda_\gamma = \left\{ \lambda_\gamma(c) : c \in \Omega, r \right\}
$$

(2.33)

the set of all degree $r$ sample inclusion indicators corresponding to $\gamma \in S, r$. Then,

$$
\zeta_r(gh) = \text{Cov} \left[ G^*(a), H^*(\rho) \mid \#(a \land \rho) = r \right] \\
= \text{Cov} \left\{ E \left[ G^*(a) \mid \lambda_\gamma \right], E \left[ H^*(\rho) \mid \lambda_\gamma \right] \mid \#(a \land \rho) = r \right\} \\
+ E \left\{ \text{Cov} \left[ G^*(a), H^*(\rho) \mid \lambda_\gamma, \#(a \land \rho) = r \right] \right\}
$$

(2.34)
Folsom shows that

\[
E[G^*(\alpha) | \lambda_{\gamma}] = \sum_{c \in [0, r]} \lambda_{\gamma}(c) \frac{G_r(c)}{\phi(c)}
= G^*_r(\gamma).
\] (2.35)

Likewise,

\[
E[H^*(\rho) | \lambda_{\gamma}] = \sum_{c \in [0, r]} \lambda_{\gamma}(c) \frac{H_r(c)}{\phi(c)}
= H^*_r(\gamma).
\] (2.36)

Now, using (2.35) and (2.36), the first term in (2.34) becomes

\[
\text{Cov}\left\{E[G^*(\alpha) | \lambda_{\gamma}], E[H^*(\rho) | \lambda_{\gamma}] \mid \#(\alpha \wedge \rho) = r\right\}
= \text{Cov}\left\{G^*_r(\gamma), H^*_r(\gamma)\right\}
= \text{E}\left\{\left[\left(G^*_r(\gamma) - \theta_g\right) \left[H^*_r(\gamma) - \theta_h\right]\right]\right\}
= \text{E}\left[\sum_{c \in [0, r]} \lambda_{\gamma}(c) \left(\frac{G_r(c)}{\phi(c)} - \theta_g\right) \sum_{c' \in [0, r]} \lambda_{\gamma}(c') \left(\frac{H_r(c')}{\phi(c')} - \theta_h\right)\right].
\]

Noting that \(\lambda_{\gamma}(c) \lambda_{\gamma}(c') = 0\) unless \(c = c'\), in which case it equals \(\lambda_{\gamma}(c)\), shows that
\[\text{Cov}\{E[G^*(\alpha) \mid \lambda_\gamma], \ E[H^*(\beta) \mid \lambda_\gamma] \mid \#(\alpha^* \beta^*)=r}\]

\[= E\left[ \sum_{c \in [0, r]} \lambda_\gamma(c) \left( \frac{G_r(c)}{\theta} - \frac{H_r(c)}{\phi(c)} \right) \right] \]

\[= \sum_{c \in [0, r]} \phi(c) \left( \frac{G_r(c)}{\phi(c)} - \theta \right) \left( \frac{H_r(c)}{\phi(c)} - \theta \right) \] \hspace{2cm} (2.37)

which is the first term in (2.34).

The second term in (2.34) is derived from

\[\text{Cov}[G^*(\alpha), H^*(\beta) \mid \lambda_\gamma, \#(\alpha^* \beta^*)=r]\]

\[= E\left\{ [G^*(\alpha) - E(G^*(\alpha) \mid \lambda_\gamma)][H^*(\beta) - E(H^*(\beta) \mid \lambda_\gamma)] \mid \lambda_\gamma, \#(\alpha^* \beta^*)=r \right\} \] \hspace{2cm} (2.38)

Next, Folsom shows that

\[G^*(\alpha) - E(G^*(\alpha) \mid \lambda_\gamma) = G^*(\alpha) - G_r^*(\gamma)\]

\[= \sum_{c \in [0, r]} \lambda_\gamma(c) \sum_{d \in [0, m_g-r]} \lambda_{\alpha-\gamma}^{(d)} \left[ \frac{G(cd)}{\phi(c)} - \frac{G_r(c)}{\phi(c)} \right] \] \hspace{2cm} (2.39)

Likewise,
\[ H^*(\rho) - E(H^*(\rho) \mid \lambda') = H^*(\rho) - H^*_r(\gamma) \]

\[ = \sum_{c \in [0, r]} \lambda' \sum_{d \in [0, m_g - r]} \sum_{e \in [0, m_h - r]} \lambda \mu \left( \frac{H(ce)}{\phi(ce)} - \frac{H_r(c)}{\phi(c)} \right) \cdot (2.40) \]

Thus, substituting (2.39) and (2.40) into (2.38) shows that

\[
\text{Cov}[G^*(a), H^*(\rho) \mid \lambda' \mu' \mu = r] = \sum_{c \in [0, r]} \lambda \sum_{d \in [0, m_g - r]} \sum_{e \in [0, m_h - r]} E \left[ \lambda \mu(d) \mu(e) \mid \lambda'(c) = 1 \right]
\]

\[ \times \left[ \frac{G(cd)}{\phi(cd)} - \frac{G_r(c)}{\phi(c)} \right] \left[ \frac{H(ce)}{\phi(ce)} - \frac{H_r(c)}{\phi(c)} \right] \]

\[ = \sum_{c \in [0, r]} \lambda \sum_{d \in [0, m_g - r]} \sum_{e \in [0, m_h - r]} \phi(d \mu c)
\]

\[ \times \left[ \frac{G(cd)}{\phi(cd)} - \frac{G_r(c)}{\phi(c)} \right] \left[ \frac{H(ce)}{\phi(ce)} - \frac{H_r(c)}{\phi(c)} \right] \cdot (2.41) \]

Taking the expectation of (2.41) yields
\[
E\{\text{Cov}[G^*(a), H^*(b) | \lambda, \#(a \wedge b) = r]\}
= \sum_{ce[u,r]} \phi(c) \sum_{de[u,m_g-r]} \phi(c) \sum_{ee[u,m_h-r]} \phi(c) \\
\times \left[ \frac{G(cd)}{\phi(c)} - \frac{G_r(c)}{\phi(c)} \right] \left[ \frac{H(ce)}{\phi(c)} - \frac{H_r(c)}{\phi(c)} \right] .
\tag{2.42}
\]

Finally, summing (2.37) and (2.42) completes the proof.

Using Theorem 2.1, it can be concluded that

\[
\text{Cov}(U_g, U_h) = \left( \begin{array}{c} n \\ m \end{array} \right)^{-1} \sum_{r=0}^{m} \binom{m}{r} \binom{n-m}{m-r} \mathcal{S}_r(gh) \tag{2.43}
\]

with \(U_g\) and \(U_h\) being the U-statistics based on the kernels \(G\) and \(H\), respectively. As a consequence of (2.43) letting \(G=H\) and \(m_g=m_h\), we have

\[
\text{Var}(U_g) = \left( \begin{array}{c} n \\ m \end{array} \right)^{-1} \sum_{r=0}^{m} \binom{m}{r} \binom{n-m}{m-r} \mathcal{S}_r . \tag{2.44}
\]

This variance expression was noted by Folsom (1984).

For use in later developments, note that Theorem 2.1 and Corollary 2.1 can be recast using the centered kernels. In this regard, define

\[
g_r(a) = \sum_{be[u,m_g-r]} \frac{\tau(a)\tau(b)}{\tau(ab)} g(ab) \tag{2.45}
\]

and

38
Substituting the lower-case quantities for their upper-case analogues and dropping $\theta$ in Theorem 2.1 and Corollary 2.1 leaves the results unchanged.

2.6 The Projection of a U-Statistic

As described in section 1.3.3, Hoeffding (1948) used the projection of a degree $m$ U-statistic onto the observations to obtain a degree one statistics which possessed the same limiting distribution as the original U-statistic. This section derives the projection of a U-statistic in an unequal probability sample.

In the case of observations from an infinite population, the projection is obtained by taking the conditional expectation of the U-statistic given the value of a particular observation. In finite population sampling, the equivalent action is to condition on which population element is associated with a particular sample unit. To this end, consider the set defined in (2.33) with $r=1$. Explicitly,

$$\lambda_\alpha = \{\lambda_\alpha(a): a = 1, \ldots, N\}$$

which is the set of all $N$ sample inclusion indicators for sampling unit $\alpha$. This set shows which population unit is selected into the sample and randomly labeled $\alpha$. With this definition, the projection of a finite population sample U-statistic is
\[ \hat{U} = \sum_{a=1}^{n} E[U \mid \lambda_a] - (n-1)\theta. \hspace{1cm} (2.48) \]

In deriving an expression for \( \hat{U} \), the following two functions will be used:

\[ \tilde{G}(a) = \sum_{b \in [0,m]} \frac{\tau(b)}{\tau(ab)} \frac{\phi(ab)}{\phi(b)} G(b) \]

\[ = \sum_{b \in [0,m] \setminus \{a\}} \phi(a\{b\}) G(b) \hspace{1cm} (2.49) \]

for \( a \in \Omega \) and

\[ \tilde{G}^*(a) = \sum_{a=1}^{N} \lambda_a(a) \frac{\tilde{G}(a)}{\phi(a)} \hspace{1cm} (2.50) \]

where \( G \) is a degree \( m \) symmetric function. With these definitions, the following lemma is proved.

**Lemma 2.1:**

If \( U \) is a degree \( m \) U-statistic based on the kernel \( G \), then

\[ E[U \mid \lambda_a] = \frac{m}{n} G_1^*(a) + \frac{n-m}{n} \tilde{G}^*(a) \hspace{1cm} (2.51) \]

where \( G_1^* \) is defined in (2.35) with \( r=1 \).
PROOF:

By definition of $U$,

$$E[U \mid \lambda_a] = \left( \frac{n}{m} \right)^{-1} \sum_{\rho \in [S,m]} E[G^{\ast}(\rho) \mid \lambda_a]. \quad (2.52)$$

Two cases arise:

(a) $a \in \rho$

(b) $a \notin \rho$

When $a \in \rho$, Folsom (1984) Section (2.3) demonstrates that

$$E[G^{\ast}(\rho) \mid \lambda_a] = G^{\ast}_1(a). \quad (2.53)$$

Next, under case (b) when $a \notin \rho$, we have that

$$E[G^{\ast}(\rho) \mid \lambda_a] = \sum_{b \in [0,m]} E[\lambda_\rho(b) \mid \lambda_a] \, G(b) / \phi(b). \quad (2.54)$$

In light of (2.13), we see that

$$E[\lambda_\rho(b) \mid \lambda_a] = E[\sum_{b'}^{R} \mu_\rho(b') \mid \lambda_a]$$

$$= \sum_{b'}^{R} E[\mu_\rho(b') \mid \lambda_a]$$

$$= \sum_{b'}^{R} \sum_{a=1}^{N} \lambda_a(a) \, E[\lambda_a(a) \mu_\rho(b')] / E[\lambda_a(a)]$$

$$= \sum_{a=1}^{N} \lambda_a(a) \, \tau(b) \, \eta(ab) / \phi(a)$$
\[
N = \sum_{a=1}^{\lambda_a(a)} \tau(b) \phi(ab)/\tau(ab)\phi(a)
\]

\[
N = \sum_{a=1}^{\lambda_a(a)} \phi(bla).
\]  \hspace{1cm} (2.55)

The above uses the fact that \(E[\lambda_a(a) \mu_{\rho}(b')]=\eta(ab)\) is constant for all \(\tau(b)\) distinct reorderings of \(b\).

Substituting (2.55) into (2.54) and using (2.49) and (2.50) yields

\[
E[G^*(\rho) | \lambda_a] = \sum_{a=1}^{N} \lambda_a(a) \sum_{b \in [\Omega,m]} \frac{\tau(b)}{\tau(ab)} \frac{\phi(ab)}{\phi(b)} \frac{G(b)}{\phi(a)}
\]

\[
= \sum_{a=1}^{N} \lambda_a(a) \tilde{G}(a)/\phi(a)
\]

\[
= G^*(a). \hspace{1cm} (2.56)
\]

Summarizing, we have so far seen that

\[
E[G^*(\rho) | \lambda_a] = \begin{cases} G_{1}^*(a) & a \in \beta \\ \tilde{G}^*(a) & a \notin \beta \\ \end{cases}
\]  \hspace{1cm} (2.57)

Referring back to (2.52), notice that there are \(\binom{n-1}{m-1}\) cases in the summation where \(a \in \beta\) and \(\binom{n-1}{m}\) cases with \(a \notin \beta\). Hence, it follows that
\[ E[U | \lambda_a] = \left( \begin{array}{c} n \end{array} \right)^{-1} \begin{pmatrix} \binom{n-1}{m-1} G_1^*(a) + \binom{n-1}{m} \tilde{G}^*(a) \\ \end{pmatrix} \\
= \frac{m}{n} G_1^*(a) + \frac{n-m}{n} \tilde{G}^*(a) \tag{2.58} \]

which completes the proof of Lemma 2.1.

With most of the work done in proving Lemma 2.1, the main result of this section is given in Theorem 2.2.

**THEOREM 2.2:**

If \( U \) is a degree \( m \) \( U \)-statistic based on the kernel \( G \), then the projection of \( U \) is a degree one statistic given by

\[
\hat{U} = \sum_{a=1}^{n} E[U | \lambda_a] - (n-1)\theta
= \frac{1}{n} \sum_{a=1}^{n} Z^*(a) - (n-1)\theta \tag{2.59}
\]

with sample kernel

\[
Z^*(a) = \sum_{a=1}^{N} \lambda_a(a) Z(a)/\phi(a) \tag{2.60}
\]

and

\[
Z(a) = m G_1(a) + (n-m) \tilde{G}(a). \tag{2.61}
\]

The proof follows immediately from Lemma 2.1.
For future reference, note that

\[ Z^*(a) = m G_1^*(a) + (n-m) \tilde{G}^*(a). \quad (2.62) \]

2.7 Verification of \( \hat{U} \)

Because of the importance of \( \hat{U} \) in the later developments of this research, a few checks on \( \hat{U} \) are included for verification and illustration.

2.7.1 Expected Value of \( \hat{U} \)

By construction, we expect that the \( E(\hat{U}) = \theta \). To verify this, consider

\[
E(\hat{U}) = \frac{1}{n} \sum_{a=1}^{n} E[Z^*(a)] - (n-1)\theta
\]

\[
= \frac{1}{n} \sum_{a=1}^{n} \sum_{a=1}^{N} E[\lambda_a(a)]Z(a)/\phi(a) - (n-1)\theta
\]

\[
= \frac{1}{n} \sum_{a=1}^{n} \sum_{a=1}^{N} Z(a) - (n-1)\theta
\]

\[
= \sum_{a=1}^{N} Z(a) - (n-1)\theta. \quad (2.63)
\]

Substituting (2.61) yields

\[
E(\hat{U}) = m \sum_{a=1}^{N} G_1(a) + (n-m) \sum_{a=1}^{N} \tilde{G}(a) - (n-1) \theta. \quad (2.64)
\]
From (2.30) with \( r=1 \), we see that

\[
\sum_{a=1}^{N} G_1(a) = \sum_{a=1}^{N} \sum_{b \in [\Omega, m-1]} \frac{\tau(a)\tau(b)}{\tau(ab)} G(ab)
\]

\[
= \sum_{c \in [\Omega, m]} G(c)
\]

\[
= \theta \quad \text{(2.65)}
\]

Next, from (2.49), we have that

\[
\sum_{a=1}^{N} \sim G(a) = \sum_{a=1}^{N} \sum_{b \in [\Omega, m]} \phi(a|b)G(b)
\]

\[
= \sum_{b \in [\Omega, m]} G(b) \sum_{a=1}^{N} \phi(a|b)
\]

\[
= \sum_{b \in [\Omega, m]} G(b)
\]

\[
= \theta \quad \text{(2.66)}
\]

since \( \sum_{a=1}^{N} \phi(a|b) = 1 \) (see Folsom (1984) Section 2.3). Hence, putting (2.65) and (2.66) into (2.64) yields

\[
E(\hat{U}) = m\theta + (n-m)\theta - (n-1)\theta
\]

\[
= \theta,
\]

as desired.
2.7.2 With Replacement Sampling

Under with replacement sampling the observations are independent and the results should simplify to the classical case. Thus, we expect that $\tilde{G}^*(a) = \theta$ (see Serfling (1980) Section 5.3.1).

Assuming with replacement sampling, the expected inclusion indicators factor as follows

$$\eta(ab) = \eta(a) \eta(b)$$

$$= \phi(a) \eta(b). \quad (2.67)$$

Using (2.67) with (2.49) and (2.50) yields,

$$\tilde{G}^*(a) = \sum_{a=1}^{N} \lambda_a(a) \sum_{b \in [0,m]} \frac{\tau(b)}{\tau(ab)} \frac{\phi(ab)}{\phi(a) \phi(b)} G(b)$$

$$= \sum_{a=1}^{N} \lambda_a(a) \sum_{b \in [0,m]} \frac{\eta(ab)}{\phi(a) \eta(b)} G(b)$$

$$= \sum_{a=1}^{N} \lambda_a(a) \sum_{b \in [0,m]} G(b)$$

$$= \sum_{b \in [0,m]} G(b)$$

$$= \theta \quad (2.68)$$

as desired.
2.7.3 Simple Random Sampling Without Replacement

Under simple random sampling without replacement (SRS) we have that \( \phi(a) = 0 \) if \( a \) contains repeated elements. Otherwise

\[
\phi(a) = \binom{n}{m}^{-1} \binom{N}{m} N^{-[m]}
\]

\[
= \binom{N}{m}^{-1}.
\]  

(2.69)

Also, under SRS

\[
G_1^s(a) = \sum_{a=1}^{N} \lambda_a(a) \sum_{b \in [0-a,m-1]} \frac{\tau(b) g(ab)}{\tau(ab) \phi(a)}
\]

\[
= N \sum_{a=1}^{N} \lambda_a(a) \sum_{b \in [0-a,m-1]} \frac{\tau(b)}{\tau(ab)} g(ab).
\]  

(2.70)

where \( \Omega-a \) is the set of population units excluding unit \( a \). Notice that the summation in (2.70) is over the set \([0-a,m-1]\) rather than \([0-a,m-1]\)' since repeats are not allowed in SRS. For this reason, we can conclude that

\[
\frac{\tau(b)}{\tau(ab)} = \frac{(m-1)!}{m!} = \frac{1}{m}.
\]  

(2.71)
Hence,

\[ G_1^*(a) = \frac{N}{m} \sum_{a=1}^{N} \lambda_a(a) \sum_{b \in [Q-a, m-1]} g(ab) \]

\[ = \binom{N}{m} \binom{N-1}{m-1}^{-1} \sum_{a=1}^{N} \sum_{b \in [Q-a, m-1]} g(ab) \]  

(2.72)

This checks with Sen (1981) Section 3.5 after scaling by \( \binom{N}{m} \) since Sen defines \( \theta \) as a mean rather than as a total.

2.8 Equivalence of \( U \) and \( \hat{U} \) in Quadratic Mean for Without Replacement Sampling

This section demonstrates that \( U \) and \( \hat{U} \) have the same large sample distribution under certain regularity conditions. This is done by studying \( E(U - \hat{U})^2 = \text{Var}(U - \hat{U}) \). We need to show that

\[ \frac{\text{Var}(U_t - \hat{U}_t)}{\text{Var}(U_t)} \to 0 \quad \text{as} \quad t \to \infty \]  

(2.73)

where \( U_t \) and \( \hat{U}_t \) are the statistics based on population \( t \) as defined in Section 2.2. From this we can conclude that both \( U \) and \( \hat{U} \) have the same asymptotic distribution (Hoeffding, 1948). Then, since \( \hat{U} \) is a degree one statistic, currently available central limit theorems can be used to establish the large sample distribution of \( \hat{U} \) and, hence, \( U \).

In order to determine when (2.73) holds, the components representation of the variance and covariance developed in Section 2.5 are used. First, note that
\[ \text{Var}(U - \hat{U}) = \text{Var}(U) + \text{Var}(\hat{U}) - 2\text{Cov}(U, \hat{U}) \quad (2.74) \]

Components representations for \( \text{Var}(U) \), \( \text{Var}(\hat{U}) \) and \( \text{Cov}(U, \hat{U}) \) will be used to show that \( \text{Var}(U_t - \hat{U}_t) \) is at most \( o(N_t^{2m-1}) \). Then, since \( \text{Var}(U_t) \) is \( O(N_t^{2m-1}) \), condition (2.73) follows.

To avoid many complications, this section will assume that the sample design is strictly without replacement. It is further assumed that the joint inclusion probabilities are positive for any set of \( m \) distinct population units. This implies that the degree \( m \) population total of interest is

\[ \theta = \sum_{a \in [0, \bar{a}]} G(a) = \sum_{1 \leq a_1 < \ldots < a_m \leq N} G(a_1, \ldots, a_m) \quad (2.75) \]

which does not allow duplicate population units in the degree \( m \) subsets. Likewise, most other summations will be over sets of distinct units since the joint inclusion probability is zero for a set with duplicated elements. Finally, also assume that \( \theta = 0 \) by using the centered kernels in the development to simplify the exposition.

First, consider the \( \text{Var}(\hat{U}) \). From Theorem 2.2, \( \hat{U} \) is a degree one U-statistic with kernel

\[ z(a) = m g_1(a) + (n-m) \tilde{g}(a) \quad (2.76) \]

with
\[ g_1(a) = \sum_{b \in [\Omega - a, m-1]} g(ab)/m \] 

(2.77)

and

\[ \tilde{\eta}(a) = \sum_{b \in [\Omega - a, m]} \phi(ab) g(b) \]

\[ = \sum_{b \in [\Omega - a, m]} \frac{\eta(ab)}{\eta(b)} g(b) \] 

(2.78)

where \( \Omega - a \) is the set of population elements excluding unit \( a \).

The variance of \( \hat{U} \) is derived in the following lemma.

**Lemma 2.2:**

The variance of the projected statistic \( \hat{U} \) is

\[
\text{Var}(\hat{U}) = \frac{m^2}{n} \sum_{a=1}^{N} g_1(a)^2 / \phi(a) + \frac{(n-m)^2}{n} \sum_{a=1}^{N} \tilde{\eta}(a)^2 / \phi(a) \\
+ \frac{2m(n-m)}{n} \sum_{a=1}^{N} g_1(a) \tilde{\eta}(a) / \phi(a) \\
+ \frac{n-1}{n} \sum_{a \neq b} \eta(ab) z(a)z(b) / \phi(a)\phi(b). 
\] 

(2.79)

**Proof:**

Equation (2.44) with \( m=1 \) shows that

\[
\text{Var}(\hat{U}) = \binom{n}{1}^{-1} \sum_{r=0}^{1} \binom{1}{r} \binom{n-1}{1-r} S_r(z) \\
= \frac{n-1}{n} S_0(z) + \frac{1}{n} S_1(z). 
\] 

(2.80)
Next, Corollary 2.1 implies that

\[ S_1(z) = \sum_{a=1}^{N} \phi(a) \left[ z(a)/\phi(a) \right]^2 \]

\[ = \sum_a \left[ m g_1(a) + (n-m) \tilde{g}(a) \right]^2 / \phi(a) \]

\[ = m^2 \sum_a g_1(a)^2 / \phi(a) + (n-m)^2 \sum_a \tilde{g}(a)^2 / \phi(a) \]

\[ + 2m(n-m) \sum_a g_1(a) \tilde{g}(a) / \phi(a) \]  \hspace{1cm} (2.81)

Furthermore, again using Corollary 2.1,

\[ S_0(z) = \sum_a \sum_{a \neq b} \phi(abi\cdot) z(a)z(b) / \phi(a)\phi(b) \]

\[ = \sum_a \sum_{a \neq b} \phi(ab) \frac{z(a)z(b)}{\tau(ab) \phi(a)\phi(b)} \]

\[ = \sum_a \sum_{a \neq b} \eta(ab)z(a)z(b) / \phi(a)\phi(b) \]  \hspace{1cm} (2.82)

with

\[ \phi(abi\cdot) = \frac{\tau(a)\tau(b)}{\tau(ab)} \phi(ab) \]

\[ = \phi(ab)/\tau(ab) \]  \hspace{1cm} (2.83)

The proof is completed by combining (2.81) and (2.82) with (2.80) to yield (2.79).
The next term to consider is the covariance between $U$ and $\hat{U}$. This is presented in the following lemma.

**LEMMA 2.3:**

The covariance between a U-statistic $U$ and its projection $\hat{U}$ is

\[
\text{Cov}(U, \hat{U}) = \frac{m^2}{n} \sum_{a=1}^{N} g_1(a)^2/\phi(a) \\
\quad + \frac{(n-m)^2}{n} \sum_{a=1}^{N} \tilde{g}(a)^2/\phi(a) \\
\quad + \frac{2m(n-m)}{n} \sum_{a=1}^{N} g_1(a)\tilde{g}(a)/\phi(a). \tag{2.84}
\]

**PROOF:**

Equation (2.43), with $m_g=m$ and $m_h=1$, implies that

\[
\text{Cov}(U, \hat{U}) = \binom{n}{1}^{-1} \sum_{r=0}^{m} \binom{m}{r} \binom{n-m}{1-r} C_r(gz) \\
\quad = \frac{n-m}{n} C_0(gz) + \frac{m}{n} C_1(gz). \tag{2.85}
\]

Next, from Theorem 2.1, the single-overlap component is

\[
C_1(gz) = \sum_{a=1}^{N} g_1(a)z(a)/\phi(a) \\
\quad = \sum_{a} g_1(a)[mg_1(a) + (n-m)\tilde{g}(a)]/\phi(a) \\
\quad = m \sum_{a} g_1(a)^2/\phi(a) + (n-m) \sum_{a} g_1(a)\tilde{g}(a)/\phi(a). \tag{2.86}
\]

52
Furthermore, again using Theorem 2.1, the zero-overlap component is

\[ S_0(gz) = \sum_{a=1}^{N} \sum_{b \in [0-a,m]} \phi(ab) z(a) g(b) / \phi(a) \phi(b) \]

\[ = \sum_{a} \sum_{b} \frac{\tau(a) \tau(b)}{\tau(ab)} \frac{\phi(ab)}{\phi(a) \phi(b)} z(a) g(b) \]

\[ = \sum_{a} \sum_{b} \frac{\tau(b)}{\tau(ab)} \frac{\phi(ab)}{\phi(a) \phi(b)} g(b) \left[ mg_1(a) + (n-m) g(a) \right] \]

\[ = m \sum_{a} \frac{g_1(a)}{\phi(a)} \sum_{b} \phi(ab) g(b) \]

\[ + (n-m) \sum_{a} \frac{g(a)}{\phi(a)} \sum_{b} \phi(ab) g(b) \]

\[ = m \sum_{a} g_1(a) \frac{g(a)}{\phi(a)} + (n-m) \sum_{a} g(a)^2 / \phi(a) . \quad (2.87) \]

Finally, the proof is completed by combining (2.86) and (2.87) with (2.85) to yield (2.84).

It is also interesting to note that

\[ \text{Cov}(U, \hat{U}) = \frac{1}{n} \sum_{a=1}^{N} z(a)^2 / \phi(a) . \quad (2.88) \]

Thus, the covariance is always positive.

The next term to be quantify is the variance of U. Using Corollary 2.1 and equation (2.44) we deduce the well known asymptotic variance formula for the variance of U.
\[ \text{AVar (U_t)} = tS_0 + m^2 t S_1 / n_t. \] (2.89)

Before proceeding, we need to reformulate the expression for \( S_0 \) and \( S_1 \). From equation (2.32), \( S_1 \) is

\[
S_1 = \sum_{a=1}^{N} g_1(a)^2 / \phi(a)
\]

\[
+ \sum_{a=1}^{N} \phi(a) \sum_{b \in [Q-a,m-1]} \sum_{d \in [Q-ab,m-1]} \phi(bd|a)
\]

\[
\times \left( \frac{g(ab)}{\phi(ab)} - \frac{g_1(a)}{\phi(a)} \right) \left( \frac{g(ad)}{\phi(ad)} - \frac{g_1(a)}{\phi(a)} \right).
\] (2.90)

Expanding the second term in \( S_1 \) and reordering the summations yields that

\[
S_1 = \sum_{a=1}^{N} g_1(a)^2 / \phi(a) - \sum_{a=1}^{N} g_1(a)^2 / \phi(a)
\]

\[
+ \sum_{a=1}^{N} \phi(a) \sum_{b \in [Q-a,m-1]} \sum_{d \in [Q-ab,m-1]} \phi(bd|a) \frac{g(ab)g(ad)}{\phi(ab)\phi(ad)}
\]

\[
= \sum_{a=1}^{N} \phi(a) \sum_{b \in [Q-a,m-1]} \sum_{d \in [Q-ab,m-1]} \frac{\eta(abd)}{\eta(ab)\eta(ad)} g(ab)g(ad).
\] (2.91)

Turning to \( S_0 \), equation (2.32) also yields that

54
\[ S_0 = \sum_{a \in [0,m]} \sum_{b \in [0-a,m]} \phi(ab) \frac{g(a)g(b)}{\phi(a)\phi(b)} \]

\[ = \sum_{a \in [0,m]} \sum_{b \in [0-a,m]} g(a)g(b) \frac{\eta(ab)}{\eta(a)\eta(b)} . \tag{2.92} \]

The important terms in evaluating the asymptotic equivalence of \( U \) and \( \hat{U} \) in quadratic mean have been developed. The basic conditions under which this is shown to be true are now stated. Assume that the sequence of populations and samples from Section 2.2 satisfy the following conditions:

(C1) \( 0 < q_1 \leq n_t/N_t \leq q_2 < 1 \)

(C2) \( \sup_t \max_a [N_t p_a] < \infty \)

(C3) \( \frac{\eta_t(ab)}{\eta_t(a) \eta_t(b)} = 1 + O(N_t^{-1}) \)

for \( a \in [0_t,m], b \in [0_t-a,m] \)

(C4) \( \frac{\phi_t(a) \eta_t(abd)}{\eta_t(ab) \eta_t(ad)} = 1 + O(N_t^{-1}) \)

for \( 1 \leq a \leq N_t, b \in [0_t-a,m-1], d \in [0_t-ab,m-1] \)

55
\[
\eta_t(ab) \over \dot{\phi}_t(a) \eta_t(b) = 1 + O(N_t^{-1})
1 \leq a \leq N_t, \ b \in [a+1, m]
\]

\[
\eta_t(ab) \over \dot{\phi}_t(a) \dot{\phi}_t(b) = 1 + O(N_t^{-1})
1 \leq a \neq b \leq N_t
\]

\[
\eta \max \over _{\lambda \leq a \leq N_t} \left( \frac{\gamma \left( \frac{t}{\phi}_t(a) \right)}{\phi_t(a)} - \theta_t \right)^2 \over \left( \frac{\gamma \left( \frac{t}{\phi}_t(a) \right)}{\phi_t(a)} - \theta_t \right)^2 + 0 \text{ as } t \rightarrow \infty.
\]

Condition (C1) implies that \( n_t \) and \( N_t \) grow at the same rate. The second condition requires that \( tP_a = O(N_t^{-1}) \) uniformly in \( a \). Both are commonly made conditions in such situations.

The next four conditions, (C3) - (C6), require that certain ratios of products of the expected inclusion frequencies, with the same units in both the numerator and the denominator, approach one. It is easy to show that all with replacement sampling schemes satisfy these conditions since the selections are independent and the terms in (C3) - (C6) all factor into products of their single draw probabilities. Also, simple random sampling without replacement satisfies the conditions. In Section 3.5 it will be shown that Sampford's method of unequal probability sampling without replacement satisfies (C3) - (C6).
The final condition, (C7), is sometimes called uniform asymptotic negligibility. This condition requires that the contribution of any individual unit to the variance grows slower than the \( \text{Var}(U_t) \). Since \( \text{Var}(U_t) = O(N^{2m-1}) \), we see that the numerator of (C7) must be \( o(N^{2m-1}) \) uniformly in \( a \).

The main goal of this section is to establish that (2.73) holds. This is proved in the following theorem.

**THEOREM 2.3:**

Define a sequence of finite populations and samples as shown in Section 2.2 with the restriction that the samples are drawn without replacement. Then, assuming that the sequence of populations and samples satisfy the conditions (C1) through (C7), it follows that

\[
\frac{\text{Var}(U_t - \hat{U}_t)}{\text{Var}(U_t)} \rightarrow 0 \quad \text{as} \quad t \rightarrow \infty
\]

**PROOF:**

A detailed proof for the degree 2 \((=m)\) case will be given. The modifications for the general \( m \) degree case will then be noted.

First note that \( \text{Var}(U_t) \) is of \( O(N_t^4/n_t) = O(N_t^3) \) for a degree 2 \( U \)-statistic total. Thus, it is needed to show that \( \text{Var}(U_t - \hat{U}_t) \) is of \( o(N_t^3) \). Toward this end, notice that Lemmas 2.2 and 2.3 imply

\[
\text{Var}(U - \hat{U}) = \text{Var}(U) + \text{Var}(\hat{U}) - 2 \text{Cov}(U, \hat{U})
\]

\[
= \text{Var}(U) - \text{Cov}(U, \hat{U}) + \frac{n-1}{n} \xi_0(z) \quad . \quad (2.93)
\]
Next, using (2.89) and the above, the asymptotic variance is

$$\text{AVar}(U_t - \hat{U}_t) = tS_0 + 4tS_1/n_t - \text{Cov}(U_t, \hat{U}_t) + tS_0(z). \quad (2.94)$$

It will be shown that the order of (2.94) is no greater than $o(N^3)$ and the theorem will follow.

When evaluating the order of $\text{Var}(U_t - \hat{U}_t)$, repeated use will be made of the fact that any term that is always non-positive can be ignored to obtain an upper bound for the order since $\text{Var}(U_t - \hat{U}_t) \geq 0$. That is, if $\text{Var}(U_t - \hat{U}_t) = A_t - B_t$ with $B_t \geq 0$ for all $t$, then

$$0 \leq \text{Var}(U_t - \hat{U}_t) = A_t - B_t \leq A_t.$$ 

Hence, the order of $\text{Var}(U_t - \hat{U}_t)$ is bounded above by $O(A_t)$.

Now, using equations (2.82), (2.84), (2.91) and (2.92), the $\text{AVar}(U_t - \hat{U}_t)$ for a degree two statistic is the sum of the following six terms which are labeled $A$ through $E$:

$$A = \sum_{a_1 \neq a_2 \neq b_1 \neq b_2} \frac{g(a_1, a_2)}{2} \frac{g(b_1, b_2)}{2} \frac{\eta_t(a_1, a_2, b_1, b_2)}{\eta_t(a_1, a_2) \eta_t(b_1, b_2)}$$

$$B = \frac{4}{n_t} \sum_a \frac{1}{\phi_t(a)} \sum_{b \neq a} \frac{g(a, b)}{2} \frac{\phi_t(a) \eta_t(a, b, d)}{\eta_t(a, b) \eta_t(a, d)}$$

$$C = -\frac{4}{n_t} \sum_a g_1(a)^2 / \phi_t(a)$$

58
\[ D = - \frac{(n_t-2)^2}{n_t} \sum_a \tilde{g}_t(a)^2/\phi_t(a) \]

\[ E = - \frac{4(n_t-2)}{n_t} \sum_a g_1(a) \sum_{b_1 \neq b_2} \frac{g(b_1,b_2)}{2} \frac{\eta_t(a,b_1,b_2)}{\phi_t(a)\phi_t(b)} \neq a \]

\[ F = \sum_a \sum_{a \neq b} z(a)z(b) \frac{\eta_t(a,b)}{\phi_t(a)\phi_t(b)} . \]

Each term is considered in sequence.

Using condition (C3), term A, which is \( \xi_0 \), becomes

\[ A = \sum_{a_1 \neq a_2} \sum_{b_1 \neq b_2} \sum \frac{g(a_1,a_2)}{2} \frac{g(b_1,b_2)}{2} \left[ 1 + O(N_t^{-1}) \right] . \]

This implies that term A is of no greater order than

\[ A_1 = \sum_{a_1 \neq a_2} \sum_{b_1 \neq b_2} \sum \frac{g(a_1,a_2)}{2} \frac{g(b_1,b_2)}{2} \]  \( \quad (2.95) \)

To further analyze \( A_1 \), consider the following decomposition

\[ \sum_{b_1 \neq b_2} \frac{1}{2} g(b_1,b_2) = \sum_{b_1 \neq b_2} \frac{1}{2} g(b_1,b_2) \neq a_1,a_2 \]

\[ - \sum_{b \neq a_1} \frac{1}{2} g(a_1,b) - \sum_{b \neq a_1} \frac{1}{2} g(b,a_1) \]

59
\[ - \sum_{b \neq a_2} \frac{1}{2} g(a_2, b) - \sum_{b \neq a_2} \frac{1}{2} g(b, a_2) \]
\[ + \frac{1}{2} g(a_1, a_2) + \frac{1}{2} g(a_2, a_1) \]
\[ = 0 - 2 \sum_{b \neq a_1} \frac{1}{2} g(a_1, b) - 2 \sum_{b \neq a_2} \frac{1}{2} g(a_2, b) + g(a_1, a) \]
\[ = -2 g_1(a_1) - 2 g_1(a_2) + g(a_1, a_2). \] (2.96)

The above is obtained by first including the terms containing either \( a_1 \) or \( a_2 \) and then subtracting them out. When the terms containing either \( a_1 \) or \( a_2 \) are included, we obtain \( \theta \) which is zero. Next, put equation (2.96) into (2.95) to get

\[ A_1 = \sum_{a_1 \neq a_2} \left( \sum_{a_2 \neq a_1} \frac{1}{2} g(a_1, a_2) \left[ -2g_1(a_1) - 2g_1(a_2) + g(a_1, a_2) \right] \right) \]
\[ = -2 \sum_{a_1} g_1(a_1) \sum_{a_2 \neq a_1} \frac{1}{2} g(a_1, a_2) \]
\[ - 2 \sum_{a_2} g_1(a_2) \sum_{a_1 \neq a_2} \frac{1}{2} g(a_1, a_2) \]
\[ + \sum_{a_1 \neq a_2} \frac{1}{2} g(a_1, a_2)^2 \]
\[ = -4 \sum_{a_1} g_1(a)^2 + \sum_{a_1 \neq a_2} \frac{1}{2} g(a_1, a_2). \] (2.97)

60
Now, notice that the first term above is always non-positive so that it can not dominate the order of \( \text{Var}(U_t - \hat{U}_t) \). The second term is \( O(N_t^2) \). Hence, term A can contribute at most \( O(N_t^2) \) to the order of \( \text{Var}(U_t - \hat{U}_t) \). This does not imply that term A, which is \( \varsigma_0 \), is \( O(N_t^2) \). However, when combined with the other terms in \( \text{Var}(U_t - \hat{U}_t) \), its maximum possible contribution is \( O(N_t^2) \).

Turning now to term B, which is \( \varsigma_1 \), first note that condition (C4) implies that

\[
B = \frac{4}{n_t} \sum_a \frac{1}{\phi_t(a)} \sum_{b \neq a} \frac{1}{2} g(a,b) \sum_{d \neq a,b} \frac{1}{2} g(a,d) \left[ 1 + O(N_t)^{-1} \right]
\]

This shows that B is of no greater order than

\[
B_1 = \frac{4}{n_t} \sum_a \frac{1}{\phi_t(a)} \sum_{b \neq a} \frac{1}{2} g(a,b) \sum_{d \neq a,b} \frac{1}{2} g(a,d) \quad . \quad (2.98)
\]

Considering the last summation above, note that

\[
\sum_{d \neq a,b} \frac{1}{2} g(a,d) = \sum_{d \neq a} \frac{1}{2} g(a,d) - \frac{1}{2} g(a,b) \\
= g_1(a) - \frac{1}{2} g(a,b) \quad (2.99)
\]

which is obtained by adding and subtracting \( \frac{1}{2} g(a,b) \). Now, putting (2.99) into (2.98) shows that

\[
B_1 = \frac{4}{n_t} \sum_a g_1(a)^2/\phi_t(a) - \frac{1}{n_t} \sum_a \sum_{a \neq b} g(a,b)^2/\phi_t(a) \quad . \quad (2.100)
\]
The second term above is $O(N_t^2)$ and always non-positive while the first term is $O(N_t^3)$. Hence $B$, or $\xi_1$, is approximately

$$B = \frac{4}{n_t} \sum a \frac{g_1(a)^2}{\phi_t(a)} + O(N_t^2). \quad (2.101)$$

The next term in the $\text{Var} (U_t - \hat{U}_t)$ is $C$. This term is left as is since it will cancel with the $O(N_t^3)$ part of $B$ in (2.101). Also, term $D$ requires no work since it is always non-positive. Thus, it cannot dominate the order of $\text{Var} (U_t - \hat{U}_t)$, and will be ignored.

Turning next to $E$, condition (C5) yields that

$$E = -\frac{4(n_t-2)}{n_t} \sum a \frac{g_1(a)}{\sqrt{b}_1 \neq b_2} \sum \frac{1}{2} g(b_1,b_2) \left[ 1 + O(N_t^{-1}) \right].$$

Thus, the order of $E$ is no greater than

$$E_1 = -4 \sum a \frac{g_1(a)}{\sqrt{b}_1 \neq b_2} \sum \frac{1}{2} g(b_1,b_2). \quad (2.102)$$

Now, the last two summations above are

$$\sum b_1 \neq b_2 \frac{1}{2} g(b_1,b_2) = \sum b_1 \neq b_2 \frac{1}{2} g(b_1,b_2) - \sum_{b \neq a} \frac{1}{2} [g(a,b) + g(b,a)]$$

$$= 0 - 2 g_1(a)$$

$$= -2 g_1(a). \quad (2.103)$$
This follows by adding and subtracting the terms involving $a$ and recalling that $\theta = 0$. Then, putting (2.103) into (2.102) shows that

$$E_1 = 8 \sum_a g_1(a)^2.$$  \hfill (2.104)

From these steps, it follows that

$$E = 8 \sum_a g_1(a)^2 + O(N_t^2).$$  \hfill (2.105)

This leaves term $F$ to consider. Condition (C6) implies that

$$F = \sum_a \sum_{a \neq b} z(a) z(b) \left[1 + O(N_t^{-1})\right]$$

which is of no greater order than

$$F_1 = \sum_a \sum_{a \neq b} z(a) z(b).$$  \hfill (2.106)

To further simplify $F_1$, with $\theta = 0$, the definition of $z(a)$ implies

$$\sum_a z(a) = 0.$$  \hfill (2.107)

With this

$$\sum_{b \neq a} z(b) = \sum_b z(b) - z(a)$$

$$= -z(a).$$  \hfill (2.108)

Finally, this implies that

$$F_1 = -\sum_a z(a)^2.$$  \hfill (2.109)
Since $F_1$ cannot be positive, it can not dominate the order of $\text{Var}(U_t - \hat{U}_t)$.

Summarizing the results obtained thus far, it has been shown that the order of the $\text{Var}(U_t - \hat{U}_t)$ is no greater than the order of

$$V_t = \frac{4}{n_t} \sum_a g_1(a)^2/\phi_t(a) - \frac{4}{n_t} \sum_a g_1(a)^2/\phi_t(a)$$

$$+ 8 \sum_a g_1(a)^2 + O(N_t^2)$$

$$= 8 \sum_a g(a)^2 + O(N_t^2). \quad (2.110)$$

In order to complete the proof of the theorem, it remains to be shown that

$$\sum_a t g_1(a)^2 \over \text{Var}(U_t) \rightarrow 0 \text{ as } t \rightarrow \infty. \quad (2.111)$$

To show that this is the case, note that condition (C7) can be restated as

$$(C7') \max_{1 \leq a \leq N} \frac{t g_1(a)^2/\phi_t(a)}{\text{Var}(U_t)} \rightarrow 0 \text{ as } t \rightarrow \infty$$

since $\theta_t$ is assumed to be zero. Thus,

$$\sum_a t g_1(a)^2/\text{Var}(U_t) = \sum_a \phi_t(a) x_t(a) \quad (2.112)$$
with
\[ x_t(a) = \frac{t_1(a)/\phi_t(a)}{\text{Var}(U_t)} \]  
(2.113)

With this definition, consider the two triangular arrays

\[ \{x_t(a) : 1 \leq a \leq N_t, \ t \geq 1\} \]

and

\[ \{\phi_t(a) : 1 \leq a \leq N_t, \ t \geq 1\} \]

By condition (C7')

\[ x_t(a) \to 0 \text{ as } t \to \infty. \]

Also, by definition of \( \phi_t \) and condition (C2)

\[ \sum_{a} |\phi_t(a)| = 1 \quad \text{for all } t \]

and

\[ \phi_t(a) \to 0 \text{ as } t \to \infty \text{ for } 1 \leq a \leq N_t. \]

The above is sufficient to conclude that (2.111) is true.

Hence, it has been shown that the \( \text{Var}(U_t - \hat{U}_t) \) is at most o\( (N_t^3) \) while \( \text{Var}(U_t) \) is \( O(N_t^3) \). Thus, the theorem is true for degree 2.

A different analysis of the non-positive terms from A, E and F of \( \text{Var}(U - \hat{U}) \) indicates that a condition like (C7) is essential to the
proof of Theorem 2.3. It is possible to extract terms which cancel with the first term of (2.105). However, this always introduces similar, but slightly more complicated terms, that require condition (C7) to complete the proof of the theorem via (2.111).

Now, to modify the proof of Theorem 2.3 for a general degree m statistic, first note that \( \text{Var}(U_t) \) is \( O(N_t^{2m-1}) \). Then, use the same logic as before to find that the order of \( \text{Var}(U_t - \hat{U}_t) \) is no greater than the order of the sum of

\[
A = -m^2 \sum_a g_1(a)^2 + O(N_t^{2m-2})
\]

\[
B = \frac{m^2}{n_t} \sum_a g_1(a)^2/\phi_t(a) + O(N_t^{2m-2})
\]

\[
C = -\frac{m^2}{n_t} \sum_a g_1(a)^2/\phi_t(a)
\]

\[
D = -\frac{(n_t-m)^2}{n_t} \sum_a \tilde{g}(a)^2/\phi_t(a)
\]

\[
E = 2m^2 \sum_a g_1(a)^2 + O(N_t^{2m-2})
\]

\[
F = -\sum_a z(a)^2
\]

Note that the highest order part of term \( A \) is always non-positive, which leave a contribution of at most \( O(N_t^{2m-2}) \). Next, terms \( B \) and \( C \) cancel to leave terms of \( O(N_t^{2m-2}) \). Also, terms \( D \) and \( F \) are non-
positive and cannot dominate the order of $\text{Var}(U_t - \hat{U}_t)$. Finally, term $E$ is handled via (2.111). Thus, $\text{Var}(U_t - \hat{U}_t)$ is at most $o(N_t^{2m-1})$ and the proof is complete.
III. SAMFORD'S METHOD OF UNEQUAL PROBABILITY SAMPLING WITHOUT REPLACEMENT

3.1 Introduction

This Chapter develops some results for Sampford's method of unequal probability sampling without replacement. A simple approximation for the rth degree inclusion probabilities is developed and a numerical example given. It is shown that Sampford's method satisfies the requirements of Section 2.8 so that it can be concluded that a U-statistics and its projection, $\hat{U}$, have the same limiting distribution.

3.2 Sampford's Method of Sample Selection

Sampford (1967) presented a method of selecting a sample of n distinct units from a population so that unit a of the population has probability $np_a$, assumed to be less than 1, of appearing in the sample. The set of relative size measures $\{p_a\}_{a=1}^N$ is assumed to satisfy $0 < np_a < 1$ for all $a$, and

$$\sum_{a=1}^{N} p_a = 1 .$$

Sampford actually presented three different methods of selecting a sample that yield the same probability structure. The easiest to describe of these methods is to select up to n units with replacement. The first draw being made with probabilities $\{p_a\}$ and all subsequent
ones with probabilities proportional to \( p_d/(1 - np_d) \). Any sample not containing \( n \) distinct units is rejected and the process repeated until a sample of \( n \) distinct units is obtained. This is the rejective version of Sampford's method.

Sampford describes his method by first defining

\[
\lambda_a = p_a/(1 - np_a). \tag{3.1}
\]

This use of the symbol \( \lambda \) should not be confused with the sample inclusion indicator defined in Section 2.3. This use of \( \lambda \) is adopted to ease reference to other work. Sampford then defines \( L_0=1 \) and

\[
L_m = \sum_{a \in \{0, m\}} \lambda_{a_1} \lambda_{a_2} ... \lambda_{a_m} \quad (1 \leq m \leq N). \tag{3.2}
\]

where the summation is over all possible sets of \( m \) drawn from the population (see Section 2.3).

The method consists of selecting a particular sample \( b_n \) of distinct units \( \{b_1, ..., b_n\} \) with probability

\[
P(b_n) = n \ K_n \lambda_{b_1} \lambda_{b_2} ... \lambda_{b_n} \ (1 - \sum_{i=1}^{n} p_{b_i}) \tag{3.3}
\]

where \( K_n \) is the constant multiplier (for a fixed population)

\[
K_n = (\sum_{t=1}^{n} L_{n-t}/n^t)^{-1}. \tag{3.4}
\]
The probability that a particular set of \( r \) units \( a = \{a_1, \ldots, a_r\} \) is included in the sample is obtained by summing (3.3) over all samples of size \( n \) that contain \( a \). This yields

\[
\pi(a) = K_n \lambda_{a_1} \ldots \lambda_{a_r} \phi(a)
\]  

(3.5)

with

\[
\phi(a) = n \sum_{b \in \mathcal{U} - a, n-r} \lambda_{b_1} \ldots \lambda_{b_{n-r}} \left[ 1 - (p_{a_1} + \ldots + p_{a_r}) - \sum_{j=1}^{n-r} p_{b_j} \right].
\]  

(3.6)

The summation is over all \( \binom{N-r}{n-r} \) unordered without replacement subsets of size \( n-r \) from the \( N-r \) population units excluding those in \( a \).

### 3.3 Approximate Multiple Inclusion Probabilities for Sampford's Method

While it is possible to exactly evaluate the joint inclusion probabilities for Sampford's method, it is not practical to do so for even moderately large sample sizes. A simple approximate expression for the joint inclusion probabilities was developed by Asok and Sukhatme (1976). Asok and Sukhatme assumed that \( n_t \) was small relative to \( N_t \) and that \( tp_a \) was of \( O(N_t^{-1}) \) as \( t \to \infty \). Their approximation for the joint inclusion probabilities \( \pi(a,b) \) is correct to \( O(N_t^{-4}) \).

For the situation germane to this study, \( n_t \) and \( N_t \) increase together. In this situation, still assuming that \( tp_a \) is \( O(N_t^{-1}) \), the approach of Asok and Sukhatme yields that the \( r^{th} \) degree inclusion probabilities are approximately of the form

70
\[ \pi_t(a) = n_t^{[r]} p_{a_1} \cdots p_{a_r} [1 + O(N_t^{-1})] \] (3.7)

with \( n^{[r]} = n!/(n-r)! \). This result is sufficient for this study.

To demonstrate (3.7), the approach of Asok and Sukhatme for the joint inclusion case is extended to the \( r \)th degree inclusion probabilities. While doing so, their basic assumption that \( n \) is small relative to \( N \) is followed and terms in the asymptotic expression for \( K_n \), (3.4), and \( \phi \), (3.6), are retained to \( O(N_t^{-2}) \). This actually retains more terms than needed for the task at hand, but yields a useful approximate expression for the degree-\( r \) inclusion probabilities to \( O(N_t^{-r+2}) \).

Equation (3.5) for \( \pi(a) \) consists of 3 terms. The first term, \( K_n \) is shown by Asok and Sukhatme in equation (2.13) to be approximately

\[
K_n = (n-1)! \left[ 1 - \frac{1}{2} n(n-1) \sum_{d=1}^{N} p_d^2 \\
- \frac{1}{3} n(n-1)(n+1) \sum_{d=1}^{N} p_d^3 \\
+ \frac{1}{8} n(n-1) (n^2-n+2) \left( \sum_{d=1}^{N} p_d^2 \right)^2 \right] . \quad (3.8)
\]

The second part of (3.5) is

\[
\lambda_{a_1} \cdots \lambda_{a_r} = \frac{p_{a_1}}{(1-np_{a_1})} \cdots \frac{p_{a_r}}{(1-np_{a_r})} . \quad (3.9)
\]
Since np_a < 1 for all a, expanding (3.9) in a Taylor's series, yields

\[ \lambda_1 \ldots \lambda_r = p_{a_1} \ldots p_{a_r} \{ 1 + np_{a_1} + n^2 p_{a_1}^2 + \ldots \} \ldots \{ 1 + np_{a_r} + n^2 p_{a_r}^2 + \ldots \} \]

\[ \approx p_{a_1} \ldots p_{a_r} \{ 1 + n(p_{a_1} + \ldots + p_{a_r}) \}

+ n^2(p_{a_1}^2 + \ldots + p_{a_r}^2) + n^2(\text{cross-products}) \]  \hspace{1cm} (3.10)

where the cross-products are

\[ \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} p_{a_i} p_{a_j} \]

The leading term is O(N^{-r}) multiplied by a function retaining term of O(N^{-2}).

Next, equation (2.14) of Asok and Sukhatme shows that the last term in (3.5) can be written as

\[ \phi(a) = n \binom{n-r}{n-r} E^* \left\{ \lambda b_1 \ldots \lambda b_{n-r} \left[ 1 - (p_{a_1} + \ldots + p_{a_r}) - \sum_{j=1}^{n-r} p_{b_j} \right] \right\} \]  \hspace{1cm} (3.11)

where E^* denotes the expectation taken over selecting (n-r) units from the population excluding the units in a by simple random sampling without replacement. This can further be rewritten as
\[ \Phi(a) = n \left\{ L_{n-r} - (p_{a_1} + \ldots + p_{a_r}) L_{n-r} - J_{n-r} \right\} \quad (3.12) \]

with

\[ L_{n-r} = \binom{N-r}{n-r} \mathbb{E} \left[ \lambda_{b_1} \ldots \lambda_{b_{n-r}} \right] \quad (3.13) \]

and

\[ J_{n-r} = \binom{N-r}{n-r} \mathbb{E} \left[ \lambda_{b_1} \ldots \lambda_{b_{n-r}} \sum_{j=1}^{n-r} P_{b_j} \right] \quad (3.14) \]

Using the lemma in Asok and Sukhatme (Equation (2.5) - (2.8)) it can be shown that \( L_{n-r} \) is approximately

\[ (n-r)! \ L_{n-r} = (\Sigma^* p_d)^{n-r} \]

\[ + \left[ n \binom{n-r}{1} (\Sigma^* p_d)^{n-r-1} - \binom{n-r}{2} (\Sigma^* p_d)^{n-r-2} \right] (\Sigma^* p_d^2) \]

\[ + \left[ n^2 \binom{n-r}{1} (\Sigma^* p_d)^{n-r-1} - 2n \binom{n-r}{2} (\Sigma^* p_d)^{n-r-2} \right] (\Sigma^* p_d^3) \]

\[ + 2 \binom{n-r}{3} (\Sigma^* p_d)^{n-r-3} \]
\[ + \left[ n^2 \binom{n-r}{2} \left( \Sigma^* p_d \right)^{n-r-2} - 3n \binom{n-r}{3} \left( \Sigma^* p_d \right)^{n-r-3} \right] + 3 \binom{n-r}{4} \left( \Sigma^* p_d \right)^{n-r-4} \left( \Sigma^* p_d^2 \right)^2 \]  

(3.15)

where \( \Sigma^* \) is the summation over the \( N-r \) units in the population excluding those in \( a \). Likewise, it can be shown that \( J_{n-r} \) is approximately

\[
\frac{(n-r)!}{(n-r)^r} J_{n-r} = (\Sigma^* p_d)^{n-r-1} (\Sigma^* p_d^2)
\]

\[
+ \left[ n(\Sigma^* p_d)^{n-r-1} - (n-r-1)(\Sigma^* p_d)^{n-r-2} \right] (\Sigma^* p_d^3)
\]

\[
+ \left[ n(n-r-1)(\Sigma^* p_d)^{n-r-2} - \binom{n-r-1}{2} (\Sigma^* p_d)^{n-r-3} \right] (\Sigma^* p_d^2)^2.
\]

(3.16)

The derivations of \( L_{n-r} \) and \( J_{n-r} \) are not difficult, but they are tedious.

Next, in order to further evaluate \( L_{n-r} \) and \( J_{n-r} \), note the following relationship
\[(\Sigma^* p_d)^k = (1 - \Sigma_{i=1}^r p_{a_i})^k\]

\[= \sum_{j=0}^{k} \binom{k}{j} 1^{k-j} (-1)^j \left(\Sigma_{i=1}^r p_{a_i}\right)^j\]

\[= 1 - k(p_{a_1} + \ldots + p_{a_r}) + \frac{1}{2} k(k-1)(p_{a_1} + \ldots + p_{a_r})^2 + O(N^{-3})\]  

(3.17)

This will be used with \(k = n-r, n-r-1, \ldots, n-r-4\). The evaluation of \(L_{n-r}\) and \(J_{n-r}\) also require the following three relations

\[(\Sigma^* p_d^2) = \Sigma p_d^2 - \Sigma_{i=1}^r p_{a_i}^2,\]  

(3.18)

\[\Sigma^* p_d^3 = \Sigma p_d^3 + O(N^{-3})\]  

(3.19)

and

\[(\Sigma^* p_d^2)^2 = (\Sigma p_d^2)^2 + O(N^{-3})\]  

(3.20)

The next step in approximating \(L_{n-r}\) is to substitute (3.17) through (3.20) into (3.15). After much tedious algebra, this yields that
\[(n-r)! \, L_{n-r} =
\]
\[
1 + \frac{1}{2} \, (n-r)(n+r+1)(\sum p_d^2) - (n-r)(p_{a_1} + \ldots + p_{a_r})
\]
\[
+ (n-r)(n-r-1) \text{(cross-products)} - (r+1)(n-r)(p_{a_1^2} + \ldots + p_{a_r^2})
\]
\[
- \frac{1}{2} \, (n-r)(n-r-1)(n+r+2)(\sum p_d^2)(p_{a_1} + \ldots + p_{a_r})
\]
\[
+ \frac{1}{3} \, (n-r) \left[ n^2 + rn + (r+1)(r+2) \right] (\sum p_d^3)
\]
\[
+ \frac{1}{8} \, (n-r)(n-r-1) \left[ n^2 + (2r+3)n + (r+2)(r+3) \right] (\sum p_d^2)^2.
\]
(3.21)

In a similar manner, \( J_{n-r} \) is found using (3.17) through (3.20) with (3.16) to yield, approximately,

\[(n-r)! \, J_{n-r} = (n-r)(\sum p_d^2) - (n-r)(p_{a_1^2} + \ldots + p_{a_r^2})
\]
\[
- (n-r)(n-r-1)(\sum p_d^2)(p_{a_1} + \ldots + p_{a_r})
\]
\[
+ (r+1)(n-r)(\sum p_d^3)
\]
\[
+ \frac{1}{2} \, (n-r)(n-r-1)(n+r+2)(\sum p_d^2)^2.
\]
(3.22)
The final term needed to evaluate $\phi(a)$ in (3.12) is approximately

$$(n-r)! \left( p_{a_1} + \ldots + p_{a_r} \right) L_{n-r} = \left( p_{a_1} + \ldots + p_{a_r} \right)$$

$$+ \frac{1}{2} (n-r)(n+r+1) \left( \sum_{d} p_d^2 \right) \left( p_{a_1} + \ldots + p_{a_r} \right)$$

$$- (n-r) \left( p_{a_1}^2 + \ldots + p_{a_r}^2 \right) - 2(n-r) \text{(cross-products)}. \quad (3.23)$$

It is now possible to evaluate $\phi(a)$. This is done by using (3.21), (3.22) and (3.23) with (3.12) and combining like terms to yield

$$\frac{(n-r)!}{n} \phi(a) = 1 + \frac{1}{2} (n-r)(n+r-1) \left( \sum_{d} p_d^2 \right)$$

$$- (n-r+1) \left( p_{a_1} + \ldots + p_{a_r} \right) + (n-r)(n-r+1) \text{(cross-products)}$$

$$- (r-1)(n-r) \left( p_{a_1}^2 + \ldots + p_{a_r}^2 \right)$$

$$- \frac{1}{2} (n-r) \left[ n^2 - (r-1)(r+1) \right] \left( \sum_{d} p_d^2 \right) \left( p_{a_1} + \ldots + p_{a_r} \right)$$

$$+ \frac{1}{3} (n-r) \left[ n^2 + rn + (r-1)(r+1) \right] \left( \sum_{d} p_d^3 \right)$$

$$+ \frac{1}{8} (n-r)(n-r-1) \left[ n^2 + (2r-1)n + (r-1)(r+2) \right] \left( \sum_{d} p_d^2 \right)^2.$$

$$\quad (3.24)$$
At last, \( \pi(a) \) in (3.5) can be evaluated. This is done by using (3.8), (3.10) and (3.24). Multiplying these three expressions, yields that \( \pi(a) \) is approximately

\[
\pi(a) = n[r] \prod_{a_1 \ldots a_r} \frac{1}{2} \left[ 1 - \frac{1}{2} r(r-1)(\Sigma p_2^2) + (r-1)(p_{a_1} + \ldots + p_{a_r}) \right] \\
- [n - r(r-1)] (\text{cross-products}) + r(r-1)(p_{a_1}^2 + \ldots + p_{a_r}^2) \\
+ (r-1) [n - \frac{1}{2} r(r+1)] (\Sigma p_3^2) (p_{a_1} + \ldots + p_{a_r}) \\
- \frac{1}{3} r(r-1)(r+1) (\Sigma p_3^3) \\
- \frac{1}{8} \left[ 4r(r-1)n - r(r-1)(r+1)(r+2) \right] (\Sigma p_4^2)^2 .
\]

Under the assumptions of Asok and Sukhatme, this approximation is valid to \( O(N^{-r+2}) \). It will prove most useful when analyzing surveys selected under Sampford's methods.

3.4 Numerical Illustration

To demonstrate the accuracy of the approximation, the example given by Asok and Sukhatme is expanded. They consider the data on 35 Scottish farms which appears in Table 5.1 in Sampford (1962) and is
reproduced here in Table 3.1. Using this data, Asok and Sukhatme calculate the exact and approximate joint inclusion probabilities for certain pairs of units for a sample of size n=3. They choose pairs of units spaced through the range of sizes of the population units. Table 2 in Asok and Sukhatme (1976) shows the approximation to very good for this situation.

To illustrate the approximation for higher order inclusion probabilities, a total sample size of n=5 is considered, and the exact and approximate inclusion probabilities of degree 2, 3 and 4 are found for certain sets of units. Following Asok and Sukhatme's lead, a set of units spaced through the range of sizes of the farms is used. Specifically, farms numbered 1, 5, 10, 15, 20, 25, 30 and 35 are studied. For this set of farms, the 28 possible joint inclusion probabilities, the 56 possible triple inclusion probabilities and the 70 possible quadruple inclusion probabilities are calculated.

The sets of inclusion probabilities are displayed in Tables 3.2, 3.3 and 3.4 for degree 2, 3 and 4, respectively. The first column in each table indicates which set of farms is being studied. The next two columns report the exact and approximate inclusion probabilities, respectively. The final column contains the percent relative difference of the approximation from the exact probability.

Studying these three tables, we see that the approximation performs very well. All of the joint and triple approximations differ from the exact value by far less than one percent. While the quadruple approximations differ from the exact values by at most 1.67 percent. It appears that the approximation is best for the joint inclusion probabilities and becomes steadily less reliable for the
Table 3.1  Recorded Acreage of Crops and Grass for 1947 for 35 Farms in Orkney

<table>
<thead>
<tr>
<th>Farm No.</th>
<th>Acreage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>62</td>
</tr>
<tr>
<td>8</td>
<td>65</td>
</tr>
<tr>
<td>9</td>
<td>65</td>
</tr>
<tr>
<td>10</td>
<td>68</td>
</tr>
<tr>
<td>11</td>
<td>71</td>
</tr>
<tr>
<td>12</td>
<td>74</td>
</tr>
<tr>
<td>13</td>
<td>78</td>
</tr>
<tr>
<td>14</td>
<td>90</td>
</tr>
<tr>
<td>15</td>
<td>91</td>
</tr>
<tr>
<td>16</td>
<td>92</td>
</tr>
<tr>
<td>17</td>
<td>96</td>
</tr>
<tr>
<td>18</td>
<td>110</td>
</tr>
<tr>
<td>19</td>
<td>140</td>
</tr>
<tr>
<td>20</td>
<td>140</td>
</tr>
<tr>
<td>21</td>
<td>156</td>
</tr>
<tr>
<td>22</td>
<td>156</td>
</tr>
<tr>
<td>23</td>
<td>190</td>
</tr>
<tr>
<td>24</td>
<td>198</td>
</tr>
<tr>
<td>25</td>
<td>209</td>
</tr>
<tr>
<td>26</td>
<td>240</td>
</tr>
<tr>
<td>27</td>
<td>274</td>
</tr>
<tr>
<td>28</td>
<td>300</td>
</tr>
<tr>
<td>29</td>
<td>303</td>
</tr>
<tr>
<td>30</td>
<td>311</td>
</tr>
<tr>
<td>31</td>
<td>324</td>
</tr>
<tr>
<td>32</td>
<td>330</td>
</tr>
<tr>
<td>33</td>
<td>356</td>
</tr>
<tr>
<td>34</td>
<td>410</td>
</tr>
<tr>
<td>35</td>
<td>430</td>
</tr>
</tbody>
</table>
Table 3.2 Exact and Approximate Joint Inclusion Probabilities for Sampford's Method with n=5 for the Data in Table 3.1

<table>
<thead>
<tr>
<th>Farms</th>
<th>Exact Probability</th>
<th>Approximate Probability</th>
<th>Percent Relative Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 5</td>
<td>0.0017519</td>
<td>0.0017552</td>
<td>-0.19</td>
</tr>
<tr>
<td>1 10</td>
<td>0.0019887</td>
<td>0.0019924</td>
<td>-0.19</td>
</tr>
<tr>
<td>1 15</td>
<td>0.0026738</td>
<td>0.0026785</td>
<td>-0.18</td>
</tr>
<tr>
<td>1 20</td>
<td>0.0041557</td>
<td>0.0041617</td>
<td>-0.14</td>
</tr>
<tr>
<td>1 25</td>
<td>0.0062968</td>
<td>0.0063019</td>
<td>-0.08</td>
</tr>
<tr>
<td>1 30</td>
<td>0.0095881</td>
<td>0.0095833</td>
<td>0.05</td>
</tr>
<tr>
<td>1 35</td>
<td>0.0136367</td>
<td>0.0136029</td>
<td>0.25</td>
</tr>
<tr>
<td>5 10</td>
<td>0.0023911</td>
<td>0.0023955</td>
<td>-0.18</td>
</tr>
<tr>
<td>5 15</td>
<td>0.0032148</td>
<td>0.0032204</td>
<td>-0.17</td>
</tr>
<tr>
<td>5 20</td>
<td>0.0049962</td>
<td>0.0050033</td>
<td>-0.14</td>
</tr>
<tr>
<td>5 25</td>
<td>0.0075695</td>
<td>0.0075756</td>
<td>-0.08</td>
</tr>
<tr>
<td>5 30</td>
<td>0.0115242</td>
<td>0.0115188</td>
<td>0.05</td>
</tr>
<tr>
<td>5 35</td>
<td>0.0163871</td>
<td>0.0163478</td>
<td>0.24</td>
</tr>
<tr>
<td>10 15</td>
<td>0.0036492</td>
<td>0.0036553</td>
<td>-0.17</td>
</tr>
<tr>
<td>10 20</td>
<td>0.0056709</td>
<td>0.0056788</td>
<td>-0.14</td>
</tr>
<tr>
<td>10 25</td>
<td>0.0085910</td>
<td>0.0085978</td>
<td>-0.08</td>
</tr>
<tr>
<td>10 30</td>
<td>0.0130777</td>
<td>0.0130718</td>
<td>0.05</td>
</tr>
<tr>
<td>10 35</td>
<td>0.0185930</td>
<td>0.0185497</td>
<td>0.23</td>
</tr>
<tr>
<td>15 20</td>
<td>0.0076223</td>
<td>0.0076323</td>
<td>-0.13</td>
</tr>
<tr>
<td>15 25</td>
<td>0.0115444</td>
<td>0.0115532</td>
<td>-0.08</td>
</tr>
<tr>
<td>15 30</td>
<td>0.0175668</td>
<td>0.0175598</td>
<td>0.04</td>
</tr>
<tr>
<td>15 35</td>
<td>0.0249634</td>
<td>0.0249102</td>
<td>0.21</td>
</tr>
<tr>
<td>20 25</td>
<td>0.0179213</td>
<td>0.0179327</td>
<td>-0.06</td>
</tr>
<tr>
<td>20 30</td>
<td>0.0272476</td>
<td>0.0272391</td>
<td>0.03</td>
</tr>
<tr>
<td>20 35</td>
<td>0.0386798</td>
<td>0.0386140</td>
<td>0.17</td>
</tr>
<tr>
<td>25 30</td>
<td>0.0411634</td>
<td>0.0411545</td>
<td>0.02</td>
</tr>
<tr>
<td>25 35</td>
<td>0.0583446</td>
<td>0.0582830</td>
<td>0.11</td>
</tr>
<tr>
<td>30 35</td>
<td>0.0882160</td>
<td>0.0882150</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Table 3.3 Exact and Approximate Triple Inclusion Probabilities for Sampford's Method with n=5 for the Data in Table 3.1

<table>
<thead>
<tr>
<th>Farms</th>
<th>Exact Probability</th>
<th>Approximate Probability</th>
<th>Percent Relative Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 5 10</td>
<td>0.0000584</td>
<td>0.0000587</td>
<td>-0.51</td>
</tr>
<tr>
<td>1 5 15</td>
<td>0.0000789</td>
<td>0.0000793</td>
<td>-0.51</td>
</tr>
<tr>
<td>1 5 20</td>
<td>0.0001238</td>
<td>0.0001244</td>
<td>-0.49</td>
</tr>
<tr>
<td>1 5 25</td>
<td>0.0001903</td>
<td>0.0001911</td>
<td>-0.43</td>
</tr>
<tr>
<td>1 5 30</td>
<td>0.0002967</td>
<td>0.0002973</td>
<td>-0.19</td>
</tr>
<tr>
<td>1 5 35</td>
<td>0.0004352</td>
<td>0.0004337</td>
<td>0.35</td>
</tr>
<tr>
<td>1 10 15</td>
<td>0.0000897</td>
<td>0.0000901</td>
<td>-0.51</td>
</tr>
<tr>
<td>1 10 20</td>
<td>0.0001407</td>
<td>0.0001414</td>
<td>-0.49</td>
</tr>
<tr>
<td>1 10 25</td>
<td>0.0002164</td>
<td>0.0002173</td>
<td>-0.43</td>
</tr>
<tr>
<td>1 10 30</td>
<td>0.0003373</td>
<td>0.0003379</td>
<td>-0.19</td>
</tr>
<tr>
<td>1 10 35</td>
<td>0.0004946</td>
<td>0.0004929</td>
<td>0.35</td>
</tr>
<tr>
<td>1 15 20</td>
<td>0.0001900</td>
<td>0.0001909</td>
<td>-0.49</td>
</tr>
<tr>
<td>1 15 25</td>
<td>0.0002921</td>
<td>0.0002933</td>
<td>-0.42</td>
</tr>
<tr>
<td>1 15 30</td>
<td>0.0004552</td>
<td>0.0004560</td>
<td>-0.18</td>
</tr>
<tr>
<td>1 15 35</td>
<td>0.0006672</td>
<td>0.0006649</td>
<td>0.35</td>
</tr>
<tr>
<td>1 20 25</td>
<td>0.0004580</td>
<td>0.0004598</td>
<td>-0.40</td>
</tr>
<tr>
<td>1 20 30</td>
<td>0.0007133</td>
<td>0.0007144</td>
<td>-0.15</td>
</tr>
<tr>
<td>1 20 35</td>
<td>0.0010449</td>
<td>0.0010409</td>
<td>0.38</td>
</tr>
<tr>
<td>1 25 30</td>
<td>0.0010944</td>
<td>0.0010952</td>
<td>-0.07</td>
</tr>
<tr>
<td>1 25 35</td>
<td>0.0016011</td>
<td>0.0015939</td>
<td>0.45</td>
</tr>
<tr>
<td>1 30 35</td>
<td>0.0024825</td>
<td>0.0024663</td>
<td>0.65</td>
</tr>
<tr>
<td>5 10 15</td>
<td>0.0001080</td>
<td>0.0001086</td>
<td>-0.51</td>
</tr>
<tr>
<td>5 10 20</td>
<td>0.0001695</td>
<td>0.0001704</td>
<td>-0.49</td>
</tr>
<tr>
<td>5 10 25</td>
<td>0.0002606</td>
<td>0.0002617</td>
<td>-0.43</td>
</tr>
<tr>
<td>5 10 30</td>
<td>0.0004062</td>
<td>0.0004069</td>
<td>-0.19</td>
</tr>
<tr>
<td>5 10 35</td>
<td>0.0005955</td>
<td>0.0005935</td>
<td>0.35</td>
</tr>
<tr>
<td>5 15 20</td>
<td>0.0002289</td>
<td>0.0002300</td>
<td>-0.49</td>
</tr>
<tr>
<td>5 15 25</td>
<td>0.0003518</td>
<td>0.0003533</td>
<td>-0.42</td>
</tr>
<tr>
<td>5 15 30</td>
<td>0.0005481</td>
<td>0.0005491</td>
<td>-0.18</td>
</tr>
<tr>
<td>5 15 35</td>
<td>0.0008034</td>
<td>0.0008006</td>
<td>0.35</td>
</tr>
<tr>
<td>5 20 25</td>
<td>0.0005516</td>
<td>0.0005538</td>
<td>-0.39</td>
</tr>
<tr>
<td>5 20 30</td>
<td>0.0008590</td>
<td>0.0008603</td>
<td>-0.15</td>
</tr>
<tr>
<td>5 20 35</td>
<td>0.0012580</td>
<td>0.0012533</td>
<td>0.38</td>
</tr>
<tr>
<td>5 25 30</td>
<td>0.0013178</td>
<td>0.0013187</td>
<td>-0.07</td>
</tr>
<tr>
<td>5 25 35</td>
<td>0.0019275</td>
<td>0.0019190</td>
<td>0.45</td>
</tr>
<tr>
<td>5 30 35</td>
<td>0.0029882</td>
<td>0.0029690</td>
<td>0.65</td>
</tr>
<tr>
<td>10 15 20</td>
<td>0.0002602</td>
<td>0.0002615</td>
<td>-0.49</td>
</tr>
<tr>
<td>10 15 25</td>
<td>0.0003999</td>
<td>0.0004015</td>
<td>-0.42</td>
</tr>
<tr>
<td>10 15 30</td>
<td>0.0006230</td>
<td>0.0006241</td>
<td>-0.18</td>
</tr>
<tr>
<td>10 15 35</td>
<td>0.0009130</td>
<td>0.0009098</td>
<td>0.35</td>
</tr>
<tr>
<td>10 20 25</td>
<td>0.0006270</td>
<td>0.0006295</td>
<td>-0.39</td>
</tr>
<tr>
<td>10 20 30</td>
<td>0.0009763</td>
<td>0.0009778</td>
<td>-0.15</td>
</tr>
<tr>
<td>10 20 35</td>
<td>0.0014296</td>
<td>0.0014242</td>
<td>0.38</td>
</tr>
</tbody>
</table>
Table 3.3 Exact and Approximate Triple Inclusion Probabilities for Sampford's Method with n=5 for the Data in Table 3.1 (CONTINUED)

<table>
<thead>
<tr>
<th>Farms</th>
<th>Exact Probability</th>
<th>Approximate Probability</th>
<th>Percent Relative Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 25 30</td>
<td>0.0014976</td>
<td>0.0014986</td>
<td>-0.07</td>
</tr>
<tr>
<td>10 25 35</td>
<td>0.0021903</td>
<td>0.0021806</td>
<td>0.44</td>
</tr>
<tr>
<td>10 30 35</td>
<td>0.0033952</td>
<td>0.0033735</td>
<td>0.64</td>
</tr>
<tr>
<td>15 20 25</td>
<td>0.0008463</td>
<td>0.0008495</td>
<td>-0.39</td>
</tr>
<tr>
<td>15 20 30</td>
<td>0.0013173</td>
<td>0.0013192</td>
<td>-0.15</td>
</tr>
<tr>
<td>15 20 35</td>
<td>0.0019282</td>
<td>0.0019209</td>
<td>0.38</td>
</tr>
<tr>
<td>15 25 30</td>
<td>0.0020203</td>
<td>0.0020216</td>
<td>-0.07</td>
</tr>
<tr>
<td>15 25 35</td>
<td>0.0029535</td>
<td>0.0029406</td>
<td>0.44</td>
</tr>
<tr>
<td>15 30 35</td>
<td>0.0045768</td>
<td>0.0045482</td>
<td>0.62</td>
</tr>
<tr>
<td>20 25 30</td>
<td>0.0031632</td>
<td>0.0031649</td>
<td>-0.05</td>
</tr>
<tr>
<td>20 25 35</td>
<td>0.0046204</td>
<td>0.0046002</td>
<td>0.44</td>
</tr>
<tr>
<td>20 30 35</td>
<td>0.0071546</td>
<td>0.0071115</td>
<td>0.60</td>
</tr>
<tr>
<td>25 30 35</td>
<td>0.0109350</td>
<td>0.0108697</td>
<td>0.60</td>
</tr>
<tr>
<td>Farms</td>
<td>Exact Probability</td>
<td>Approximate Probability</td>
<td>Percent Relative Difference</td>
</tr>
<tr>
<td>-------</td>
<td>------------------</td>
<td>-------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>1 5 10 15</td>
<td>0.0000017</td>
<td>0.0000017</td>
<td>-0.80</td>
</tr>
<tr>
<td>1 5 10 20</td>
<td>0.0000027</td>
<td>0.0000027</td>
<td>-0.89</td>
</tr>
<tr>
<td>1 5 10 25</td>
<td>0.0000042</td>
<td>0.0000043</td>
<td>-0.94</td>
</tr>
<tr>
<td>1 5 10 30</td>
<td>0.0000067</td>
<td>0.0000068</td>
<td>-0.72</td>
</tr>
<tr>
<td>1 5 10 35</td>
<td>0.0000102</td>
<td>0.0000102</td>
<td>0.22</td>
</tr>
<tr>
<td>1 5 15 20</td>
<td>0.0000037</td>
<td>0.0000037</td>
<td>-0.93</td>
</tr>
<tr>
<td>1 5 15 25</td>
<td>0.0000057</td>
<td>0.0000058</td>
<td>-0.97</td>
</tr>
<tr>
<td>1 5 15 30</td>
<td>0.0000091</td>
<td>0.0000092</td>
<td>-0.73</td>
</tr>
<tr>
<td>1 5 15 35</td>
<td>0.0000139</td>
<td>0.0000138</td>
<td>0.23</td>
</tr>
<tr>
<td>1 5 20 25</td>
<td>0.0000091</td>
<td>0.0000092</td>
<td>-1.01</td>
</tr>
<tr>
<td>1 5 20 30</td>
<td>0.0000145</td>
<td>0.0000146</td>
<td>-0.73</td>
</tr>
<tr>
<td>1 5 20 35</td>
<td>0.0000219</td>
<td>0.0000219</td>
<td>0.29</td>
</tr>
<tr>
<td>1 5 25 30</td>
<td>0.0000225</td>
<td>0.0000227</td>
<td>-0.62</td>
</tr>
<tr>
<td>1 5 25 35</td>
<td>0.0000341</td>
<td>0.0000339</td>
<td>0.48</td>
</tr>
<tr>
<td>1 5 30 35</td>
<td>0.0000543</td>
<td>0.0000537</td>
<td>1.07</td>
</tr>
<tr>
<td>1 10 15 20</td>
<td>0.0000042</td>
<td>0.0000042</td>
<td>-0.95</td>
</tr>
<tr>
<td>1 10 15 25</td>
<td>0.0000065</td>
<td>0.0000066</td>
<td>-0.98</td>
</tr>
<tr>
<td>1 10 15 30</td>
<td>0.0000104</td>
<td>0.0000105</td>
<td>-0.74</td>
</tr>
<tr>
<td>1 10 15 35</td>
<td>0.0000158</td>
<td>0.0000157</td>
<td>0.23</td>
</tr>
<tr>
<td>1 10 20 25</td>
<td>0.0000103</td>
<td>0.0000104</td>
<td>-1.02</td>
</tr>
<tr>
<td>1 10 20 30</td>
<td>0.0000165</td>
<td>0.0000166</td>
<td>-0.73</td>
</tr>
<tr>
<td>1 10 20 35</td>
<td>0.0000249</td>
<td>0.0000249</td>
<td>0.30</td>
</tr>
<tr>
<td>1 10 25 30</td>
<td>0.0000256</td>
<td>0.0000258</td>
<td>-0.62</td>
</tr>
<tr>
<td>1 10 25 35</td>
<td>0.0000388</td>
<td>0.0000386</td>
<td>0.49</td>
</tr>
<tr>
<td>1 10 30 35</td>
<td>0.0000618</td>
<td>0.0000611</td>
<td>1.08</td>
</tr>
<tr>
<td>1 15 20 25</td>
<td>0.0000140</td>
<td>0.0000141</td>
<td>-1.04</td>
</tr>
<tr>
<td>1 15 20 30</td>
<td>0.0000223</td>
<td>0.0000225</td>
<td>-0.73</td>
</tr>
<tr>
<td>1 15 20 35</td>
<td>0.0000338</td>
<td>0.0000337</td>
<td>0.32</td>
</tr>
<tr>
<td>1 15 25 30</td>
<td>0.0000347</td>
<td>0.0000350</td>
<td>-0.62</td>
</tr>
<tr>
<td>1 15 25 35</td>
<td>0.0000526</td>
<td>0.0000523</td>
<td>0.51</td>
</tr>
<tr>
<td>1 15 30 35</td>
<td>0.0000837</td>
<td>0.0000827</td>
<td>1.12</td>
</tr>
<tr>
<td>1 20 25 30</td>
<td>0.0000550</td>
<td>0.0000553</td>
<td>-0.57</td>
</tr>
<tr>
<td>1 20 25 35</td>
<td>0.0000831</td>
<td>0.0000826</td>
<td>0.60</td>
</tr>
<tr>
<td>1 20 30 35</td>
<td>0.0001322</td>
<td>0.0001306</td>
<td>1.23</td>
</tr>
<tr>
<td>1 25 30 35</td>
<td>0.0002054</td>
<td>0.0002023</td>
<td>1.49</td>
</tr>
<tr>
<td>5 10 15 20</td>
<td>0.0000050</td>
<td>0.0000051</td>
<td>-0.96</td>
</tr>
<tr>
<td>5 10 15 25</td>
<td>0.0000079</td>
<td>0.0000079</td>
<td>-0.99</td>
</tr>
<tr>
<td>5 10 15 30</td>
<td>0.0000125</td>
<td>0.0000126</td>
<td>-0.74</td>
</tr>
<tr>
<td>5 10 15 35</td>
<td>0.0000190</td>
<td>0.0000190</td>
<td>0.24</td>
</tr>
<tr>
<td>5 10 20 25</td>
<td>0.0000124</td>
<td>0.0000126</td>
<td>-1.03</td>
</tr>
<tr>
<td>5 10 20 30</td>
<td>0.0000199</td>
<td>0.0000200</td>
<td>-0.73</td>
</tr>
<tr>
<td>Farms</td>
<td>Exact Probability</td>
<td>Approximate Probability</td>
<td>Percent Relative Difference</td>
</tr>
<tr>
<td>-------</td>
<td>------------------</td>
<td>------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>5 10 20 35</td>
<td>0.0000301</td>
<td>0.0000300</td>
<td>0.30</td>
</tr>
<tr>
<td>5 10 25 30</td>
<td>0.0000309</td>
<td>0.0000311</td>
<td>-0.62</td>
</tr>
<tr>
<td>5 10 25 35</td>
<td>0.0000468</td>
<td>0.0000466</td>
<td>0.49</td>
</tr>
<tr>
<td>5 10 30 35</td>
<td>0.0000745</td>
<td>0.0000737</td>
<td>1.09</td>
</tr>
<tr>
<td>5 15 20 25</td>
<td>0.0000169</td>
<td>0.0000170</td>
<td>-1.05</td>
</tr>
<tr>
<td>5 15 20 30</td>
<td>0.0000269</td>
<td>0.0000271</td>
<td>-0.74</td>
</tr>
<tr>
<td>5 15 20 35</td>
<td>0.0000408</td>
<td>0.0000406</td>
<td>0.32</td>
</tr>
<tr>
<td>5 15 25 30</td>
<td>0.0000419</td>
<td>0.0000422</td>
<td>-0.62</td>
</tr>
<tr>
<td>5 15 25 35</td>
<td>0.0000634</td>
<td>0.0000631</td>
<td>0.52</td>
</tr>
<tr>
<td>5 15 30 35</td>
<td>0.0001099</td>
<td>0.0000998</td>
<td>1.13</td>
</tr>
<tr>
<td>5 20 25 30</td>
<td>0.0000663</td>
<td>0.0000667</td>
<td>-0.57</td>
</tr>
<tr>
<td>5 20 25 35</td>
<td>0.0001003</td>
<td>0.0000997</td>
<td>0.61</td>
</tr>
<tr>
<td>5 20 30 35</td>
<td>0.0001594</td>
<td>0.0001575</td>
<td>1.24</td>
</tr>
<tr>
<td>5 25 30 35</td>
<td>0.0002477</td>
<td>0.0002440</td>
<td>1.50</td>
</tr>
<tr>
<td>10 15 20 25</td>
<td>0.0000192</td>
<td>0.0000194</td>
<td>-1.06</td>
</tr>
<tr>
<td>10 15 20 30</td>
<td>0.0000306</td>
<td>0.0000309</td>
<td>-0.74</td>
</tr>
<tr>
<td>10 15 20 35</td>
<td>0.0000464</td>
<td>0.0000463</td>
<td>0.33</td>
</tr>
<tr>
<td>10 15 25 30</td>
<td>0.0000477</td>
<td>0.0000480</td>
<td>-0.62</td>
</tr>
<tr>
<td>10 15 25 35</td>
<td>0.0000722</td>
<td>0.0000718</td>
<td>0.53</td>
</tr>
<tr>
<td>10 15 30 35</td>
<td>0.0001148</td>
<td>0.0001135</td>
<td>1.14</td>
</tr>
<tr>
<td>10 20 25 30</td>
<td>0.0000754</td>
<td>0.0000759</td>
<td>-0.57</td>
</tr>
<tr>
<td>10 20 25 35</td>
<td>0.0001141</td>
<td>0.0001134</td>
<td>0.61</td>
</tr>
<tr>
<td>10 20 30 35</td>
<td>0.0001814</td>
<td>0.0001792</td>
<td>1.25</td>
</tr>
<tr>
<td>10 25 30 35</td>
<td>0.0002819</td>
<td>0.0002776</td>
<td>1.51</td>
</tr>
<tr>
<td>15 20 25 30</td>
<td>0.0001022</td>
<td>0.0001028</td>
<td>-0.56</td>
</tr>
<tr>
<td>15 20 25 35</td>
<td>0.0001546</td>
<td>0.0001536</td>
<td>0.64</td>
</tr>
<tr>
<td>15 20 30 35</td>
<td>0.0002457</td>
<td>0.0002426</td>
<td>1.28</td>
</tr>
<tr>
<td>15 25 30 35</td>
<td>0.0003817</td>
<td>0.0003757</td>
<td>1.55</td>
</tr>
<tr>
<td>20 25 30 35</td>
<td>0.0006027</td>
<td>0.0005926</td>
<td>1.67</td>
</tr>
</tbody>
</table>
degree 3 and 4 inclusion probabilities, however, it is still very good for all three cases. This point is further illustrated in Table 3.5, which presents the mean absolute percent relative difference between the exact and approximate inclusion probabilities for the cases considered. The means are all less than one percent, but display an increasing trend from degree 2, to 3, to 4.

3.5 Large Sample Normality for Sampford's Method

Using the approximation for $\pi(a)$ in (3.25), it can be shown that Sampford's method satisfies conditions (C3) through (C6) of Section 2.8. With this shown, Theorem 2.3 applies and we conclude that $U$ and $\hat{U}$ are equivalent in quadratic mean.

Inspection of the approximation in (3.25) shows that as $n_t$ increases with $N_t$, $\pi_t(a)$ has the form shown in (3.7). Next, note that by definition (see Section 2.3) when the $r$ elements of $a = (a_1, \ldots, a_r)$ are all distinct,

$$\eta(a) = \phi(a)/r!$$

$$= \frac{1}{r} \pi(a)/r!$$

$$= n^{-[r]} \pi(a).$$

(3.26)

Thus, combining (3.7) and (3.26) yields

$$\eta_t(a) = t^{p_{a_1}} \cdots t^{p_{a_r}} \left[1 + o(N^{-1}_t)\right].$$

(3.27)
Table 3.5 Mean Absolute Percent Relative Difference Between the Exact and Approximate Inclusion Probabilities with n=5 for the Cases in Tables 3.2, 3.3 and 3.4

<table>
<thead>
<tr>
<th>Degree of Inclusion</th>
<th>Number of Cases</th>
<th>Mean Absolute Percent Relative Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>28</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td>56</td>
<td>0.37</td>
</tr>
<tr>
<td>4</td>
<td>70</td>
<td>0.79</td>
</tr>
</tbody>
</table>
Now when a and b are as given under condition (C3) in Section 2.8, we see that

\[
\frac{\eta_t(ab)}{\eta_t(a) \eta_t(b)} = \frac{t^{P_1} \cdots t^{P_m} t^{P_{b_1}} \cdots t^{P_{b_m}} [1 + O(N^{-1})]}{t^{P_1} \cdots t^{P_m} [1 + O(N^{-1})] t^{P_{b_1}} \cdots t^{P_{b_m}} [1 + O(N^{-1})]}
\]

\[
= \frac{1 + O(N^{-1})}{[1 + O(N^{-1})][1 + O(N^{-1})]}
\]

\[
= \frac{1 + O(N^{-1})}{1 + O(N^{-1})}
\]

\[
= 1 + O(N^{-1})
\]

(3.28)

Hence, condition (C3) holds.

Next, when a, b and d are as given under condition (C4), we have that

\[
\frac{\phi_t(a) \eta_t(abd)}{\eta_t(ab) \eta_t(ad)} = \frac{t^{P_a} t^{P_{b_1}} \cdots t^{P_{b_{m-1}}} t^{P_{d_1}} \cdots t^{P_{d_{m-1}}} [1 + O(N^{-1})]}{t^{P_1} t^{P_{b_1}} \cdots t^{P_{b_{m-1}}} [1 + O(N^{-1})] t^{P_a} t^{P_{d_1}} \cdots t^{P_{d_{m-1}}} [1 + O(N^{-1})]}
\]

\[
= \frac{1 + O(N^{-1})}{[1 + O(N^{-1})][1 + O(N^{-1})]}
\]

\[
= 1 + O(N^{-1})
\]

(3.29)
Thus, condition (C4) is satisfied.

Likewise, when \( a \) and \( b \) are as given under condition (C5), we see that

\[
\frac{\eta_t(ab)}{\phi_t(a) \eta_t(b)} = \frac{t^a t_b^1 \cdots t_b^{m-1} [1 + O(N_t^{-1})]}{t^a t_b^1 \cdots t_b^{m-1} [1 + O(N_t^{-1})]}
\]

\[
= 1 + O(N_t^{-1}). \tag{3.30}
\]

This shows that (C5) holds for Sampford's method.

Finally, condition (C6) is seen to be true from (3.28) with \( m = 1 \).

Thus, Sampford's method satisfies (C3) through (C6) and \( U \) and \( \hat{U} \) are equivalent in quadratic mean. With this shown, we can now appeal to Visek's (1979) central limit theorem for Sampford's method (see Section 1.4) to infer the asymptotic normality of \( U \).
IV. NUMERICAL SIMULATION

The results developed in Chapters II demonstrates that the distribution of a U-statistic estimated from an unequal probability sample converges to a normal law as the sequence of samples and populations become infinitely large. In practice, we are allowed only one such population and sample. The advisability of using the normal distribution as a basis for inference in this situation is assessed in the numerical simulation presented in this Chapter. The simulation proceeds by selecting 1,000 independent samples using Sampford's method from a population of U.S. counties. Two different U-statistics are estimated from each sample and the empirical distribution of the 1,000 estimates for each statistic is compared with a normal distribution.

The two U-statistics used in this simulation -- Kendall's rank correlation and the with replacement variance component -- differ in that one has a narrow constricted range, while the other may take on any finite nonnegative value. The rank correlation is more likely to benefit from nearly equal probabilities of selection, while the variance component is more suited to unequal selection probabilities proportional to the size of the observation. Three sets of size measures, used to generate the selection probabilities, are considered. The first set covers a very wide range and is related to
the size of the observations. The other two sets of size measures cover progressively smaller ranges.

4.1 The Data

The data for this simulation are taken from the 1986 Area Resource File (ARF) obtained from the Health Resources and Services Administration of the U.S. Public Health Service (1987). The ARF contains one record for each of the 3,080 counties in the U.S., a subset of which will serve as the study population, with the counties being the sampling units. The 1980 U.S. Census population of each county, the 0.57 root of the population and the natural logarithm of the population are used as the size measures for selecting the samples. The 1984 count of the number of short-term general hospitals and short-term general hospital beds in the county are the analysis variables. The study population was restricted to the 2,000 smallest counties, as measured by their 1980 population count, to eliminate the possibility of large self-representing (selection probability greater than one) units and to control the cost of selecting the samples.

Table 4.1 presents some basic descriptive statistics on the study population for the three size measures and two analysis variables. The table contains the mean, standard deviation, the minimum, the first quartile, median, the third quartile, the maximum and, for the three size measures, the ratio of the maximum to minimum values. Even though only the 2,000 smallest counties are included, we see that they are still very diverse in total population size. They range in size from 91 persons to 35,376 with a ratio of largest to smallest of 389. The other two size measures are used to generate samples from a less diverse set of probabilities. The 0.57 root transformation was chosen
Table 4.1 Descriptive Statistics on the Study Population of 2,000 U.S. Counties

<table>
<thead>
<tr>
<th></th>
<th>Total Population</th>
<th>Root Population</th>
<th>Log Population</th>
<th>Hospitals</th>
<th>Hospital Beds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>14,839</td>
<td>226.8</td>
<td>9.36</td>
<td>0.95</td>
<td>62.2</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>8,885</td>
<td>84.6</td>
<td>0.79</td>
<td>0.70</td>
<td>64.2</td>
</tr>
<tr>
<td>Minimum</td>
<td>91</td>
<td>13.1</td>
<td>4.51</td>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>First Quartile</td>
<td>7,721</td>
<td>164.4</td>
<td>8.95</td>
<td>1.00</td>
<td>19.0</td>
</tr>
<tr>
<td>Median</td>
<td>13,403</td>
<td>225.2</td>
<td>9.50</td>
<td>1.00</td>
<td>50.0</td>
</tr>
<tr>
<td>Third Quartile</td>
<td>21,164</td>
<td>292.1</td>
<td>9.96</td>
<td>1.00</td>
<td>88.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>35,376</td>
<td>391.5</td>
<td>10.47</td>
<td>4.00</td>
<td>1,006</td>
</tr>
<tr>
<td>Ratio of Maximum to Minimum</td>
<td>389</td>
<td>29.9</td>
<td>2.32</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
to provided a set of size measures that had an approximate 30 to 1 range of sizes, while the log transformation provides a set of sizes measures with a less than 2.5 to 1 range. These two transformations of the size measures were mainly chosen to restrict the variability of the selection probabilities and not to determine an optimum size measure.

Further information about the relationships among these variables is given by the Pearson correlation matrix in Table 4.2. Notice that all three of the size measures are well correlated with the analysis variables. The usual intuition for degree one statistics, see Cochran (1977) Section 9A.4, is that if the size measure is proportional to the analysis variable, the variability of the estimate will be reduced. Thus, the size measures in this study would be considered potentially "good" in current survey practice.

4.2 Design of the Simulation

Six sets of 1,000 independently replicated samples were drawn from the 2,000 study counties using the rejective version of Sampford's method described in Section 3.2. The six sets result from using each of the three size measures with sample sizes of 50 and 100. I attempted to consider sample sizes greater than 100, but, as the sample size increases, the rejective version of Sampford's method almost always rejects a candidate with replacement sample because of at least one duplicated unit in the sample.

For each of the 6,000 samples, two U-statistics were estimated. The first U-statistic considered was Kendall's (1938) rank correlation, often called τa. A general description of this statistic is given in Kendall (1970) or in Conover (1980). To formulate this as
<table>
<thead>
<tr>
<th></th>
<th>Total Population</th>
<th>Root Population</th>
<th>Log Population</th>
<th>Hospitals</th>
<th>Hospital Beds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td>1.00</td>
<td>0.99</td>
<td>0.91</td>
<td>0.38</td>
<td>0.55</td>
</tr>
<tr>
<td>Root Population</td>
<td>0.99</td>
<td>1.00</td>
<td>0.96</td>
<td>0.39</td>
<td>0.54</td>
</tr>
<tr>
<td>Log Population</td>
<td>0.91</td>
<td>0.96</td>
<td>1.00</td>
<td>0.39</td>
<td>0.49</td>
</tr>
<tr>
<td>Hospitals</td>
<td>0.38</td>
<td>0.39</td>
<td>0.39</td>
<td>1.00</td>
<td>0.65</td>
</tr>
<tr>
<td>Hospital Beds</td>
<td>0.55</td>
<td>0.54</td>
<td>0.49</td>
<td>0.65</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 4.2 Pearson Correlation Matrix Among the Study Variables for 2,000 U.S. Counties
a U-statistic, consider two bivariate observations -- \((X_1, Y_1)\) and \((X_2, Y_2)\). The kernel for \(\tau_a\) is

\[
g\left([X_1, Y_1], [X_2, Y_2]\right) = \left(\binom{2000}{2}\right)^{-1} \text{sgn}(X_1 - X_2) \text{sgn}(Y_1 - Y_2) \tag{4.1}
\]

with

\[
\begin{align*}
-1 & \text{ if } z < 0 \\
0 & \text{ if } z = 0 \\
1 & \text{ if } z > 0 
\end{align*}
\]

This kernel will be averaged over all pairs of observations to find the rank correlation between the number of hospitals and the number of hospital beds in a county.

The second U-statistic used in this simulation is the with replacement variance component given by

\[
\sigma_{WR}^2 = \sum_{a=1}^{N} \phi(a) \left(\frac{y_a}{\phi(a)} - Y_+\right)^2 \tag{4.2}
\]

where \(Y_+\) is the population total. This statistic is useful in designing new surveys and is discussed by Folsom (1984). Divided by the sample size, this is the variance of the estimated total under with replacement sampling. This statistic was chosen because it was likely to benefit from probability proportional to size sampling if the analysis variable, \(Y\), is proportional to the size measure. The with replacement variance component will be found for the number of hospital beds in a county. A kernel for this statistic is
\[ g(a,b) = \phi(a) \phi(b) \left[ \frac{y_a}{\phi(a)} - \frac{y_b}{\phi(b)} \right]^2. \] (4.3)

In the following, the relative with replacement variance component will be reported. This is obtained by dividing by the squared total, \( Y^2 \). This does not affect the results, but simplifies the presentation.

4.3 Findings

The actual estimates were obtained using equation (2.22). The joint inclusion probabilities required in the estimation were approximated using (3.25). This provides a good example of how the development in Chapter III can be used in practice. Also, since the true population values can be calculated, a further test of the approximation can be obtained by comparing the mean of all the samples with the true value to see that an unbiased estimator results.

Tables 4.3 and 4.4 display the mean and standard error over all of the estimates for the rank correlation and relative with replacement variance component, respectively. Notice that, in every case, the mean estimate is very close to the true value (also in the tables). Hence, we see that the approximation for the joint inclusion probability is appropriate. Next, note that, for the rank correlation (Table 4.3), the standard error of the mean estimate decreases as the size measure is changed from the total population size, to the root population size, to the log of the population size. The converse is true for the relative with replacement variance component in Table 4.4. In the latter case, the relative standard error is generally smallest for the total population size measure, intermediate for the root population size measure, and largest for the log population size.
Table 4.3 Mean and Standard Error for Kendall's Rank Correlation* from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure

<table>
<thead>
<tr>
<th>Size Measure</th>
<th>Sample Size</th>
<th>Mean</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td>50</td>
<td>0.239</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.242</td>
<td>0.069</td>
</tr>
<tr>
<td>Root Population</td>
<td>50</td>
<td>0.242</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.240</td>
<td>0.038</td>
</tr>
<tr>
<td>Log Population</td>
<td>50</td>
<td>0.240</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.240</td>
<td>0.023</td>
</tr>
</tbody>
</table>

*True value = 0.234
<table>
<thead>
<tr>
<th>Size Measure</th>
<th>Sample Size</th>
<th>Mean</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.699</td>
<td>0.447</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.690</td>
<td>0.303</td>
</tr>
<tr>
<td>Root Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.670</td>
<td>0.421</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.688</td>
<td>0.321</td>
</tr>
<tr>
<td>Log Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.933</td>
<td>0.668</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.958</td>
<td>0.499</td>
</tr>
</tbody>
</table>

*True value
measure. This fits with the general intuition since the kernel for the rank correlation only takes on the values -1, 0, and 1, which will be more "proportional" to the less diverse transformed size measures. On the other hand, since the number of hospital beds in a county should grow with the population of the county, we expect that the with replacement variance component should enjoy the benefit of proportional to size sample selection.

Next, as a first attempt to assess how "normal" the distribution of the estimates is, the skewness and kurtosis coefficients of each set of 1,000 estimates is presented in Tables 4.5 and 4.6. Recall that skewness measures how symmetric the distribution is with the normal distribution having a skewness of 0. Kurtosis measures how "heavy" the tails of the distribution are with the normal distribution having a kurtosis of 3. For Kendall's rank correlation in Table 4.5 with the log population size measure, the skewness and kurtosis of the sample estimates are very close the values for a normal distribution for both sample sizes of 50 and 100. They move progressively further from the normal distribution values as the size measure is changed to the root of population size to the total population size. This again fits with the preliminary expectations. Turning to the with replacement variance component estimates in Table 4.6, we see that all of the coefficients differ markedly from the normal distribution values. However, as expected, they are closer to the normal values for the total population size measure than for the other size measures.

A graphical depiction of the frequency distribution of the estimates is given by the histograms in Figures A.1 through A.12 in
Table 4.5 Skewness and Kurtosis for Kendall's Rank Correlation from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure

<table>
<thead>
<tr>
<th>Size Measure</th>
<th>Sample Size</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td>50</td>
<td>1.30</td>
<td>6.25</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.25</td>
<td>6.38</td>
</tr>
<tr>
<td>Root Population</td>
<td>50</td>
<td>0.70</td>
<td>4.69</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.16</td>
<td>3.17</td>
</tr>
<tr>
<td>Log Population</td>
<td>50</td>
<td>-0.41</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.47</td>
<td>3.55</td>
</tr>
</tbody>
</table>
Table 4.6 Skewness and Kurtosis for Relative Replacement Variance Component from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure

<table>
<thead>
<tr>
<th>Size Measure</th>
<th>Sample Size</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>2.00</td>
<td>6.87</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.32</td>
<td>4.41</td>
<td></td>
</tr>
<tr>
<td>Root Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>3.08</td>
<td>13.40</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>2.02</td>
<td>6.84</td>
<td></td>
</tr>
<tr>
<td>Log Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>4.05</td>
<td>22.18</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>2.66</td>
<td>10.31</td>
<td></td>
</tr>
</tbody>
</table>

101
the Appendix. The data displayed in this Appendix has been standardized by subtracting the true value of the population statistic from each estimate and then dividing by the empirical standard error of the 1,000 independent estimates. The first six histograms, for Kendall's rank correlation, are fairly symmetric. Contrast this with the next six histograms, for the with replacement variance component, which show a highly skewed distribution with a long right hand tail.

A more specific graphical comparison with the normal distribution is given in Figures A.13 through A.24. Here normal probability plots compare the standardized empirical values with the quantiles of the standard normal distribution. Each plot contains a 45° line for reference. If the observed estimates were truly normal, they would produce a straight line with slope $\sigma$ and intercept $\mu$. Since the empirical values have been standardized, they will correspond to the reference line if they approximate a normal distribution. See Sievers (1986) or Wilk and Gnanadesikan (1968) for a detailed interpretation of such plots.

Figures A.13 through A.18 contain the probability plots for the rank correlation. Notice that with the root and log population size measures (Figures A.15 - A.18), the empirical distribution conforms quite well to the standard normal. Figures A.13 and A.14 shows that the empirical distribution deviates somewhat from the normal when the total population size measure is used. This observation is further reenforced by the Kolmogorov-Smirnov (K-S) tests given in Table 4.7. The K-S test statistics were compared with the two-sided asymptotic critical values of levels 0.200, 0.100, 0.050, 0.025, and 0.010 to determine the range of the significance level of each test. This
Table 4.7 Kolmogorov-Smirnov (K-S) Test of Normality for Kendall's Rank Correlation from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure

<table>
<thead>
<tr>
<th>Size Measure</th>
<th>Sample Size</th>
<th>K-S Statistic</th>
<th>P-Value Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.084</td>
<td>&lt;.010</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.067</td>
<td>&lt;.010</td>
</tr>
<tr>
<td>Root Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.047</td>
<td>.050 - .025</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.024</td>
<td>&gt;.200</td>
</tr>
<tr>
<td>Log Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.038</td>
<td>.200 - .100</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.037</td>
<td>.200 - .100</td>
</tr>
</tbody>
</table>
shows that the empirical distribution of the rank correlation is significantly different from that of a normal distribution when using total population as a measure of size (P-value < 0.01). However, when using the other size measures, the empirical distribution differs significantly, at the 0.05 level, from the normal distribution only for samples of size 50 with the root population size measure.

Turning to the with replacement variance component, Figures A.19 through A.24 show very clearly that the empirical distribution differs markedly from a normal distribution for this statistic. This is further demonstrated by the K-S tests shown in Table 4.8. In every case, the test rejects (P-value < 0.01) that the estimates are from a normal distribution.

The final numerical findings of this study are given in Tables 4.9 and 4.10 where the empirical tail and confidence interval coverage probabilities determined for each set of 1,000 estimates are shown. These were obtained by comparing the standardized empirical values with the 0.050, the 0.025, and the 0.010 upper and lower quantiles of the standard normal distribution to find the proportion of times that the empirical values were in the tails of the distribution.

For Kendall's rank correlation (Table 4.9), the tail probabilities are very near their nominal levels for both sample sizes and are fairly symmetric when using the log population size measure. With the root population size measure, the probabilities are also well behaved. They are skewed toward the upper tail, but the overall confidence interval is near the proper nominal levels. Sampling using the total population size measure yields the most highly skewed intervals. However, even these intervals are reasonable for the total coverage probabilities of the intervals.
Table 4.8 Kolmogorov-Smirnov (K-S) Test of Normality for Relative With Replacement Variance Component from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure

<table>
<thead>
<tr>
<th>Size Measure</th>
<th>Sample Size</th>
<th>K-S Statistic</th>
<th>P-Value Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td>50</td>
<td>0.204</td>
<td>&lt;.01</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.167</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Root Population</td>
<td>50</td>
<td>0.229</td>
<td>&lt;.01</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.198</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Log Population</td>
<td>50</td>
<td>0.217</td>
<td>&lt;.01</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.215</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>
Table 4.9 Empirical Tail Probabilities for Kendall's Rank Correlation from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure

<table>
<thead>
<tr>
<th>Size Measure Nominal Tail Level</th>
<th>$n=50$</th>
<th></th>
<th></th>
<th>$n=100$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower Tail</td>
<td>Upper Tail</td>
<td>Both Tails</td>
<td>Lower Tail</td>
<td>Upper Tail</td>
<td>Both Tails</td>
</tr>
<tr>
<td>Total Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td>0.013</td>
<td>0.063</td>
<td>0.076</td>
<td>0.005</td>
<td>0.070</td>
<td>0.075</td>
</tr>
<tr>
<td>0.025</td>
<td>0.004</td>
<td>0.047</td>
<td>0.051</td>
<td>0.000</td>
<td>0.052</td>
<td>0.052</td>
</tr>
<tr>
<td>0.010</td>
<td>0.000</td>
<td>0.031</td>
<td>0.031</td>
<td>0.000</td>
<td>0.032</td>
<td>0.032</td>
</tr>
<tr>
<td>Root Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td>0.017</td>
<td>0.057</td>
<td>0.074</td>
<td>0.029</td>
<td>0.072</td>
<td>0.101</td>
</tr>
<tr>
<td>0.025</td>
<td>0.006</td>
<td>0.038</td>
<td>0.044</td>
<td>0.014</td>
<td>0.040</td>
<td>0.054</td>
</tr>
<tr>
<td>0.010</td>
<td>0.002</td>
<td>0.029</td>
<td>0.031</td>
<td>0.003</td>
<td>0.015</td>
<td>0.018</td>
</tr>
<tr>
<td>Log Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td>0.045</td>
<td>0.056</td>
<td>0.101</td>
<td>0.036</td>
<td>0.075</td>
<td>0.111</td>
</tr>
<tr>
<td>0.025</td>
<td>0.024</td>
<td>0.019</td>
<td>0.043</td>
<td>0.020</td>
<td>0.028</td>
<td>0.048</td>
</tr>
<tr>
<td>0.010</td>
<td>0.011</td>
<td>0.006</td>
<td>0.017</td>
<td>0.011</td>
<td>0.011</td>
<td>0.022</td>
</tr>
</tbody>
</table>
Moving on to the with replacement variance component, the situation is different in Table 4.10. As noted when considering the frequency histograms in Figures A.7 - A.12, the distribution of the empirical values is very skewed which results in none of the standardized estimates falling into the lower tail. However, the total probability of the confidence interval when using the total population size measure is near the correct nominal level. When sampling with the other two size measures, the situation is much worse. Again, none of the estimates fall in the lower tail and the empirical level of the confidence interval differs substantially from the nominal level.

4.4 Conclusions

The results of this simulation study are mixed. The validity of the large sample theory developed in Chapter II is evident from the results for Kendall's rank correlation. It is clear that the large sample distribution of estimates for this statistic under Sampford's method is approaching that of a normal distribution. The rate of convergence appears to be related to the variability of the selection probabilities. Less variable selection probabilities seem to enhance the rate of convergence. The potential advantages of probability proportional to size sampling are mitigated by the restricted range of the kernel. However, even with highly variable selection probabilities, the normal distribution provides an adequate approximation for sample sizes of 50 to 100.

On the other hand, a warning is sounded by the results for the with replacement variance component. For this statistic, there is some evidence that the large sample distribution of the estimates is
Table 4.10 Empirical Tail Probabilities for Relative With Replacement Variance Component from 1,000 Replicated Samples of Sizes 50 and 100 by Type of Size Measure

<table>
<thead>
<tr>
<th>Size Measure Nominal Tail Level</th>
<th>n=50</th>
<th></th>
<th></th>
<th>n=100</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td>Both</td>
<td>Lower</td>
<td>Upper</td>
<td>Both</td>
</tr>
<tr>
<td>Total Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td>0.000</td>
<td>0.106</td>
<td>0.106</td>
<td>0.000</td>
<td>0.080</td>
<td>0.080</td>
</tr>
<tr>
<td>0.025</td>
<td>0.000</td>
<td>0.092</td>
<td>0.092</td>
<td>0.000</td>
<td>0.047</td>
<td>0.047</td>
</tr>
<tr>
<td>0.010</td>
<td>0.000</td>
<td>0.052</td>
<td>0.052</td>
<td>0.000</td>
<td>0.029</td>
<td>0.029</td>
</tr>
<tr>
<td>Root Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td>0.000</td>
<td>0.061</td>
<td>0.061</td>
<td>0.000</td>
<td>0.086</td>
<td>0.086</td>
</tr>
<tr>
<td>0.025</td>
<td>0.000</td>
<td>0.047</td>
<td>0.047</td>
<td>0.000</td>
<td>0.082</td>
<td>0.082</td>
</tr>
<tr>
<td>0.010</td>
<td>0.000</td>
<td>0.040</td>
<td>0.040</td>
<td>0.000</td>
<td>0.074</td>
<td>0.074</td>
</tr>
<tr>
<td>Log Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td>0.000</td>
<td>0.038</td>
<td>0.038</td>
<td>0.000</td>
<td>0.059</td>
<td>0.059</td>
</tr>
<tr>
<td>0.025</td>
<td>0.000</td>
<td>0.027</td>
<td>0.027</td>
<td>0.000</td>
<td>0.056</td>
<td>0.056</td>
</tr>
<tr>
<td>0.010</td>
<td>0.000</td>
<td>0.025</td>
<td>0.025</td>
<td>0.000</td>
<td>0.056</td>
<td>0.056</td>
</tr>
</tbody>
</table>
moving toward a normal distribution. However, for sample sizes of 50 to 100 under Sampford's method, the distribution of the estimates is still highly skewed with a long right hand tail. The slower rate of convergence for this statistic is not completely surprising since variances tend to generate a chi-square distribution with a long right hand tail. Such a distribution is slow to approach normality in large samples. It was hoped that unequal probability selection proportional to total population size would hasten the approach to normality when compared to the more restricted range size measures. While there is some slight evidence that this is the case, the effect was far from substantial. The empirical distribution of the estimates deviated greatly from that of a normal distribution.
V. RECOMMENDATIONS FOR FUTURE RESEARCH

5.1 Weak Convergence to a Gaussian Process

The literature in Section 1.3.2 presents some of the recent developments concerning the weak convergence of a sequence of U-statistics to a Brownian motion process. This approach was taken by Sen (1972) for equal probability sampling from a finite population without replacement and by Sen (1980) for successive sampling with varying probabilities. One advantage of this approach is that it easily accommodates the problem of a random sample size. This would be especially useful when studying subgroups of the population where the sample size is a random variable.

To demonstrate weak convergence to a Gaussian process, it is first necessary to show the convergence of the finite dimensional distributions (Billingsley, 1986). The added property of either tightness or relative compactness (Billingsley, 1968) is also required to complete the proof of weak convergence. The research in this dissertation, which shows asymptotic normality, extends immediately to the convergence of the finite dimensional distributions. It remains to show either tightness or relative compactness. These properties may or may not hold. The covariance structure under general unequal probability sampling is more complicated than under equal probability sampling. In the case of equal probability sampling, a reverse martingale structure holds which allows the verification of tightness
without any extra regularity conditions. However, in the absence of a reverse martingale structure, the proof of tightness may be quite involved and may demand additional regularity conditions. As noted in Section 1.5, I explored a reverse martingale property for unequal probability sampling. This pointed out a basic difference between equal and unequal probability sampling. In equal probability sampling, the data are exchangeable in that all permutations of the data values share the same joint distribution. This is not necessarily the case in unequal probability sampling because each value is associated with its own probability of being observed. A conditional reverse martingale property can be shown by conditioning on the sample and using the exchangeability of the random relabeling distribution defined in Section 2.2. It may be possible to use this as a starting point for further development.

5.2 Variance Estimation

The main focus of this research was on determining the asymptotic distribution of a U-statistic in unequal probability sampling. To make the results fully useful in practice, a consistent variance estimator is required.

An unbiased estimator is given by Folsom (1984) as

$$\text{var}(U) = \left(\frac{n}{m}\right)^{-2} \sum_{a \in [S,m]} \sum_{b \in [S,m]} n(a)n(b)w(a,b) \left[ \frac{q(a)}{\bar{f}(a)} - \frac{q(b)}{\bar{f}(b)} \right]^2$$

(5.1)

where \( n(a) \) and \( n(b) \) are the number of times that the degree \( m \) subsets \( a \) and \( b \) are included in the sample, see (2.16), and
\[ w(a,b) = \frac{E[n(a)] E[n(b)]}{E[n(a) n(b)]} - 1. \]

The ordering implied in the above summation requires that the two subsets are not equal and that any pair of subsets appears only once in the summation. This estimator is analogous to the Yates-Grundy-Sen estimator for a degree one total.

While this estimator is unbiased, it is difficult to implement because it requires summing \( \binom{n}{m} \left[ \binom{n}{m} - 1 \right]/2 \) terms. For example, if \( n=200 \) and \( m=2 \), this is almost 198 million terms. In addition, this would also require calculating almost 65 million quadruple inclusion probabilities. Even with an approximation for the multiple inclusion probabilities like that in (3.25), such an estimator is currently impractical except in special situations.

One situation where Folsom's unbiased estimator might be practical is in stratified designs. If the parameter is defined as a weighted sum across the strata, the variance can be estimated separately for each stratum and then summed to obtain the overall variance estimate. Typically, the within stratum sample sizes will be small enough to make the estimator computationally feasible.

Another option would be to sample the terms in (5.1) rather than to calculate them all. If we view the \( \binom{n}{m} \left[ \binom{n}{m} - 1 \right]/2 \) terms as a population, it should be possible to obtain an excellent estimate of the population total, the estimated variance of \( U \), with a sample of only a few thousand terms. A systematic sampling plan could be easily implemented in this situation.
In another vein, replicated samples provide an alternative variance estimation option as shown by McCarty (1966). Under this approach the sample is designed to consist of several independent subsamples. The variability among the subsample estimates can be used to estimate the overall variance. This is the approach used in the numerical simulation.

Another method of variance estimation that might prove fruitful is Sen's (1960) components approach. Folsom (1984) shows that this leads to an asymptotically unbiased estimator for with replacement sampling and unequal probabilities. The extension of this method to without replacement sampling would be computationally more efficient than Folsom's unbiased estimator.

A similar recommendation can be made for further research into jackknife variance estimators for U-statistics in unequal probability samples. Sen (1960 and 1977) and Arvesen (1969) studied this approach for i.i.d. observations. Folsom (1984) investigated jackknifing under with replacement selection of primary sampling units in unequal probability designs. The jackknife approach could lead to a computational efficient estimator.


5.3 Extension to Other Sample Designs

The current research is specific to single stage unequal probability designs. Specific examples are given for Sampford's
method of selection. Extensions to more complex sample designs are needed. This can take two forms.

First, under the single stage framework, research into optimal properties for the multiple inclusion probabilities is needed. In the numerical simulation, the effect on the rate of convergence to normality of changes in the size measure was noted. In fact, it is possible for the large sample distribution to not be normal under certain designs for some populations. As an analogy, note that for i.i.d. observations when the population is stationary of order 1 rather than 0 (see Section 1.3.1), the asymptotic distribution is a weighted sum of single degree of freedom chi-square random variables, (Serfling, 1980).

Second, extensions to stratified and multistage or multiphase designs are needed. Folsom (1984) considered U-statistic estimators and variance functions for multistage designs. This work can form the basis for extending the current asymptotic distribution results to multistage designs.
REFERENCES


Figure A.1 Frequency Distribution
Kendall's Rank Correlation
1,000 Replicated Samples of Size 50
PPS to Population Size
Figure A.2 Frequency Distribution
Kendall's Rank Correlation
1,000 Replicated Samples of Size 100
PPS to Population Size
Figure A.3 Frequency Distribution
Kendall's Rank Correlation
1,000 Replicated Samples of Size 50
PPS to Root Population Size
Figure A.4 Frequency Distribution
Kendall's Rank Correlation
1,000 Replicated Samples of Size 100
PPS to Root Population Size
Figure A.5 Frequency Distribution
Kendall's Rank Correlation
1,000 Replicated Samples of Size 50
PPS to Log Population Size
Figure A.6 Frequency Distribution
Kendall's Rank Correlation
1,000 Replicated Samples of Size 100
PPS to Log Population Size
Figure A.7 Frequency Distribution
With Replacement Variance Component
1,000 Replicated Samples of Size 50
PPS to Population Size
Figure A.8 Frequency Distribution
With Replacement Variance Component
1,000 Replicated Samples of Size 100
PPS to Population Size
Figure A.9 Frequency Distribution
With Replacement Variance Component
1,000 Replicated Samples of Size 50
PPS to Root Population Size
Figure A.10 Frequency Distribution
With Replacement Variance Component
1,000 Replicated Samples of Size 100
PPS to Root Population Size
Figure A.11 Frequency Distribution
With Replacement Variance Component
1,000 Replicated Samples of Size 50
PPS to Log Population Size
Figure A.12 Frequency Distribution
With Replacement Variance Component
1,000 Replicated Samples of Size 100
PPS to Log Population Size
Figure A.13 Normal Probability Plot
Kendall's Rank Correlation
1,000 Replicated Samples of Size 50
PPS to Population Size
Figure A.14 Normal Probability Plot
Kendall's Rank Correlation
1,000 Replicated Samples of Size 100
PPS to Population Size
Figure A.15 Normal Probability Plot
Kendall's Rank Correlation
1,000 Replicated Samples of Size 50
PPS to Root Population Size
Figure A.16 Normal Probability Plot
Kendall's Rank Correlation
1,000 Replicated Samples of Size 100
PPS to Root Population Size
Figure A.17 Normal Probability Plot
Kendall’s Rank Correlation
1,000 Replicated Samples of Size 50
PPS to Log Population Size
Figure A.18 Normal Probability Plot
Kendall's Rank Correlation
1,000 Replicated Samples of Size 100
PPS to Log Population Size
Figure A.19 Normal Probability Plot
With Replacement Variance Component
1,000 Replicated Samples of Size 50
PPS to Population Size
Figure A.20 Normal Probability Plot
With Replacement Variance Component
1,000 Replicated Samples of Size 100
PPS to Population Size
Figure A.21 Normal Probability Plot
With Replacement Variance Component
1,000 Replicated Samples of Size 50
PPS to Root Population Size

Legend: —— Empirical Quantiles
         — Normal Reference Line
Figure A.22 Normal Probability Plot
With Replacement Variance Component
1,000 Replicated Samples of Size 100
PPS to Root Population Size
Figure A.23 Normal Probability Plot
With Replacement Variance Component
1,000 Replicated Samples of Size 50
PPS to Log Population Size
Figure A.24 Normal Probability Plot
With Replacement Variance Component
1,000 Replicated Samples of Size 100
PPS to Log Population Size