



A response surface approach to assessing the relative efficiency of adaptive cluster sampling  
by Philip James Turk

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of  
Philosophy in Statistics  
Montana State University  
© Copyright by Philip James Turk (2003)

**Abstract:**

In this paper, we provide motivation and background behind adaptive cluster sampling (ACS). Several initial sampling designs are discussed with respect to theory and comparison to conventional sampling. A review of the literature for ACS is given for the years 1990 through 2002. Factors and issues determining the relative efficiency of ACS are discussed. We discuss further developments including initial unequal probability sampling, other types of estimators, detectability, bootstrapping, and limiting sampling effort and cost.

Using a designed experiment, multiple factors known to influence the efficiency of ACS were studied via response surface methodology where data were generated through computer simulation. This approach allowed us to characterize significant interaction and quadratic effects as opposed to ignoring them as has been done in the literature. Using simple random sampling without replacement as a comparison, two responses were considered: the relative efficiency for the Horvitz-Thompson estimator of  $T$  (REHT) and the relative efficiency for the Hansen-Hurwitz estimator of  $T$  (REHH). Response surface plots at given conditions and grid searches were used to find factorial settings yielding variance-optimal relative efficiencies. Conditions that optimize REHT were harder to characterize than for REHH. Different simulation approaches from the literature were compared with ours. A concern regarding any investigation and the validity of conclusions regarding important factors is the dependence on the simulation approach used to generate the study area population. The two estimation procedures were compared both through the response surfaces and through distributions of data gathered under varying population settings. Horvitz-Thompson estimation was relatively more efficient than Hansen-Hurwitz estimation. However, the variability of REHT was greater than the variability of REHH, most notably due to extremely low  $\text{Var} [THT]$  in some populations that were neither patchy nor rare. This low variance came at the price of a large final sample size. Finally, other factors and issues are presented with respect to ideas towards future research.

A RESPONSE SURFACE APPROACH TO ASSESSING THE RELATIVE  
EFFICIENCY OF ADAPTIVE CLUSTER SAMPLING

by

Philip James Turk

A dissertation submitted in partial fulfillment  
of the requirements for the degree

of

Doctor of Philosophy

in

Statistics

MONTANA STATE UNIVERSITY  
Bozeman, Montana

April 2003

D378  
T847

APPROVAL

of a dissertation submitted by

Philip James Turk

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. John Borkowski John J. Borkowski 4/18/03  
(Signature) Date

Approved for the Department of Statistics

Dr. Ken Bowers Kenneth Z. Bowers 4/18/03  
(Signature) Date

Approved for the College of Graduate Studies

Dr. Bruce McLeod Bruce L. McLeod 4-21-03  
(Signature) Date

## STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U. S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Signature



Date

7/18/03

## ACKNOWLEDGEMENTS

The preparation of a thesis is a long, grueling endeavor which is facilitated by the help of others. I would like to thank my advisor, Dr. John Borkowski, for his patient wisdom in being able to know when to offer guidance and when to let me explore various frontiers on my own. Dr. Robert Boik and Dr. Bill Quimby were readers, a tedious task for which I am appreciative. Thanks go to Dr. Steve Cherry, Dr. Warren Esty, and Dr. Jodi Carson for serving on my committee. Finally, I would be remiss if I did not give thanks for my "support system"; my friends who never let me give up.

## TABLE OF CONTENTS

LIST OF TABLES .....	vii
LIST OF FIGURES .....	xii
1. INTRODUCTION.....	1
Motivation and Background .....	1
Fixed-Population Sampling Theory and Notation .....	6
ACS With an Initial SRSWOR - Theory .....	12
ACS With an Initial SRSWR - Theory .....	17
ACS Versus Conventional Sampling - Simple Random Sampling .....	18
ACS With an Initial Sample of Primary Units - Theory .....	20
ACS Versus Conventional Sampling - Cluster and Systematic Sampling ..	24
Stratified ACS - Theory.....	25
ACS Versus Conventional Sampling - Stratified Random Sampling .....	29
2. A REVIEW OF ADAPTIVE CLUSTER SAMPLING: 1990-2002 .....	31
Practical Issues Determining Relative Efficiency of ACS .....	31
Choice of Sampling Design .....	31
Within-Network Variability and Related Issues .....	32
Critical Value .....	34
Neighborhood Definition .....	35
Estimator Type.....	37
Unit Size .....	38
Sample Size and Cost - Theory.....	38
Geographic Rarity .....	41
Further Developments .....	43
Initial Unequal Probability Sampling .....	44
Rao-Blackwell Estimators .....	46
Other Estimators .....	53
Order Statistics .....	55
Double Sampling and Multivariate ACS.....	58
Detectability in ACS .....	61
Applications of Bootstrapping .....	63
Optimality .....	65
Limiting Sampling Effort and Cost in ACS .....	65
Some Problems With Current Thinking .....	71

3.	METHODS .....	74
	The Initial Study and Experimental Setup .....	85
	The Fixed Covariate Analysis (FCA) .....	90
	The Random Covariate Analysis (RCA).....	93
	Confirmation Trials.....	95
4.	RESULTS .....	97
	Results for $RE_{HT}^*$ and $RE_{HH}$ Data - The Fixed Covariate Analysis .....	113
	Results for $RE_{HT}^*$ and $RE_{HH}$ Data - The Random Covariate Analysis ...	131
5.	DISCUSSION .....	151
	The Choice of 250 Populations and the Use of the Median .....	151
	Simulation Approaches and Edge Effect .....	153
	An Initial and General Comparison of $RE_{HT}$ Versus $RE_{HH}$ .....	163
	The Variability of $RE_{HT}$ .....	168
	A General Overview of the Grid Searches .....	176
	Interaction Analysis at Specific Population Settings .....	180
	Controllable Factors - Sample Size, Critical Value and Area.....	181
	Sample Size .....	181
	Critical Value .....	185
	Area .....	190
	Cluster Morphology .....	196
	Shape.....	197
	Direct .....	199
	Spread.....	200
	Number of Parents and Children .....	204
	The Random Covariates - Initial Results .....	209
	Within-Network Variability.....	213
	Number of Study Area Networks .....	216
	Predicted Number of Parents .....	221
	Rarity .....	224
	Discussion of Confirmation Runs .....	230
	Other Factors and Issues With Ideas Towards Future Research .....	232
	REFERENCES CITED .....	246
	APPENDICES .....	251
	APPENDIX A - R Code .....	252
	APPENDIX B - SAS Code.....	273
	APPENDIX C - Example Populations .....	298

## LIST OF TABLES

Table	Page
1. Factors used in the study .....	75
2. Settings used to replicate study of Christman (1996a) .....	79
3. Cluster center locations .....	81
4. Original offspring locations .....	82
5. Matrix generated for example population (Figure 1) .....	82
6. Settings used in the initial study .....	85
7. Design matrix of uncoded settings used in the FCA .....	88
8. Design matrix of uncoded levels used in the RCA. All data shown are median values, $N = 250$ .....	97
9. Initial study example population characteristics .....	101
10. Numerical summary of relative efficiencies and random covari- ates from simulation for initial study where $N = 500$ . Here, na = not applicable. For the $RE_{HT}$ distribution under the Extreme 1 setting, $\mu$ and Max where so large in magnitude due to the presence of outliers, they were not included in the table .....	109
11. Numerical summary of relative efficiencies from replication of Christman (1996a) .....	109
12. Numerical summary of simulation results for 72 trials. na = not applicable .....	113
13. SAS computer output for the reduced FCA model for $RE_{HT}^*$ . Effects unique to the model for this response are preceded by # .....	114

14. SAS computer output for the reduced FCA model for  $RE_{HH}$ . Effects unique to the model for this response are preceded by #..... 115
15. FCA global grid search results for maximum and minimum predicted  $RE_{HT}^*$  and  $RE_{HH}$ . A factor marked with an asterisk indicates a change took place between minimum and maximum level settings for the predicted response. Range is the range of predicted responses. no = not observed; na = not applicable; CI = confidence interval ..... 116
16. Centerpoint FCA of  $RE_{HT}^*$ . Variables not included in the effect are set to centerpoint levels. Factor levels are for the respective minimum and maximum predicted  $RE_{HT}^*$ . The levels are uncoded and are in the factor order given in the Effect column ..... 117
17. Centerpoint FCA of  $RE_{HH}$ . Variables not included in the effect are set to centerpoint levels. Factor levels are for the respective minimum and maximum predicted  $RE_{HH}$ . The levels are uncoded and are in the factor order given in the Effect column ..... 124
18. Conditional grid search results of the FCA with fixed covariate levels for both the maximum and minimum predicted  $RE_{HT}^*$  and  $RE_{HH}$ . A controllable factor marked with an asterisk indicates a change occurred between minimum and maximum level settings for the predicted response. Range is the range of predicted responses. no = not observed; na = not applicable; CI = confidence interval ..... 125
19. Conditional FCA of  $RE_{HT}^*$  with fixed covariates set to reasonable levels. Factor levels are for the respective minimum and maximum predicted  $RE_{HT}^*$ . The levels are uncoded and are in the factor order given in the Effect column. The setting of the remaining third controllable factor is noted in the Effect column. "@" indicates the effect was included in the model. "\*" indicates a correspondence with either the maximum or minimum obtained from the conditional grid search (see Table 18)..... 125

20. Conditional FCA of  $RE_{HH}$  with fixed covariates set to reasonable levels. Factor levels are for the respective minimum and maximum predicted  $RE_{HH}$ . The levels are uncoded and are in the factor order given in the Effect column. The setting of the remaining third controllable factor is noted in the Effect column. “@” indicates the effect was included in the model. “\*” indicates a correspondence with either the maximum or minimum obtained from the conditional grid search (see Table 18)..... 131
21. SAS computer output for the reduced RCA model for  $RE_{HT}^*$ . Effects unique to the model for this response are preceded by #..... 132
22. SAS computer output for the reduced RCA model for  $RE_{HH}$ . Effects unique to the model for this response are preceded by #..... 133
23. The RCA global grid search results for maximum and minimum predicted  $RE_{HT}^*$  and  $RE_{HH}$ . A factor marked with an asterisk indicates a change took place between minimum and maximum level settings for the predicted response. Range is the range of predicted responses. no = not observed; na = not applicable; CI = confidence interval ..... 134
24. The centerpoint RCA of  $RE_{HT}^*$ . Variables not included in the effect are set to centerpoint levels. Factor levels are for the respective minimum and maximum predicted  $RE_{HT}^*$ . The levels are uncoded and are in the factor order given in the Effect column ..... 135
25. The centerpoint RCA of  $RE_{HH}$ . Variables not included in the effect are set to centerpoint levels. Factor levels are for the respective minimum and maximum predicted  $RE_{HH}$ . The levels are uncoded and are in the factor order given in the Effect column ..... 135

26. Conditional RCA of  $RE_{HT}^*$ . Any remaining random covariate not included in the effect is set to a reasonable level. Factor levels are for the respective minimum and maximum predicted  $RE_{HT}^*$ . The levels are uncoded and are in the factor order given in the Effect column. The setting of any remaining random covariate is noted in the Effect column. "@" indicates the effect was included in the model. "\*" indicates a correspondence with either the maximum or minimum obtained from the global grid search (see Table 23) ..... 136
27. Conditional RCA of  $RE_{HH}$ . Any remaining random covariate not included in the effect is set to a reasonable level. Factor levels are for the respective minimum and maximum predicted  $RE_{HH}$ . The levels are uncoded and are in the factor order given in the Effect column. The setting of any remaining random covariate is noted in the Effect column. "@" indicates the effect was included in the model. "\*" indicates a correspondence with either the maximum or minimum obtained from the global grid search (see Table 23). "†" indicates a value that was actually lower than the minimum obtained from the global grid search..... 150
28. Study population characteristics - example 1 ..... 162
29. Study population characteristics - example 2 ..... 164
30. Trial 50 settings used in the FCA..... 169
31. Trial 50 results for the 14th population ..... 171
32. Initial study example population characteristics ..... 173
33. FCA global grid search results for maximum and minimum predicted  $RE_{HT}^*$  and  $RE_{HH}$ . A factor marked with an asterisk indicates a change took place between minimum and maximum level settings for the predicted response. na = not applicable ..... 177

34. Conditional grid search results of the FCA with fixed covariate levels for both the maximum and minimum predicted  $RE_{HT}^*$  and  $RE_{HH}$ . A controllable factor marked with an asterisk indicates a change occurred between minimum and maximum level settings for the predicted response. na = not applicable ..... 178
35. The RCA global grid search results for maximum and minimum predicted  $RE_{HT}^*$  and  $RE_{HH}$ . A factor marked with an asterisk indicates a change took place between minimum and maximum level settings for the predicted response. na = not applicable ..... 179
36. Summary of medians and ranges from simulation for initial study,  $N = 500$  ..... 210
37. Results of confirmation runs. The FCA PIs immediately to the right of the RE columns are the 95% prediction intervals from the FCA. The RCA PIs, analogously, are the 95% prediction intervals from the RCA with levels of the four random covariates listed in the Setting column. Observations that were not captured by their respective prediction intervals are marked with an asterisk ..... 231
38. Pervarw vs. Spread - simulation study. Values in the table are the median Pervarw values for 100 populations, the coefficients of (multiple) determination, and  $P$ -values for the test for the significance of the linear (quadratic) association ..... 241

## LIST OF FIGURES

Figure	Page
1. An Example Population .....	84
2. A Nominal Population .....	100
3. An Extreme 1 Population .....	100
4. An Extreme 2 Population .....	100
5. Relative Efficiency Distributions, Nominal Setting, $N = 500$ .....	103
6. Relative Efficiency Distributions, Extreme 1 Setting, $N = 500$ .....	104
7. Relative Efficiency Distributions, Extreme 2 Setting, $N = 500$ .....	105
8. Random Covariate Distributions, Nominal Setting, $N = 500$ .....	106
9. Random Covariate Distributions, Extreme 1 Setting, $N = 500$ .....	107
10. Random Covariate Distributions, Extreme 2 Setting, $N = 500$ .....	108
11. Relative Efficiency Distributions for 72 Trials .....	110
12. Random Covariate Distributions for 72 Trials .....	111
13. Plot of $RE_{HT}^*$ Versus $RE_{HH}$ for 72 Trials With Reference Line $RE_{HT}^* = 1/RE_{HH}$ .....	112
14. FCA Centerpoint Plot of Shape*Spread for $RE_{HT}^*$ .....	118
15. FCA Centerpoint Plot of Shape*Critical for $RE_{HT}^*$ .....	118
16. FCA Centerpoint Plot of Direct*Critical for $RE_{HT}^*$ .....	118
17. FCA Centerpoint Plot of Spread*Parents for $RE_{HT}^*$ .....	118
18. FCA Centerpoint Plot of Spread*Area for $RE_{HT}^*$ .....	119

19.	FCA Centerpoint Plot of Spread*Sample for $RE_{HT}^*$ .....	119
20.	FCA Centerpoint Plot of Parents*Area for $RE_{HT}^*$ .....	119
21.	FCA Centerpoint Plot of Parents*Kids for $RE_{HT}^*$ .....	119
22.	FCA Centerpoint Plot of Parents*Sample for $RE_{HT}^*$ .....	120
23.	FCA Centerpoint Plot of Area*Sample for $RE_{HT}^*$ .....	120
24.	FCA Centerpoint Plot of Kids*Sample for $RE_{HT}^*$ .....	120
25.	FCA Centerpoint Plot of Shape*Spread for $RE_{HH}$ .....	121
26.	FCA Centerpoint Plot of Spread*Parents for $RE_{HH}$ .....	121
27.	FCA Centerpoint Plot of Spread*Area for $RE_{HH}$ .....	121
28.	FCA Centerpoint Plot of Spread*Critical for $RE_{HH}$ .....	121
29.	FCA Centerpoint Plot of Spread*Kids for $RE_{HH}$ .....	122
30.	FCA Centerpoint Plot of Spread*Sample for $RE_{HH}$ .....	122
31.	FCA Centerpoint Plot of Parents*Area for $RE_{HH}$ .....	122
32.	FCA Centerpoint Plot of Parents*Kids for $RE_{HH}$ .....	122
33.	FCA Centerpoint Plot of Area*Critical for $RE_{HH}$ .....	123
34.	FCA Centerpoint Plot of Area*Kids for $RE_{HH}$ .....	123
35.	FCA Centerpoint Plot of Kids*Sample for $RE_{HH}$ .....	123
36.	FCA Centerpoint Plot of Area for $RE_{HH}$ .....	123
37.	FCA Conditional Plot of Area*Critical for $RE_{HT}^*$ , Sample = 10 ....	126
38.	FCA Conditional Plot of Area*Critical for $RE_{HT}^*$ , Sample = 10 ....	126
39.	FCA Conditional Plot of Area*Sample for $RE_{HT}^*$ , Critical = 1 .....	126
40.	FCA Conditional Plot of Area*Sample for $RE_{HT}^*$ , Critical = 1 .....	126

41.	FCA Conditional Plot of Area*Sample for $RE_{HT}^*$ , Critical = 2	.....	127
42.	FCA Conditional Plot of Area*Sample for $RE_{HT}^*$ , Critical = 2	.....	127
43.	FCA Conditional Plot of Critical*Sample for $RE_{HT}^*$ , Area = 10	....	127
44.	FCA Conditional Plot of Critical*Sample for $RE_{HT}^*$ , Area = 10	....	127
45.	FCA Conditional Plot of Critical*Sample for $RE_{HT}^*$ , Area = 30	....	128
46.	FCA Conditional Plot of Critical*Sample for $RE_{HT}^*$ , Area = 30	....	128
47.	FCA Conditional Plot of Area*Critical for $RE_{HH}$ , Sample = 10	....	129
48.	FCA Conditional Plot of Area*Critical for $RE_{HH}$ , Sample = 10	....	129
49.	FCA Conditional Plot of Area*Critical for $RE_{HH}$ , Sample = 50	....	129
50.	FCA Conditional Plot of Area*Critical for $RE_{HH}$ , Sample = 50	....	129
51.	FCA Conditional Plot of Area*Sample for $RE_{HH}$ , Critical = 1	.....	130
52.	FCA Conditional Plot of Area*Sample for $RE_{HH}$ , Critical = 1	.....	130
53.	FCA Conditional Plot of Critical*Sample for $RE_{HH}$ , Area = 10	....	130
54.	FCA Conditional Plot of Critical*Sample for $RE_{HH}$ , Area = 30	....	130
55.	RCA Centerpoint Plot of Pervarw*Numnets for $RE_{HT}^*$	.....	137
56.	RCA Centerpoint Plot of Pervarw for $RE_{HT}^*$	.....	137
57.	RCA Centerpoint Plot of Rarity for $RE_{HT}^*$	.....	137
58.	RCA Centerpoint Plot of Pervarw*Numnets for $RE_{HH}$	.....	138
59.	RCA Centerpoint Plot of Pervarw*Predp for $RE_{HH}$	.....	138
60.	RCA Centerpoint Plot of Numnets*Predp for $RE_{HH}$	.....	138
61.	RCA Centerpoint Plot of Numnets for $RE_{HH}$	.....	138
62.	RCA Centerpoint Plot of Rarity for $RE_{HH}$	.....	139

63.	RCA Conditional Plot of Pervarw*Rarity for $RE_{HT}^*$ , Numnets = 73.0 .....	140
64.	RCA Conditional Plot of Pervarw*Rarity for $RE_{HT}^*$ , Numnets = 73.0 .....	140
65.	RCA Conditional Plot of Numnets*Rarity for $RE_{HT}^*$ , Pervarw = 0.000 .....	140
66.	RCA Conditional Plot of Numnets*Rarity for $RE_{HT}^*$ , Pervarw = 0.000 .....	140
67.	RCA Conditional Plot of Numnets*Rarity for $RE_{HT}^*$ , Pervarw = 0.547 .....	141
68.	RCA Conditional Plot of Numnets*Rarity for $RE_{HT}^*$ , Pervarw = 0.547 .....	141
69.	RCA Conditional Plot of Pervarw*Numnets for $RE_{HT}^*$ , Rarity = 1.009 .....	141
70.	RCA Conditional Plot of Pervarw*Numnets for $RE_{HT}^*$ , Rarity = 1.009 .....	141
71.	RCA Conditional Plot of Pervarw*Numnets for $RE_{HT}^*$ , Rarity = 4.521 .....	142
72.	RCA Conditional Plot of Pervarw for $RE_{HT}^*$ , Numnets = 73.0, Rarity = 1.009 .....	142
73.	RCA Conditional Plot of Pervarw for $RE_{HT}^*$ , Numnets = 73.0, Rarity = 4.521 .....	142
74.	RCA Conditional Plot of Rarity for $RE_{HT}^*$ , Numnets = 73.0, Pervarw = 0.000 .....	142
75.	RCA Conditional Plot of Rarity for $RE_{HT}^*$ , Numnets = 73.0, Pervarw = 0.547 .....	143
76.	RCA Conditional Plot of Pervarw*Numnets for $RE_{HH}$ , Predp = 0.8, Rarity = 1.009 .....	144

77.	RCA Conditional Plot of Pervarw*Numnets for $RE_{HH}$ , Predp = 0.8, Rarity = 1.009 .....	144
78.	RCA Conditional Plot of Pervarw*Numnets for $RE_{HH}$ , Predp = 5.3, Rarity = 3.819 .....	144
79.	RCA Conditional Plot of Pervarw*Predp for $RE_{HH}$ , Numnets = 900.0, Rarity = 1.009 .....	144
80.	RCA Conditional Plot of Pervarw*Predp for $RE_{HH}$ , Numnets = 900.0, Rarity = 1.009 .....	145
81.	RCA Conditional Plot of Pervarw*Predp for $RE_{HH}$ , Numnets = 73.0, Rarity = 3.819 .....	145
82.	RCA Conditional Plot of Pervarw*Rarity for $RE_{HH}$ , Numnets = 900.0, Predp = 0.8 .....	145
83.	RCA Conditional Plot of Pervarw*Rarity for $RE_{HH}$ , Numnets = 900.0, Predp = 0.8 .....	145
84.	RCA Conditional Plot of Pervarw*Rarity for $RE_{HH}$ , Numnets = 73.0, Predp = 5.3 .....	146
85.	RCA Conditional Plot of Numnets*Predp for $RE_{HH}$ , Pervarw = 0.547, Rarity = 1.009 .....	146
86.	RCA Conditional Plot of Numnets*Predp for $RE_{HH}$ , Pervarw = 0.000, Rarity = 3.819 .....	146
87.	RCA Conditional Plot of Numnets*Rarity for $RE_{HH}$ , Pervarw = 0.547, Predp = 0.8 .....	146
88.	RCA Conditional Plot of Numnets*Rarity for $RE_{HH}$ , Pervarw = 0.547, Predp = 0.8 .....	147
89.	RCA Conditional Plot of Numnets*Rarity for $RE_{HH}$ , Pervarw = 0.000, Predp = 5.3 .....	147
90.	RCA Conditional Plot of Predp*Rarity for $RE_{HH}$ , Pervarw = 0.547, Numnets = 900.0 .....	147

91.	RCA Conditional Plot of Predp*Rarity for $RE_{HH}$ , Pervarw = 0.547, Numnets = 900.0 .....	147
92.	RCA Conditional Plot of Predp*Rarity for $RE_{HH}$ , Pervarw = 0.000, Numnets = 73.0 .....	148
93.	RCA Conditional Plot of Numnets for $RE_{HH}$ , Pervarw = 0.547, Predp = 0.8, Rarity = 1.009 .....	148
94.	RCA Conditional Plot of Numnets for $RE_{HH}$ , Pervarw = 0.000, Predp = 5.3, Rarity = 3.819 .....	148
95.	RCA Conditional Plot of Rarity for $RE_{HH}$ , Pervarw = 0.547, Numnets = 900.0, Predp = 0.8 .....	148
96.	RCA Conditional Plot of Rarity for $RE_{HH}$ , Pervarw = 0.000, Numnets = 73.0, Predp = 5.3 .....	149
97.	Plot of $M_a$ for $RE_{HT}$ , Nominal Setting, $N = 500$ .....	154
98.	Plot of $M_a$ for $RE_{HH}$ , Nominal Setting, $N = 500$ .....	154
99.	Plot of $M_a$ for $RE_{HT}$ , Extreme 1 Setting, $N = 500$ .....	154
100.	Plot of $M_a$ for $RE_{HH}$ , Extreme 1 Setting, $N = 500$ .....	154
101.	Plot of $M_a$ for $RE_{HT}$ , Extreme 2 Setting, $N = 500$ .....	155
102.	Plot of $M_a$ for $RE_{HH}$ , Extreme 2 Setting, $N = 500$ .....	155
103.	Plot of $M_a$ for $RE_{HT}^*$ , Nominal Setting, $N = 500$ .....	155
104.	Plot of $M_a$ for $\frac{\sigma_W^2}{\sigma^2}$ , Nominal Setting, $N = 500$ .....	155
105.	Plot of $M_a$ for Numnets, Nominal Setting, $N = 500$ .....	156
106.	Plot of $M_a$ for $\hat{\lambda}_p$ , Nominal Setting, $N = 500$ .....	156
107.	Plot of $M_a$ for $\frac{E[\nu]}{n_1}$ , Nominal Setting, $N = 500$ .....	156
108.	Plot of $M_a$ for $RE_{HT}^*$ , Extreme 1 Setting, $N = 500$ .....	156
109.	Plot of $M_a$ for $\frac{\sigma_W^2}{\sigma^2}$ , Extreme 1 Setting, $N = 500$ .....	157

110. Plot of $M_a$ for Numnets, Extreme 1 Setting, $N = 500$ .....	157
111. Plot of $M_a$ for $\hat{\lambda}_p$ , Extreme 1 Setting, $N = 500$ .....	157
112. Plot of $M_a$ for $\frac{E[\nu]}{n_1}$ , Extreme 1 Setting, $N = 500$ .....	157
113. Plot of $M_a$ for $RE_{HT}^*$ , Extreme 2 Setting, $N = 500$ .....	158
114. Plot of $M_a$ for $\frac{\sigma_w^2}{\sigma^2}$ , Extreme 2 Setting, $N = 500$ .....	158
115. Plot of $M_a$ for Numnets, Extreme 2 Setting, $N = 500$ .....	158
116. Plot of $M_a$ for $\hat{\lambda}_p$ , Extreme 2 Setting, $N = 500$ .....	158
117. Plot of $M_a$ for $\frac{E[\nu]}{n_1}$ , Extreme 2 Setting, $N = 500$ .....	159
118. Study Population in Study Area B - Example 1 .....	160
119. Resulting Study Populations in Study Area A - Example 1 .....	161
120. Study Population in Study Area B - Example 2 .....	163
121. Resulting Study Populations in Study Area A - Example 2 .....	163
122. The 14th Population for Trial 50 Displaying Only Units Con- taining One Or More Elements.....	170
123. Trial 50 Results.....	174

## ABSTRACT

In this paper, we provide motivation and background behind adaptive cluster sampling (ACS). Several initial sampling designs are discussed with respect to theory and comparison to conventional sampling. A review of the literature for ACS is given for the years 1990 through 2002. Factors and issues determining the relative efficiency of ACS are discussed. We discuss further developments including initial unequal probability sampling, other types of estimators, detectability, bootstrapping, and limiting sampling effort and cost.

Using a designed experiment, multiple factors known to influence the efficiency of ACS were studied via response surface methodology where data were generated through computer simulation. This approach allowed us to characterize significant interaction and quadratic effects as opposed to ignoring them as has been done in the literature. Using simple random sampling without replacement as a comparison, two responses were considered: the relative efficiency for the Horvitz-Thompson estimator of  $\tau$  ( $RE_{HT}$ ) and the relative efficiency for the Hansen-Hurwitz estimator of  $\tau$  ( $RE_{HH}$ ). Response surface plots at given conditions and grid searches were used to find factorial settings yielding variance-optimal relative efficiencies. Conditions that optimize  $RE_{HT}$  were harder to characterize than for  $RE_{HH}$ . Different simulation approaches from the literature were compared with ours. A concern regarding any investigation and the validity of conclusions regarding important factors is the dependence on the simulation approach used to generate the study area population. The two estimation procedures were compared both through the response surfaces and through distributions of data gathered under varying population settings. Horvitz-Thompson estimation was relatively more efficient than Hansen-Hurwitz estimation. However, the variability of  $RE_{HT}$  was greater than the variability of  $RE_{HH}$ , most notably due to extremely low  $Var[\hat{\tau}_{HT}]$  in some populations that were neither patchy nor rare. This low variance came at the price of a large final sample size. Finally, other factors and issues are presented with respect to ideas towards future research.

## CHAPTER 1

## INTRODUCTION

Motivation and Background

Adaptive sampling is a type of sampling scheme in which the procedure for selecting units to include in the sample may depend on values of the variable of interest observed during the survey, i.e. the sampling is “adapted” to the data (Thompson and Seber 1996). A specific type of adaptive scheme is adaptive cluster sampling (ACS). ACS operates under the rule that when the observed value of a randomly selected unit satisfies some condition of interest,  $C$  (for example  $y_i \geq c_a$ , where  $c_a$  is a fixed *critical value*), other additional units in some pre-defined accompanying neighborhood are added to the sample. In turn, if these additional units satisfy  $C$ , then their unit neighborhoods are added to the sample as well, and so on. This process stops when no further units satisfying  $C$  are encountered. “In general terms, it means that if you find what you are looking for at a particular location you sample in the vicinity of that location with the hope of obtaining even more information.” (Salehi and Seber 1997a, p. 959).

An ACS design is distinguished from the classical probability based sampling designs in that the sample selection depends on the observed unit values of the variable of interest. Historically, the ACS designs arose from informative designs (Cassel,

Särndal, and Wretman 1977), sequential statistical methods, network sampling and snowball sampling and this history, along with references, is well summarized in Thompson (1992). The most important application for ACS occurs for those populations where the number of individuals is small and the individuals show a tendency to cluster. In these cases, were one to use classical designs, the majority of measurements will be 0 and many clusters will be missed. Consequently, variances associated with estimation of the population mean or total will be large and resulting confidence intervals for corresponding parameters will be wide (Seber and Thompson 1994). Fields of study where ACS has been effectively and successfully used and where such populations naturally occur include, but are not confined to, ecology, biology, epidemiology, environmental sciences, demography and geology (Brown 1994; Dryver and Thompson 1998; Seber and Thompson 1994; Thompson 1990; Thompson and Seber 1996). For example, many populations of animals and plants have aggregation tendencies due to such factors as schooling, flocking, dispersal patterns, and environmental patchiness (Thompson 1991a).

There are numerous differences between ACS and the classical designs. With the classical designs, the entire selection of units in the sample is made prior to making any observations while with ACS the selection of units in the sample depends on observed values of the variable of interest. Let the sample  $s$  be a set or sequence of labels identifying the units for observation and  $P(s)$  be the sample's selection probability. For many classical designs,  $P(s)$  is nonzero and constant, i.e. it does not depend on the

data. On the other hand, for ACS,  $P(s)$  is sometimes zero or nonconstant, i.e. it does depend on the data. Moreover, ACS assigns higher sample selection probabilities to samples that include units where  $y_i \geq c_a$ . A similar situation exists regarding the unit marginal inclusion probabilities, or  $\pi_i$ s. Many classical plans are self-weighting i.e.  $\pi_i = \pi \forall i = 1, 2, \dots, N$  units. However, for ACS, this is no longer the case because the  $\pi_i$ s depend on the data (through  $C$ ) and are therefore no longer constant (Seber and Thompson 1994). Another important difference is that classical estimators used in conjunction with ACS are biased. However, modified estimators exist for ACS that are design-unbiased for the population mean along with design-unbiased estimators of their variances. That is, the unbiasedness depends on the way the sample is selected and does not depend on any assumptions about the population. In other words, if  $t$  is the estimator and  $t_s$  is the estimate computed when sample  $s$  is selected, then we have:

$$E[t] = \sum_s t_s P(s) = \mu$$

where the summation is taken over all possible samples. Also, the Rao-Blackwell method has been used to obtain smaller variance design-unbiased estimators for ACS. It should be noted that although design-unbiased estimators are the ones generally used when conducting ACS, estimators based on models are used on occasion. Frequentist model-based and likelihood model-based estimation along with stochastic population theory is thoroughly discussed in Thompson and Seber (1996). Only the

design-based approach is discussed in this dissertation and the reader should bear this in mind when terms like “unbiased”, for example, are used.

There are numerous advantages that ACS can and/or does have over the classical designs. In many instances ACS can be a more efficient design, i.e. the variances of estimators are smaller for an equivalent amount of sampling effort. Estimators derived via ACS make use of unit labels which can result in lower variance than estimators that do not make use of labels (Seber and Thompson 1994; Thompson and Seber 1996): Indeed, virtually every paper cited in this dissertation deals to some extent with the conditions under which ACS is a more efficient design for an equivalent amount of sampling effort. Because the location and shape of clusters of individuals is oftentimes not known prior to the survey, ACS can be used in situations where stratification may not be possible (Thompson 1991b). ACS also can find local maxima or areas of high abundance, thus increasing the yield of the sample. This attribute can be of particular importance, for example, when sampling rare or endangered species where gathering as much information as possible about individuals is desired (Seber and Thompson 1994). Another major advantage of the ACS design is its flexibility in terms of its construction. One can modify the initial sample size, the unit size, the configuration of the neighborhood, and the sampling condition  $C$  relative to the situation to achieve the most efficient design. Furthermore a variety of methods for choosing the initial sample exist, such as simple random sampling

with replacement (SRSWR) and simple random sampling without replacement (SR-SWOR) (Thompson 1990), strip sampling, systematic sampling (Thompson 1991a), stratified sampling (Thompson 1991b), sampling via probability proportional to size (Pontius 1997; Roesch 1993; Smith, Conroy, and Brakhage 1995), and simple latin square sampling (Borkowski 1999). Finally, two other potential advantages from a cost standpoint are that the average distances between sampled units are lessened and the quadrat locations are easier to find (Brown and Manly 1998; Salehi and Seber 1997a).

There are, however, some disadvantages to ACS relative to the classical designs. The final number of distinct sampled units and both initial and final number of distinct sampled networks are random. Consequently, it can be difficult to determine and control the total sampling effort and cost of the survey in advance, which is necessary for some surveys. Several papers in recent years have dealt with this issue and they are discussed in the next chapter (Brown and Manly 1998; Christman and Lan 1998; Felix-Medina and Thompson 1999; Salehi and Seber 1997a,b). Work has been done in developing the theory for using a pilot study to design an ACS survey with a given efficiency or expected cost (Salehi and Seber 1997b). Ironically, it is the very flexibility of ACS that makes finding variance optimal designs complicated. For example, the selection of  $C$  can be critical to the efficiency of the ACS design. As with the classical designs, there exists a positive probability that a sample will yield a zero abundance estimate with this issue having been addressed by Christman and

Lan (1998). Another important distinction between the two types of designs is that not all the information from sampled units is used in ACS. Specifically, *edge units*, or those sampled units that do not satisfy  $C$  but are on the boundary of a unit that does, are used only if they are part of the initial sample. This has been addressed by Dryver and Thompson (1998). Finally, the modified ACS estimators are not optimal (from a variance standpoint) among the set of possible estimators that use unit labels.

### Fixed-Population Sampling Theory and Notation

We will now consider the setting of sampling from a fixed, finite population. Most of the following discussion, in terms of theory and notation, is attributed to Thompson and Seber (1996).

Consider a region divided into  $N$  units  $(u_1, u_2, \dots, u_N)$  of equal area indexed by their respective labels  $(1, 2, \dots, N)$ . Assume that the units form a partition of the region. Here, a label identifies the location of the respective unit. Unit  $i$  will be associated with a response of interest,  $y_i, i = 1, 2, \dots, N$ . Thus, we have a vector of population  $y$ -values,  $\mathbf{y} = (y_1, y_2, \dots, y_N)'$ . Moreover, in the fixed population view, the population (or study population) is composed of  $\mathbf{y}$  where  $\mathbf{y}$  is considered to be a fixed set of unknown constants. We may write  $\boldsymbol{\theta} = \mathbf{y}$  to represent the values as a vector of unknown parameters.

A selection-ordered sample of size  $n$  is a sequence  $s_0 = (i_1, i_2, \dots, i_n)$  of  $n$  of the unit labels where the set of all possible samples is denoted by  $\mathcal{S}$ . The objective under

this setting is to select a sample, observe the  $y$ -values for the units in the sample, and estimate some function  $z(\boldsymbol{\theta})$ . Typical examples of population quantities to be estimated include the population total  $\phi(\boldsymbol{\theta}) = \sum_{i=1}^N y_i = \tau$ , the population mean  $\phi(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N y_i = \mu$ , and the finite-population variance  $\phi(\boldsymbol{\theta}) = \sum_{i=1}^N \frac{(y_i - \mu)^2}{N-1} = \sigma^2$ .

The *data*  $d_0$  consists of the  $y$ -values for the units in the sample, together with the associated unit labels, namely the selection-ordered pairs. Thus,  $d_0 = (s_0, \mathbf{y}_0)$ , where  $\mathbf{y}_0$  is the set of selection-ordered sample  $y$ -values; that is,  $\mathbf{y}_0 = (y_i : i \in s_0)$ . Furthermore, we will also be interested in  $s$ , consisting of the reduced set of  $\nu$  *distinct* labels in  $s_0$  sequentially-ordered from the smallest to largest label for uniqueness. Additionally,  $\mathbf{y}_s$  is the corresponding set of  $y$ -values in the same order as  $s$ . Also by definition,  $d_s = (s, \mathbf{y}_s)$ . It is clear that given  $s_0$  we need only  $\mathbf{y}_s$  to provide all the information about  $d_0$ . Thus, in some cases it will be convenient to redefine  $d_0$  to be equal to  $(s_0, \mathbf{y}_s)$  since both representations contain exactly the same information.

The sampling design is the procedure by which the sample is selected and is the sample selection probability function  $P_\theta(s_0)$  where the notation emphasizes the dependence of the procedure on  $\boldsymbol{\theta}$ . The sampling design also may be specified as the conditional probability  $P(s_0|\mathbf{y})$  of selecting the sample  $s_0$  such that  $P(s_0|\mathbf{y}) \geq 0$  and  $\sum_{s_0 \in \mathcal{S}} P(s_0|\mathbf{y}) = 1$  for all for all  $s_0 \in \mathcal{S}$ . In our case, we are only interested in adaptive designs, i.e. those designs such that  $P_\theta(s_0) = P(s_0|\mathbf{y}) = P(s_0|\mathbf{y}_s)$ . In other words, the procedure for selecting units depends on values of the variable of interest, but only through units included in the sample.

Also it is helpful to consider the *unordered* (with respect to both sequence and selection), reduced set  $s_R = i_1, i_2, \dots, i_\nu$  of  $\nu$  distinct labels in the sample. Additionally,  $y_R$  is the corresponding set of  $y$ -values and, by definition,  $d_R = (s_R, y_R)$ . Note that the *reduction* from  $d_0$  to either  $d_s$  or  $d_R$  is achieved by ignoring information about order and the multiplicities of the units in the sample.

Let  $\theta \in \Theta$ , where the parameter space  $\Theta$  is a subset of  $\mathfrak{R}^N$ . A parameter vector  $\theta$  is said to be *consistent* or *compatible* with datum  $d_s$  (or, equivalently, with datum  $d_R$ ) if the  $i$ th component of  $y_s$  (or, equivalently, of  $y_R$ ), equals the  $s_i$ th component of  $\theta$  for all units  $i$  in the sample. That is,  $\theta$  and  $d_s$  (or  $d_R$ ) are consistent if the  $y$  and  $\theta$  values coincide for all sample units, i.e.  $\theta$  gives rise through the design to the given  $d_s$ . Consistency depends only on the values of the distinct units in the sample, not on the order or multiplicity of selection. Therefore, if  $\theta$  is consistent with the reduced data  $d_s$  (or with  $d_R$ ) it also will be consistent with the original data  $d_0$ . We associate with every  $d_s$  a subset  $\Theta_{d_s}$  of  $\Theta$  whose vectors are consistent with  $d_s$  i.e.  $\Theta_{d_s}$  is the set of  $\theta$  that give rise through the design to the given  $s$ . Since consistency depends only on the values of the distinct units in the sample, consistency with  $d_0$  is equivalent to consistency with the reduced data  $d_s$  (and  $d_R$ ), and  $\Theta_{d_0} = \Theta_{d_s} = \Theta_{d_R}$ .

An example demonstrating the use of all the aforementioned notation can be found in Thompson and Seber (1996, p. 33-34).

We note for ACS that although the final sample size is random, the probability of getting the final sample is the same as that for the initial sample since the augmented

or adapted part of the final sample is not randomly selected and is chosen through  $C$ .

Previously, it was stated that  $P_\theta(s_0)$  depends on  $\theta$ . To clarify this statement, we note that on the set  $\Theta_{d_R}$ ,  $P_\theta(s_0)$  is constant and positive for a given  $d_0$ , but 0 otherwise. Thus,  $P_\theta(s_0)$  does not truly directly depend on  $\theta$ . Rather, it depends on  $\theta$  only through  $\Theta_{d_R}$ . As an example for an adaptive design, suppose we use simple random sampling and continue sampling until a  $y$ -value exceeds some prechosen level  $c$ . The probability of obtaining a particular sample will depend on  $\theta$ , specifically the  $y$ -values in the sample. Suppose that just the first unit in the population has a  $y$ -value exceeding  $c$ . Therefore,  $P(s_0|\mathbf{y}_s) = 1/N$  for  $s_0 = (1)$ . However, if  $\theta$  was such that only the second unit in the population had a  $y$ -value exceeding  $c$ , then  $P(s_0|\mathbf{y}_s) = 0$  for  $s_0 = (1)$  as  $\theta$  would no longer be compatible with the sample consisting of just the first unit. Hence for each  $d_0$  there exists a  $\Theta_{d_R}$  such that  $P(s_0|\mathbf{y}_s)$  does not depend on  $\theta$  for  $\theta \in \Theta_{d_R}$  and  $P(s_0|\mathbf{y}_s) = 0$  for  $\theta \notin \Theta_{d_R}$ . Since  $P(s_0|\mathbf{y}_s)$  does not depend on unobserved components of  $\theta$ , it is constant as a function of  $\theta$  for all  $\theta \in \Theta_{d_R}$ .

With adaptive designs, the probability of obtaining a specific datum  $d_0 = (s_0, \mathbf{y}_s)$  is  $P_\theta(D_0 = d_0)$  where:

$$P_\theta(D_0 = d_0) = P(s_0|\mathbf{y}_s)I_{\Theta_{d_R}}(\theta), \theta \in \Theta \quad (1.1)$$

where  $I_{\Theta_{d_R}}(\theta)$  is an indicator function taking the value 1 whenever  $\theta \in \Theta_{d_R}$  and 0 otherwise. We can also view this probability distribution as the likelihood function of the parameter  $\theta$  i.e.  $L(\theta; d_0) = P_\theta(D_0 = d_0)$ . Note that given the value of  $d_0$ ,

$P(s_0|y_s)$  is fixed and the likelihood takes only two values. Thus, the likelihood is maximized when  $\theta \in \Theta_{d_R}$  so that any value of  $\theta$  consistent with the data will be a maximum likelihood estimator.

We now state the following theorems.

**THEOREM 1.1.** *For an adaptive design satisfying equation 1.1,  $D_R$  is a minimally sufficient statistic for  $\theta$ .*

**THEOREM 1.2.** *Let  $T(D_0)$  be any estimator of  $\phi(\theta)$  that is not a function of  $D_R$ . Define  $T_{RB} = E[T|D_R]$ . Then:*

1.  $T_{RB}(D_R)$  is an estimator of  $\phi(\theta)$ .
2.  $E[T_{RB}] = E[T]$
3.  $MSE[T_{RB}] \leq MSE[T]$  with strict inequality for all  $\theta \in \Theta$  such that  $P_\theta(T \neq T_{RB}) > 0$ .

The reader is referred to Thompson and Seber (1996) for the proofs of these two theorems.

A corollary to the preceding theorem is:

**COROLLARY 1.3.** *If  $T$  is unbiased for  $\phi(\theta)$ , then  $Var[T_{RB}] \leq Var[T]$  with strict inequality for all  $\theta \in \Theta$  such that  $P_\theta(T \neq T_{RB}) > 0$ .*

The reader will recognize the preceding corollary as being the Rao-Blackwell theorem (Blackwell 1947; Rao 1945).

Several Rao-Blackwell estimators have been proposed over recent years (Dryver and Thompson 1998; Salehi 1999; Salehi and Seber 1997a; Thompson and Seber 1996). Unfortunately, none of these estimators are uniformly better than the others with respect to their variances. The following theorem helps explain why.

**THEOREM 1.4.** *The statistic  $D_R$  is not complete.*

The reader is referred to Cassel et al. (1977) for a proof.

If we consider any sampling designs in the fixed-population setting (adaptive and conventional), an UMVUE does not exist because it would have to have zero variance with this being due to the fact that  $D_R$  is not complete (Lehmann 1983). The incompleteness of  $D_R$  in the finite population sampling situation can be attributed to the presence of the unit labels (Thompson and Seber 1996). Lehmann (1983) showed that if the data do not include labels, then the order statistics are complete and so an UMVUE exists for those sampling schemes where the probabilities  $P_\theta(s_R)$  are known and positive for all  $s_R$ . However, deliberately ignoring the labels is not advised and has been a topic of controversy (Cassel et al. 1977; Chaudhuri and Vos 1988). For the purposes of this dissertation, this controversy is a moot point for two reasons. First, the only estimators we will be focusing on are label dependent. In other words, these estimators depend on specific knowledge of which labels are sampled because these

unit labels make it possible to determine which survey units to adaptively sample, i.e. where the survey units are, and how to calculate the estimates. Second, the estimators from the ACS designs that are label dependent generally perform much better than those estimators that do not use labels, the presence of a UMVUE within the latter class of estimators notwithstanding (Thompson and Seber 1996).

### ACS With an Initial SRSWOR - Theory

Consider a population consisting of  $N$  rectangularly-shaped units that can be arranged in a grid format. Assume an initial simple random sample of size  $n_1$  is taken without replacement from the population. In the Motivation and Background section of this chapter, the underlying process of ACS is described. Recall that the process stops when no further units satisfying  $C$  are encountered. At this point, a *cluster* of units is obtained that contains a boundary of units called *edge* units, where an edge unit is any unit not satisfying  $C$  but in the neighborhood of one that does (Thompson 1990). The final sample then consists of  $n_1$  clusters. These clusters are not necessarily distinct since two non-edge units in the same cluster could have been selected in the initial sample. If a unit selected in the initial sample does not satisfy  $C$ , then it is trivially considered to be a cluster of size 1.

Neighborhoods can be defined in a variety of ways although the first-order neighborhood consisting of the unit itself and the four adjacent units sharing a common boundary (to the north, south, east and west) is by far the most prevalent. This

form of neighborhood is appropriate for studies in which the  $y$ -values that satisfy  $C$  tend to cluster and are not oriented in any particular direction (Christman 2000). Technically, the units in the neighborhood don't have to be physically contiguous. For example, the neighborhood may be defined through a social relationship between units (Dryver and Thompson 1998). However, the neighborhood relationship is assumed to be symmetric. Also, the neighborhoods do not depend on  $\theta$ .

The *network*  $A_i$  for unit  $i$  is defined to be the cluster generated by unit  $i$  but with its edge units removed. A selection of any unit in  $A_i$  leads to the selection of all of  $A_i$ , i.e. the network is symmetric. By definition, any unit that does not satisfy  $C$  is trivially considered to be a network of size 1, i.e. is a *trivial network*. Thus, any cluster of unit size greater than 1 can be decomposed into a network whose units satisfy  $C$  and further networks (edge units) of size 1 that do not satisfy  $C$ . It is important to note the distinction between clusters and networks. Clusters are not necessarily disjoint since they may have overlapping edge units. On the other hand, the entire population of  $N$  units can be partitioned into a set of disjoint and exhaustive networks.

Once again, note that if the  $y$ -value of a sampled unit satisfies a certain condition  $C$ , for example  $y_i \geq c_a$ , then the rest of the unit's neighborhood is added to the sample. Therefore, unit  $i$  will be included in the final sample either if any unit of  $A_i$  is selected as part of the initial sample or if any unit of a network for which unit  $i$  is

an edge unit is selected. Thus, we have the marginal inclusion probability for unit  $i$ :

$$\pi_i = 1 - \left[ \binom{N - m_i - a_i}{n_1} / \binom{N}{n_1} \right] \quad (1.2)$$

where  $m_i$  denotes the number of units in  $A_i$ , and  $a_i$  denotes the number of units in networks of which unit  $i$  is an edge unit. If unit  $i$  satisfies  $C$ , then  $a_i = 0$ , while if unit  $i$  does not satisfy  $C$ , then  $m_i = 1$ .

The problem immediately becomes apparent that  $a_i$  is unknown for a unit  $i$  not satisfying  $C$ . Thus,  $\pi_i$ , the probability that unit  $i$  is included in the sample, cannot be computed from the sample data. Instead, we consider the empirically derived "partial" marginal inclusion probability:

$$\pi'_i = 1 - \left[ \binom{N - m_i}{n_1} / \binom{N}{n_1} \right] \quad (1.3)$$

Consequently, we are now dealing with a sample of  $n_1$  networks, not necessarily distinct. We can interpret  $\pi'_i$  as being the probability that at least one unit in the initial sample intersects network  $A_i$ . Later we will see that the estimators used in ACS make use of observations not satisfying  $C$  only when they are included in the initial sample. Consequently,  $\pi'_i$  can also be interpreted as being the probability that unit  $i$  is used in the estimator.

In 1990, Thompson (1990) developed an estimator based on a modification of the estimator of Horvitz and Thompson (1952) using the partial inclusion probabilities:

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{i=1}^N \frac{y_i I_i}{\pi'_i} \quad (1.4)$$

where  $I_i$  takes on the value 1 with probability  $\pi'_i$  if the initial sample intersects  $A_i$ , and 0 otherwise. It can be used when sampling is either WR or WOR.

Importantly, note that observations that do not satisfy  $C$  are ignored if they are not included in the initial sample.

Another way of expressing equation 1.4 is:

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{k=1}^K \frac{y_k^* I_k}{\alpha_k} \quad (1.5)$$

where  $y_k^*$  is the sum of the  $y$ -values for the  $k$ th network,  $K$  is the total number of distinct networks in the population and  $I_k$  takes a value 1 with probability  $\alpha_k$  if the initial sample intersects the  $k$ th network, and 0 otherwise. For each unit  $i$  in the  $k$ th network,  $\pi'_i$  will be the same, i.e. equal to  $\alpha_k$ . If there are  $x_k$  units in the  $k$ th network, then:

$$\alpha_k = 1 - \left[ \binom{N - x_k}{n_1} / \binom{N}{n_1} \right] \quad (1.6)$$

It should be emphasized, for the same reason as previously mentioned, that  $\alpha_k$  is not the actual network inclusion probability. Rather,  $\alpha_k$  is referred to as the *marginal initial intersection probability*, or the probability that at least one unit in the initial sample intersects  $A_k$ .

If there are  $x_k$  units in the  $k$ th network and  $x_j$  units in the  $j$ th network, then the probability (or *joint initial intersection probability*) that networks  $j$  and  $k$  are both intersected by at least one unit each in the initial sample is:

$$\alpha_{jk} = 1 - \left[ \binom{N - x_j}{n_1} + \binom{N - x_k}{n_1} - \binom{N - x_j - x_k}{n_1} \right] / \binom{N}{n_1} \quad (1.7)$$

The following results were derived by Thompson (1990).

$$E[\hat{\mu}_{HT}] = \mu \quad (1.8)$$

$$Var[\hat{\mu}_{HT}] = \frac{1}{N^2} \left[ \sum_{j=1}^K \sum_{k=1}^K y_j^* y_k^* \left( \frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_j \alpha_k} \right) \right] \quad (1.9)$$

with unbiased estimator of the variance:

$$\widehat{Var}[\hat{\mu}_{HT}] = \frac{1}{N^2} \left[ \sum_{j=1}^K \sum_{k=1}^K y_j^* y_k^* \left( \frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_j \alpha_k} \right) I_j I_k \right] \quad (1.10)$$

where, by definition,  $\alpha_{jj} = \alpha_j$ .

It should be noted that  $\hat{\mu}_{HT}$  will have low variance if  $\alpha_k \propto y_k^*$ . In this case,  $\hat{\mu}_{HT}$  will be relatively constant and will therefore have little variability (Thompson and Seber 1996). Also note that  $\alpha_{jk}$  must be greater than 0 for all pairs of networks  $j \neq k$ .

Thompson (1990) also developed an estimator based on a modification of the estimator of Hansen and Hurwitz (1943), which can be used when sampling is either WR or WOR, namely:

$$\hat{\mu}_{HH} = \frac{1}{n_1} \sum_{i=1}^N \frac{y_i f_i}{m_i} \quad (1.11)$$

where  $f_i$  is defined to be the number of times that the  $i$ th unit in the sample appears in the estimator. The modification is necessary because the draw-by-draw selection probabilities,  $p_i = (m_i + a_i)/N$ , will generally not be known for every unit in the sample. As was the case with  $\hat{\mu}_{HT}$ , observations that do not satisfy  $C$  are ignored if they are not included in the initial sample. Another expression of  $\hat{\mu}_{HH}$  written in

terms of the  $n_1$  networks (not necessarily unique) intersected by the initial sample is:

$$\hat{\mu}_{HH} = \frac{1}{n_1} \sum_{i=1}^{n_1} w_i = \bar{w} \quad (1.12)$$

where  $w_i = \frac{1}{m_i} \sum_{j \in A_i} (y_j)$ , is the mean of the  $m_i$  observations in  $A_i$ .

The following results were derived by Thompson (1990).

$$E[\hat{\mu}_{HH}] = \mu \quad (1.13)$$

$$Var[\hat{\mu}_{HH}] = \frac{N - n_1}{N n_1 (N - 1)} \sum_{i=1}^N (w_i - \mu)^2 \quad (1.14)$$

An alternative expression is:

$$Var[\hat{\mu}_{HH}] = \frac{N - n_1}{N n_1} \sigma_B^2 \quad (1.15)$$

where  $\sigma_B^2$  is the between-network variance.

An unbiased estimator of equation 1.14 is:

$$\widehat{Var}[\hat{\mu}_{HH}] = \frac{N - n_1}{N n_1 (n_1 - 1)} \sum_{i=1}^{n_1} (w_i - \hat{\mu}_{HH})^2 \quad (1.16)$$

### ACS With an Initial SRSWR - Theory

All the preceding estimators from the previous section can be used with only minor modifications when the initial random sample is conducted with replacement. Specifically, both the marginal and joint initial intersection probabilities will be modified for Horvitz-Thompson estimation and  $\widehat{Var}[\hat{\mu}_{HH}]$  will be modified for Hansen-Hurwitz estimation. The reader is referred to Thompson and Seber (1996, p. 100) for further details.

ACS Versus Conventional Sampling - Simple Random Sampling

The sample mean of the *final* ACS sample,  $\bar{y}$ , as well as the average of the *final* ACS sample cluster means, are both biased estimators of  $\mu$  since each unit has a different selection probability (Thompson 1990). Further, the sample variance of the *final* ACS sample, is biased for  $\sigma^2$ . Of course, the *initial* sample mean selected by simple random sampling,  $\bar{y}_1$ , selected with or without replacement, is an unbiased estimator of  $\mu$ . Note that  $\bar{y}_1$  also does not make use of the unit labels.

Thompson (1990) demonstrated, via a specific example with an initial sample of fixed size  $n_1$  selected via SRSWOR from a small population, that an ACS strategy can be more efficient than a simple random sampling strategy. In other words, the variance of the estimators  $\hat{\mu}_{HT}$  in equation 1.9 and  $\hat{\mu}_{HH}$  in equation 1.14 can both be less than the variance of  $\bar{y}$  with a SRSWOR design. For comparative purposes, the expected effective final sample size  $E[\nu]$ , under ACS, was calculated and used in lieu of  $n_1$  in the formula for the variance of  $\bar{y}$ :

$$Var[\bar{y}_{SRS}] = \left( \frac{N - n_1}{N} \right) \frac{\sigma^2}{n_1} \quad (1.17)$$

In another example, Thompson used a rare, clustered population produced via simulation, setting  $C$  as  $y_i \geq 1$ . He compared the two design strategies as in the previous paragraph for a variety of initial sample sizes  $n_1$ . Relative efficiency of a given estimator  $\hat{\mu}$  was calculated as:

$$\text{efficiency}(\hat{\mu}) = Var[\bar{y}] / Var[\hat{\mu}] \quad (1.18)$$

This formula will also hold for  $\hat{\tau}$ , since  $\hat{\tau} = N\hat{\mu}$ . Obviously,  $\hat{\mu}$  is considered a more efficient estimator when  $\text{efficiency}(\hat{\mu}) > 1$ . The results showed that both  $\hat{\mu}_{HT}$  and  $\hat{\mu}_{HH}$  became increasingly more efficient as  $n_1$  increased, with the efficiency increase being far greater for  $\hat{\mu}_{HT}$  than for  $\hat{\mu}_{HH}$ . Next, the  $y$ -values were set to indicate the presence or absence of point objects in the units. Interestingly, the efficiency of  $\hat{\mu}_{HH}$  was worse (less than 1) regardless of the initial sample size. However, the efficiency of  $\hat{\mu}_{HT}$  was worse only for smaller  $n_1$ , but for larger  $n_1$  the efficiency of  $\hat{\mu}_{HT}$  was better (greater than 1) once again.

Although the design-unbiasedness of ACS does not depend on the type of population being sampled, the relative efficiency of ACS does. Thompson (1990) showed that ACS using a simple random sample without replacement of size  $n_1$  and  $\hat{\mu}_{HH}$  will have lower variance than the sample mean of a simple random sample of size  $n$  if and only if:

$$\left(\frac{1}{n_1} - \frac{1}{n}\right) \sigma^2 < \frac{N - n_1}{Nn_1(N - 1)} \sum_{k=1}^K \sum_{i \in B_k} (y_i - w_{k(i)})^2 \quad (1.19)$$

$$= \frac{N - n_1}{Nn_1} \sigma_W^2 \quad (1.20)$$

where  $k(i)$  is the  $k$ th network that includes unit  $i$ ,  $B_k$  denotes the set of units that comprise the  $k$ th network, and  $w_{k(i)}$  is the average of the  $y$ -values of the units in the  $k$ th network that includes unit  $i$ . Note that in this context  $n$  is also the size of the final adaptive sample. We see that the term on the right side of equation 1.19 contains the within-network variance component of the total population variance, or

$\sigma_W^2$ . Thus, equation 1.19 says that ACS using  $\hat{\mu}_{HH}$  will be more efficient than simple random sampling if the within-network variance of the population is “high”. This is similar to the case with the decision to use conventional cluster sampling. One would want to form clusters such that the  $y$ -values within each cluster are as variable as possible but the cluster total values are as similar as possible. Smith et al. (1995) expressed Equation 1.19 as follows:

$$\sigma^2 < \frac{f_D(1 - f_1)}{(f_D - f_1)} \sigma_W^2 \quad (1.21)$$

where the initial and final sampling fractions are  $f_1 = \frac{n_1}{N}$  and  $f_D = \frac{n}{N}$ .

In light of equation 1.19, we now have an explanation for the result that  $\hat{\mu}_{HH}$  is less efficient than  $\bar{y}$  for all initial sample sizes when the population consisted of  $y$ -values that were either 0 or 1. For this population, the within-network variance would be 0 since every network would consist of a single unit with  $y_i = 0$  or a group of one or more units each with  $y_i = 1$ . Thus, for this type of population, the ACS design using  $\hat{\mu}_{HH}$  cannot do better than the simple random sample using  $\bar{y}$ . Similarly, Christman (1996b) demonstrated that, using an initial SRSWOR,  $\hat{\mu}_{HT}$  under ACS cannot be more efficient than  $\bar{y}$  when all the networks are of size 1 or  $C$  is so restrictive that no unit meets the condition. This result could be attributed to the same aforementioned reason.

ACS With an Initial Sample of Primary Units - Theory

Two ACS designs utilizing both primary and secondary units were described by Thompson (1991a), each using modifications of the estimators of Horvitz and Thompson (1952) and Hansen and Hurwitz (1943). The first ACS design occurs when an initial SRSWOR is taken of  $n_1$  primary units where the primary units are strips or clusters (from a conventional standpoint) of an equal number of secondary units. All secondary units are sampled and further adaptive sampling is done at the secondary level.

The second ACS design occurs when an initial simple random sample without replacement is taken of  $n_1$  primary units where the primary units are formed from  $k$  systematic selections. These primary units contain an equal number of non-contiguous secondary units, all of which are sampled. Further adaptive sampling is done at the secondary level.

Let  $N$  be the number of primary units, each consisting of  $M$  secondary units. Unit  $(i, j)$  represents the  $j$ th secondary unit in the  $i$ th primary unit with associated  $y$ -value,  $y_{ij}$ . Some classical results are as follows (Scheaffer, Mendenhall, and Ott 1990).

A conventional unbiased estimator of  $\mu$  is:

$$\bar{y}_1 = \frac{1}{Mn_1} \sum_{i=1}^{n_1} y_i. \quad (1.22)$$

where  $y_i = \sum_{j=1}^M y_{ij}$ . Its variance is:

$$Var[\bar{y}_1] = \frac{1}{M^2 n_1} \sigma_Y^2 \left(1 - \frac{n_1}{N}\right) \quad (1.23)$$

where  $\sigma_Y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - M\mu)^2$ . An unbiased estimator of the variance in equation 1.23 is:

$$\widehat{Var}[\bar{y}_1] = \frac{1}{M^2 n_1} s_Y^2 \left(1 - \frac{n_1}{N}\right) \quad (1.24)$$

where  $s_Y^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (y_i - M\bar{y}_1)^2$ . Note that for an initial systematic sample with only one starting point ( $n_1 = 1$ ), an unbiased estimator of the variance is not available. It should be noted that both the conventional sample mean and the estimator in equation 1.22 used in conjunction with the *final* ACS sample are biased for  $\mu$ .

The modified Horwitz-Thompson estimator used in this setting is:

$$\hat{\mu}_{HT} = \frac{1}{MN} \sum_{k=1}^K \frac{y_k^* I_k}{\alpha_k} \quad (1.25)$$

There are, however, some differences from the simple random sampling case presented earlier in this chapter. Referring to equation 1.6, now,  $x_k$  represents the number of *primary* units in the population that intersect the  $k$ th network.  $I_k$  is an indicator function taking value 1 with probability  $\alpha_k$  if at least one primary unit of the initial sample intersects the  $k$ th network, and 0 otherwise, where  $\alpha_k$  is defined as before.  $\alpha_{jk}$  is the probability that at least one primary unit of the initial sample intersects both networks  $j$  and  $k$  where  $x_{jk}$  is defined to be the number of primary units that

intersect both networks:

$$\alpha_{jk} = 1 - \frac{\left[ \binom{N-x_j}{n_1} + \binom{N-x_k}{n_1} - \binom{N-x_j-x_k+x_{jk}}{n_1} \right]}{\binom{N}{n_1}} \quad (1.26)$$

The following results were derived by Thompson (1991a):

$$E[\hat{\mu}_{HT}] = \mu \quad (1.27)$$

$$Var[\hat{\mu}_{HT}] = \frac{1}{M^2 N^2} \left[ \sum_{j=1}^K \sum_{k=1}^K y_j^* y_k^* \left( \frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_j \alpha_k} \right) \right] \quad (1.28)$$

with unbiased estimator:

$$\widehat{Var}[\hat{\mu}_{HT}] = \frac{1}{M^2 N^2} \left[ \sum_{j=1}^K \sum_{k=1}^K y_j^* y_k^* \left( \frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_{jk} \alpha_j \alpha_k} \right) I_j I_k \right] \quad (1.29)$$

For an initial systematic sample where  $n_1 = 1$ ,  $\alpha_{jk} = 0$  for some  $j$  and  $k$ , underscoring the fact that an unbiased estimator of the variance is not available for such a design (Thompson 1991a).

In this setting, the modified Hansen-Hurwitz estimator used is:

$$\hat{\mu}_{HH} = \frac{1}{M n_1} \sum_{k=1}^K \frac{b_k y_k^*}{x_k} \quad (1.30)$$

where  $b_k$  is the number of times network  $k$  is intersected by the initial sample of primary units.

The following results were also derived by Thompson (1991a):

$$E[\hat{\mu}_{HH}] = \mu \quad (1.31)$$

$$Var[\hat{\mu}_{HH}] = \frac{\sigma_W^2}{n_1} \left( 1 - \frac{n_1}{N} \right) \quad (1.32)$$

where  $\sigma_W^2 = \frac{1}{N-1} \sum_{i=1}^N (w_i - \mu)^2$ , where  $w_i = \frac{1}{M} \sum_{k=1}^K \frac{I_{ik} y_k^*}{x_k}$ , where  $I_{ik} = 1$  if the  $i$ th primary unit intersects the  $k$ th network, and 0 otherwise.

An unbiased estimator of  $Var[\hat{\mu}_{HH}]$  is:

$$\widehat{Var}[\hat{\mu}_{HH}] = \frac{s_W^2}{n_1} \left(1 - \frac{n_1}{N}\right) \quad (1.33)$$

where  $s_W^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (w_i - \hat{\mu}_{HH})^2$ .

### ACS Versus Conventional Sampling - Cluster and Systematic Sampling

Thompson (1991a) used a rare, clustered population produced via simulation to examine two primary unit structures. Whenever one or more point objects were encountered in the primary unit, further adaptive sampling was conducted at the secondary level. The first example used initial long, thin strip plots while the second example used a spatial systematic initial sample with two randomized starting points. Once again for comparative purposes, the variance was computed, via equation 1.23, for the sample mean  $\bar{y}$  of a simple random sample of primary units with sample size equal to the expected effective primary unit sample size under the adaptive design in place of  $n_1$ . Comparisons were done for various initial sample sizes of primary units ranging from 1 up to a sampling fraction of .5.

The first example showed results similar to those from the initial simple random sample results. For all initial sample sizes, both  $\hat{\mu}_{HT}$  and  $\hat{\mu}_{HH}$  were relatively more efficient than  $\bar{y}$  and this disparity became more pronounced as  $n_1$  increased.  $\hat{\mu}_{HT}$

was more efficient than  $\hat{\mu}_{HH}$  for all initial sample sizes greater than size 1 and this disparity became more pronounced as  $n_1$  increased.

The second example yielded dramatic results. Variances using initial systematic samples were much lower than those variances obtained using initial long, thin strip primary plots. This was attributed by Thompson to the spatial process used to simulate the population data (Thompson 1991a). Adaptive strategies using  $\hat{\mu}_{HT}$  and  $\hat{\mu}_{HH}$  were far more efficient relative to simple random sampling without replacement for the population used and this disparity became more pronounced with increasing  $n_1$ . In fact, the relative efficiency of  $\hat{\mu}_{HT}$  was 1.52 for  $n_1 = 1$  and went to infinity as  $n_1$  increased to 8. The reason this occurred was because the adaptive strategy had zero variance for  $n_1 > 6$  because the  $\alpha_k$ s became 1 for all the networks in the population used in this example. Once again, the results showed  $\hat{\mu}_{HT}$  was more efficient relative to  $\hat{\mu}_{HH}$ , with the disparity becoming more pronounced with increasing  $n_1$ .

### Stratified ACS - Theory

Stratification is used when there is a known source of variability such that homogeneous units can be formed into "blocks" or strata. Used appropriately, stratification can reduce the variance of an estimator. We assume simple random samples without replacement are taken of the units in *each* stratum. One can then adaptively sample at the unit level. A problem occurs when a cluster lies in more than one stratum. One possible solution might be that if such clusters are truncated at the stratum

boundaries, then the stratum estimators are independent and can be combined to provide an overall weighted estimator of  $\mu$ . However, it will be shown that truncating clusters leads to a loss of efficiency and consequently using an estimator that allows for clusters to overlap stratum boundaries is desired (Thompson 1991b).

Let the population of  $N$  units be partitioned into  $H$  strata, with  $N_h$  units in the  $h$ th stratum,  $h = 1, 2, \dots, H$ . Unit  $(h, i)$  is defined to be the  $i$ th unit in the  $h$ th stratum with  $y$ -value  $y_{hi}$ . If an initial simple random sample without replacement of  $n_h$  units are taken from each stratum  $h$ , then we define the initial total sample size to be  $n_0 = \sum_{h=1}^H n_h$ . Some classical results are as follows (Scheaffer et al. 1990).

An unbiased estimator of  $\mu$  is:

$$\bar{y}_0 = \sum_{h=1}^H \frac{N_h}{N} \bar{y}_h \quad (1.34)$$

where  $\bar{y}_h = \sum_{i=1}^{n_h} \frac{y_{hi}}{n_h}$ . Its variance is:

$$Var[\bar{y}_0] = \frac{1}{N^2} \sum_{h=1}^H N_h(N_h - n_h) \frac{\sigma_h^2}{n_h} \quad (1.35)$$

where  $\sigma_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \mu_h)^2$ , where  $\mu_h$  is the population mean for stratum  $h$ .

Were  $\bar{y}_0$  to be used in conjunction with the *final* ACS sample, then it would be biased for  $\mu$ .

Thompson (1991b) proposed several estimators and derived a variety of results to be used in conjunction with stratified ACS. The first ACS estimator we will consider is:

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{k=1}^K \frac{y_k^* J_k}{\alpha_k} \quad (1.36)$$

where the  $K$  distinct networks are indexed *without* regard to stratum boundaries and  $I_k$  equals 1 with probability  $\alpha_k$  if the initial sample of size  $n_0$  intersects network  $k$ , and 0 otherwise. The derivation of  $\alpha_k$ , the marginal initial intersection probability, is now more complex because we now must consider probabilities of intersecting network  $k$  with initial samples in *each* of the strata. Because an initial unit in a given stratum can only intersect that part of each network in the stratum, we define  $x_{hk}$  to be the number of units in stratum  $h$  that lie in network  $k$ . Thus we have:

$$\alpha_k = 1 - \left[ \prod_{h=1}^H \binom{N_h - x_{hk}}{n_h} / \binom{N_h}{n_h} \right] \quad (1.37)$$

and the probability that the initial sample intersects both networks  $j$  and  $k$ , or the joint initial intersection probability for both networks  $j$  and  $k$ :

$$\alpha_{jk} = 1 - (1 - \alpha_j) - (1 - \alpha_k) + \left[ \prod_{h=1}^H \binom{N_h - x_{hj} - x_{hk}}{n_h} / \binom{N_h}{n_h} \right] \quad (1.38)$$

The following result was derived:

$$E[\hat{\mu}_{HT}] = \mu \quad (1.39)$$

We also have the result that  $Var[\hat{\mu}_{HT}]$  and its unbiased estimator are the same as equations 1.9 and 1.10, respectively.

In order to discuss the estimators based on a modification of Hansen and Hurwitz (1943), we will need additional notation. Let  $A_{hi}$  be the network containing unit  $(h, i)$ ,  $A_{ghi}$  be the part of  $A_{hi}$  in stratum  $g$ , and  $m_{ghi}$  be the number of units in  $A_{ghi}$ .

We then have the first of three similar estimators:

$$\hat{\mu}_{HH} = \sum_{h=1}^H \frac{N_h}{N} \bar{w}_h \quad (1.40)$$

where  $\bar{w}_h = \frac{\sum_{i=1}^{n_h} w_{hi}}{n_h}$ , where  $w_{hi} = \frac{n_h}{N_h} \frac{Y_{hi}}{\sum_{g=1}^H \frac{n_g}{N_g} m_{ghi}}$ , where  $Y_{hi}$  is the sum of the  $y$ -observations in  $A_{hi}$ .

Thompson (1991b) derived the following results:

$$E[\hat{\mu}_{HH}] = \mu \quad (1.41)$$

$$Var[\hat{\mu}_{HH}] = \frac{1}{N^2} \sum_{h=1}^H N_h (N_h - n_h) \frac{\sigma_h^2}{n_h} \quad (1.42)$$

where  $\sigma_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (w_{hi} - \bar{W}_h)^2$ , where  $\bar{W}_h = \frac{\sum_{i=1}^{N_h} w_{hi}}{N_h}$ .

An unbiased estimator of  $Var[\hat{\mu}_{HH}]$  is obtained by simply replacing  $\sigma_h^2$  in equation 1.42 with  $s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (w_{hi} - \bar{w}_h)^2$ .

A second estimator, based on the stratified ‘‘multiplicity’’ estimator of network sampling, was proposed by Thompson, who also gave a variety of literature citations on the history of this type of estimator (Thompson 1992):

$$\hat{\mu}'_{HH} = \sum_{h=1}^H \frac{N_h}{N} \bar{w}'_h \quad (1.43)$$

where  $\bar{w}'_h = \frac{\sum_{i=1}^{n_h} w'_{hi}}{n_h}$ , where  $w'_{hi} = \frac{Y_{hi}}{\sum_{g=1}^H m_{ghi}}$ .

This estimator also is unbiased for  $\mu$ .  $Var[\hat{\mu}'_{HH}]$  and its unbiased estimator,  $\widehat{Var}[\hat{\mu}'_{HH}]$ , are calculated via equation 1.42, using  $w'_{hi}$ , instead of  $w_{hi}$ .

































































































































































































































































































































































































































































































































































