Estimation in Creel Surveys Under Non-Standard Conditions

by

Philip Carl Dauk

B.Sc., University of British Columbia, 1968

M.Sc., University of Illinois, 1972

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
PH.D.
in the Department
of
Mathematics and Statistics

© Philip Carl Dauk 2001
SIMON FRASER UNIVERSITY
April 2000

All rights reserved. This work may not be
reproduced in whole or in part, by photocopy
or other means, without the permission of the author.

APPROVAL

Name: Philip Carl Dauk

Degree: Ph.D.

Title of thesis: Estimation in Creel Surveys Under Non-Standard Conditions

Examining Committee: Dr. G.A.C. Graham
Chair

Dr. Carl Schwarz
Senior Supervisor

Dr. Randy Sitter

Dr. Rick Routledge

Dr. Richard Lockhart
Internal Examiner

Dr. Kenneth Hugh Pollock
External Examiner
Department of Statistics
North Carolina State University

Date Approved:

Abstract

Complemented angler surveys are often used to estimate catch from sport fishers. These surveys consist of two components: one component to estimate catch per unit effort (CPUE) and a second component to estimate effort. Their product is then an estimator for the total catch.

In the first part of the thesis, a generalization of the standard CPUE survey to account for declining efficiency over time caused by, for example, gear saturation, is examined. This is a violation of the standard assumption of a constant catch rate. Unbiased estimators, their variances, and estimated variances are derived for roving surveys. This is applied to a gill net fishery in the Fraser River, British Columbia, Canada.

In many cases, the effort survey cannot be completely randomized. In the second part of the thesis, estimators, their variances, and estimated variances are derived for this situation using a ratio estimator for both an access and roving survey. This is also applied to a gill net fishery in the Fraser River.

Third, the above method is extended to cases where the access survey is a complex, multi-stage, stratified design. Again, estimators, their variances, and estimated variances are derived. This is applied in a critical review of the Georgia Strait Creel Survey - a yearly survey conducted by the Department of Fisheries and Oceans (Canada).

Acknowledgments

I wish to express my sincerest gratitude to all faculty, staff and fellow students of the Department of Mathematics and Statistics at Simon Fraser University for all their support and assistance. In particular, I express gratitude to my supervisor Dr. Carl Schwarz and committee members Dr. Rick Routledge and Dr. Randy Sitter.

It has been my very good fortune to have Dr. Schwarz as my thesis advisor. His contributions to this work cannot be overstated. Throughout, it was his knowledge and experience that kept the project focused and able to move past issues that to me were difficult. Providing patient guidance and encouragement, he has been mentor and friend.

Special mention is also due to Derek Bingham and Changbao Wu with whom I entered the program (both have already graduated). Derek was a study partner for the comprehensive exams and it is in large part his encouragement at the most difficult of times to which I owe my success. Changbao always gave freely and generously of his time and talents. During my studies, and in particular during my thesis work, Changbao was a constant source of insight and assistance.

I would also like to acknowledge two people who have been a source of inspiration in the past: Dr. James Zidek who taught me as an undergraduate at the University of British Columbia and awakened my interest in statistics and Dr. Colin Blyth who, while on sabbatical from Queen's University, taught me as a master's student at the University of Illinois and has since always showed interest and support in my undertakings over the years.

Finally, and not least, financial support from Simon Fraser University, and NSERC through the research grant of Dr. Schwarz, is gratefully acknowledged.

Dedication

My parents

Carl and Helen

Contents

Approval Page ii

Abstract iii

Acknowledgments iv

Dedication v

List of Tables vii

List of Figures viii

1 Introduction 1

1.1 The Basic Approach 1

1.2 Literature Review 2

1.3 The Chapters 7

2 Declining Catch Rates 10

2.1 Introduction and Motivation 10

2.2 Notation 12

2.3 Survey Design, Assumptions and Model 12

2.4 Total Effort Estimation 15

2.5 The Catch Rate Estimator 15

2.5.1 The Ratio of Means Estimator 16

2.5.2 The Mean of Ratios Estimator 19

2.6 Variance of $\hat{R}$ (λ and C_0 known) 22

2.6.1 The Form of $Var\left(\hat{R} \mid \lambda, C_0\right)$ 22

2.6.2 The Asymptotic Variance of $\hat{R}$ 31

2.6.3 A Sample Estimator for $Var\left(\hat{R} \mid \lambda, C_0\right)$ 32

2.7 Estimation of λ and C_0 34

2.7.1 Estimation of $\hat{R}$ 36

2.7.2 Estimation of $Var\left(\hat{R}\right)$ 36
2.8 Overdispersion 38
2.9 Variance of Total Catch 38
2.10 Measuring Model Effectiveness 39
2.11 Bias in $\hat{R}$ 42
2.12 Simulation Results 43
2.13 The Fraser River Study 48
2.14 Conclusions and Discussion 52
2.15 Acknowledgments 52
3 Restricted Randomization in the Effort Survey 53
3.1 Introduction and Motivation 53
3.2 Notation 55
3.3 Survey Design and Assumptions 56
3.4 Estimators for the Access Designs 57
3.4.1 A Single Overflight Survey 58
3.4.2 Multiple Overflight Surveys 61
3.5 Estimators for the Roving Designs 68
3.5.1 A Single Overflight Survey 69
3.5.2 Multiple Overflight Surveys 77
3.6 Simulation Results 87
3.7 The Fraser River Study 93
3.8 Conclusions and Discussion 96
3.9 Acknowledgments 98
4 Applications in Complex Surveys 99
4.1 Introduction 99
4.2 Notation 101
4.3 Survey Design and Assumptions 102
4.4 The Estimation of C 104
4.4.1 The Unbiasedness of $\hat{R}$ 105
4.4.2 The Variance of $\hat{R}$ 106
4.4.3 Jackknife Estimates 113
4.4.4 Design Variations 114
4.5 The Variance of Catch 115

4.6 Simulation Results 116
4.7 The Georgia Strait Creel Survey 125
4.8 Conclusions and Discussion 141
4.9 Acknowledgments 144
5 Summary and Future Research 145
Appendices
A Catch Distributions for the Markov Model 151
B Results for the Markov Model as $C_0 \to \infty$ 158
C Approximating the True Inclusion Probabilities 161
Bibliography 175

List of Tables

2.1 Defining characteristics of the simulated population. 44

2.2 Effects of varying the roving survey window for two fishing time preference scenarios on the performance of $\hat{R}$, estimators of its variance and the confidence interval coverage. (Simulated population with 500 replicated samples.) 47

2.3 Effects of varying the minimum time required for sample inclusion on bias and variance. (Simulated population.) 48

2.4 Summary of results from the Fraser River study. 51

3.1 Defining characteristics of the simulated population. 87

3.2 Performance comparison of the access survey catch estimators and their variance estimators using 500 replicated samples (simulated population). 90

3.3 Performance comparison of the roving survey catch estimators and their variance estimators using 500 replicated samples (simulated population). 92

3.4 Catch estimates and their estimated variances for a Monday through Friday period over a continuous stretch of water on the Fraser River gill net fishery (1996). 95

4.1 Simulated data: Episode counts for each scenario. 118

4.2 Simulated data: Population catch by scenario. 120

4.3 Simulated data: Performance of the $\hat{R}$ estimator for differing sampling rates on Launch Site, Interview Blocks and fishing Episodes (L=25%; H=75%). Catch rates vary by Episode in Scenario I, vary by Block in Scenario II, vary by Launch in Scenario III and vary by Block and Launch in Scenario IV. (500 replicated simulations.) 123

4.4 Simulated data: Comparison of variance and variance estimates of $\hat{R}$ for differing sampling rates on Launch Site, Interview Blocks and fishing Episodes (L=25%; H=75%). $V_s(\hat{R})$ denotes the variance of the simulated $\hat{R}$ values, $\widehat{V}(\hat{R})$ the mean of formula estimates and $\widehat{V}_j(\hat{R})$ the mean of jackknife estimates (without a finite population correction factor). Catch rates vary by Episode in Scenario I, vary by Block in Scenario II, vary by Launch in Scenario III and vary by Block and Launch in Scenario IV. (500 replicated simulations.) 124
4.5 Georgia Strait Creel Survey: Comparison of July 1998 Chinook catch estimates using various methods of analysis with estimates made by DFO. The first comparison applies the proposed multistage technique using unequal site and block selection probabilities based on observed frequencies while the second uses equal selection probabilities. The third comparison assumes a simple random sample of interviews. 132
4.6 Georgia Strait Creel Survey: Comparison of August 1998 Chinook catch estimates using various methods of analysis with estimates made by DFO. The first comparison applies the proposed multistage technique using unequal site and block selection probabilities based on observed frequencies while the second uses equal selection probabilities. The third comparison assumes a simple random sample of interviews. 133
4.7 Georgia Strait Creel Survey: Comparison of July 1998 Rockfish catch estimates using various methods of analysis with estimates made by DFO. The first comparison applies the proposed multistage technique using unequal site and block selection probabilities based on observed frequencies while the second uses equal selection probabilities. The third comparison assumes a simple random sample of interviews. 134
4.8 Georgia Strait Creel Survey: Comparison of August 1998 Rockfish catch estimates using various methods of analysis with estimates made by DFO. The first comparison applies the proposed multistage technique using unequal site and block selection probabilities based on observed frequencies while the second uses equal selection probabilities. The third comparison assumes a simple random sample of interviews. 135
4.9 Georgia Strait Creel Survey: Components of variance analysis 1998 for selected months and species. 137

C.1 Block inclusion probabilities: Estimates (initial selection probabilities × sample size). . . . 162
C.2 Block inclusion probabilities: Exact values. . . . 163
C.3 Block inclusion probabilities: Approximation (based on samples of size m=30). 163
C.4 Block inclusion probabilities: Approximation (based on samples of size m=100). 164
C.5 Block inclusion probabilities: Approximation (based on samples of size m=500). 164
C.6 Launch inclusion probabilities Part I: Estimates (initial selection probabilities × sample size). . . . 165
C.7 Launch inclusion probabilities Part II: Estimates (initial selection probabilities × sample size). . . . 166
C.8 Launch inclusion probabilities Part III: Estimates (initial selection probabilities × sample size). . . . 167
C.9 Launch inclusion probabilities Part IV: Estimates (initial selection probabilities × sample size). . . . 168
C.10 Launch inclusion probabilities Part I: Approximation (based on samples of size m=500). . . . 169
C.11 Launch inclusion probabilities Part II: Approximation (based on samples of size m=500). . . . 170
C.12 Launch inclusion probabilities Part III: Approximation (based on samples of size m=500). . . . 171
C.13 Launch inclusion probabilities Part IV: Approximation (based on samples of size m=500). . . . 172
C.14 Simulated data: Performance of the $\hat{R}$ estimator for differing sampling rates on Launch Site, Interview Blocks and fishing Episodes (L=25%; H=75%). Catch rates vary by Episode in Scenario I, vary by Block in Scenario II, vary by Launch in Scenario III and vary by Block and Launch in Scenario IV. (500 replicated simulations.) . . . 173

C.15 Simulated data: Comparison of variance and variance estimates of $\hat{R}$ for differing sampling rates on Launch Site, Interview Blocks and fishing Episodes (L=25%; H=75%). $V_s(\hat{R})$ denotes the variance of the simulated $\hat{R}$ values, $\widehat{V}(\hat{R})$ the mean of formula estimates and $\widehat{V}_j(\hat{R})$ the mean of jackknife estimates (without a finite population correction factor). Catch rates vary by Episode in Scenario I, vary by Block in Scenario II, vary by Launch in Scenario III and vary by Block and Launch in Scenario IV. (500 replicated simulations.) 174

List of Figures

2.1 Fishing episodes by length for the simulated population (total count = 225). 44

2.2 Catch vs. Time Fished for completed episodes of the simulated population. Also shown are plots of $C_0\left(1-e^{-\frac{\lambda t}{C_0}}\right)$, the expected catch at time t for true (upper line) and least square estimates (lower line) of λ and C_0. 45

2.3 Edited data for selected study areas located on the Fraser River, Canada. Lines show the expected catch using the constant catch rate model (straight lines) and the proposed saturation model (curved lines). 49

3.1 Fishing episodes by length for the simulated population (total count = 225). 88

3.2 Catch vs. Time Fished for completed episodes of the simulated population. Also shown is a plot of the expected catch using the least squares fit. 88

3.3 Count of active fishing episodes by minute of the fishing day. Overflights occur at minute 720 for the single overflight scenario or at minute 480 and minute 1080 for the multiple overflight scenario. 89

4.1 Simulated Data: Distribution of episode lengths for each scenario (total count=2,670). 116

4.2 Simulated Data: Fishing Time Preference Curve. 117

4.3 Simulated Data: Effort profile for each scenario. 119

4.4 Simulated Data: Fish caught by number of times counted in the overflight surveys. (Lines connect means of catch in each group.) 121

Chapter 1

Introduction

1.1 The Basic Approach

Many of the decisions in recreational fisheries management are based on estimates of catch. Since these decisions can be controversial and have major impact on the viability of the resource, the estimates of catch must be defensible. Such estimates are typically obtained through survey methods, historically known as creel surveys. Much attention over the past 40 years has gone to the development of these techniques (Pollock et al., 1994).

One strategy for estimating total catch C in larger studies is to employ two separate but complementary surveys that independently estimate the total fishing effort E and the catch rate R, also commonly referred to as catch per unit of fishing effort and denoted as $CPUE$. The independence of the estimators for total effort and catch rate provides for unbiasedness of the estimator for total catch if the individual estimators are also unbiased. That is, with $C = E \cdot R$, where $R = C/E$,

$$\begin{aligned} E\left[\hat{C}\right] &= E\left[\hat{E}\right] E\left[\hat{R}\right] \\ &= E \cdot \frac{C}{E} \\ &= C\,. \end{aligned}$$

The independence of the total effort and catch rate estimators has practical as well as theoretical advantages. One such advantage is overall design flexibility, which can be a useful feature, since survey designs which are optimal for estimating total effort may not necessarily be optimal for estimating catch rate.

A number of methodologies for estimating the catch rate have been developed (Pollock et al., 1994). In general, these methods involve some form of angler contact, usually interviewing. Interviews are categorized in two ways. “Access” interviews are those that are conducted after the fishing episodes have been completed. Typically these are taken as the fishermen leave the fishing resource, passing some point of access. “Roving” interviews on the other hand, are those interviews that are conducted at some time during the fishing episode. These occur as the fishermen are intercepted by a roving interviewer. An important distinguishing feature then is that “access design” surveys deal with information from completed fishing episodes while “roving design” surveys deal with information from incomplete fishing episodes.

1.2 Literature Review

Designs based on complemented surveys have become a mainstay for creel survey work in the sports fishery sector. Rose and Hassler (1969) cite early work by Eschmeyer (1942) using a sample to obtain catch per fishing trip and then multiplying this by an estimated number of trips, and similar work by Tarzwell and Miller (1943) using the same technique but stratifying their data by area, month, and weather conditions. Variations evolved with the increasing need for more accurate estimates. For their estimation of the 1961 and 1962 catch for North Carolina’s dolphin fishery, Rose and Hassler (1969) used estimates of catch per boat day and boat days. Continued demand for creel survey work over the years has led to increased complexity and substantial improvements.

The benchmark work by Robson (1960, 1961) presented models based on catch per hour and fishing hours that detailed the conditions needed for unbiased sampling and estimation in creel survey work. Subsequent research and field work encompassed not only catch estimation but also related social, economic, and political issues and has resulted in a large and growing body of knowledge that has become increasingly dispersed.

> “In recognition of [a need for consolidation of knowledge] ..., the American Fisheries Society (AFS) and the Division of Federal Aid of the U.S. Fish and Wildlife Service undertook a three-part program to produce a book of fisheries survey techniques. The first step was to convene an International Symposium and Workshop on Creel and Angler Surveys in Fisheries Management, which was held in Houston, Texas, on March 26-31, 1990. This conference brought 300 biologists,

> managers, statisticians, economists, sociologists, and theoreticians together for 5 days of intensive presentations and discussions, and it exposed a great deal of new research and recent experience relevant to fisheries surveys. The second step was peer review and publication of the symposium's 528-page proceedings, *Creel and Angler Surveys in Fisheries Management* (American Fisheries Society Symposium 12, 1991) [edited by Guthrie and seven coeditors, 1991]. The third and final step is publication of this techniques book [*Angler Survey Methods and their Applications in Fisheries Management,* Pollock et. al., 1994], which draws heavily on work presented in the symposium and proceedings, as well as on other sources familiar to us." (Pollock et al., 1994).

The combined volumes, *Creel and Angler Surveys in Fisheries Management* (American Fisheries Society Symposium 12, 1991) and *Angler Survey Methods and their Applications in Fisheries Management* (Pollock et al., 1994), provide a chronology to time of publication as well as a comprehensive summary of the research, developments, and techniques currently available.

Estimating fishing effort often involves a count of anglers which can be broadly classified as "instantaneous" or "progressive" (Hoenig et al., 1993; Pollock et al., 1994). Instantaneous counts can be made using aerial overflights, a fast-moving vehicle, or a visual vantage point. If performed at a randomly chosen moment from some time interval, then the product of the count and the length of the time interval gives a measure of effort expressed in units of the time interval (e.g. fishing hours). This estimate of effort can be improved by using the average of a number of counts taken at randomly selected times (Pollock et al., 1994; Malvestuto, 1996).

Aerial surveys are often used for instantaneous counts in larger study areas with difficult terrain (Gunderson, 1993; Pollock et al., 1994). Despite the costs of aircraft time, such surveys can be efficient and cost effective. Operated with minimal personnel, they can provide spatial and temporal information, as well as aid in such functions as identification of access sites. Proper use of aerial surveys requires adequate planning. Routes or transects and their timings must be designed to best accomplish the intended task. Sighting and determining activity can be an important issue, especially during poor weather. Methods of estimating such visibility bias are reviewed by Pollock and Kendall (1987).

Progressive counts are made by an agent travelling along a pre–chosen route that includes the entire fishing resource under study. Robson (1961) describes the agent travelling at a constant rate and making a number of passes through the study area. Hoenig et al. (1993) specifies the requirements for proper use of this method:

> " (1) the starting location along the survey agent's route is chosen randomly,
> (2) the direction of travel is chosen randomly (from the two alternatives), and
> (3) the survey agent's speed of travel is greater than that of all anglers while the anglers are fishing (but not necessarily when they are traveling from one location to another)."

Hoenig et al. (1993) also argues that one pass can be used to estimate effort, providing that the start time is chosen in such a way that all times have an equal probability of being selected. The product of the progressive counts and the time required for the pass through the study area give an estimate of effort over the "pass time" and so must be expanded by the number of "pass times" in the fishing day.

Past studies have combined angler interviewing with progressive count taking, which then exposes the estimates of effort to problems of bias (Robson, 1991; Wade et al., 1991 and Pollock et al., 1994). The amount of bias, potentially a severe underestimation, stems from missed counts during interview times (Pollock et al., 1994). Wade et al. (1991) suggest a procedure involving checkpoints along the route which force the interviewer to a time schedule.

"Bus-route-type access surveys" or simply bus route surveys offer an alternative to instantaneous and progressive counts for estimating effort when dealing with large geographic survey areas. Travelling to all access points using a route with scheduled arrival, wait, and departure times, a survey agent records the amount of time that an angler's vehicle is at each site. Robson and Jones (1989) developed an estimator for total effort based on a geometric probability of encountering anglers. Chen and Woolcock (1999) propose a condition on the waiting times that ensures unbiasedness in the estimator (*i.e.* that the sum of the travel time to, and wait time at, each access point is the same and equal to the total circuit time divided by the number of access points).

Another approach for estimating total effort, expressed as angler trips, was used by McNeish and Trial (1991). Interview data was used to construct angler activity curves expressed as the proportion of the day's anglers active at each hour. Instantaneous counts made at non-randomly selected times were then divided by the proportions at the times of overflights to estimate the total number of angler trips for that day. McNeish and Trial

(1991) note that a similar method of estimating effort was discussed by Parker (1956). The Georgia Strait Creel Survey (Department of Fisheries and Oceans, Canada) also uses interview data to construct effort profiles (English et al., 1986; Hardie et al. 1999; Shardlow et al. 1989).

The natural estimator for the catch rate involves a ratio; but until work by Pollock et al. (1997) there was no clear criterion for choosing between a mean of ratios or a ratio of means type of estimator. Attention, however, was given to the performance of various estimators, e.g. by Crone and Malvestuto (1991). As pointed out by Hoenig et al. (1997) in reference to roving surveys, there was no consensus as to which estimator should be used.

> "The ratio of means estimator has been advocated for roving creel surveys by ... Malvestuto, Davies and Shelton (1978), Phippen and Bergersen (1991), Dent and Wagner(1991), ... among others. In contrast, the mean of ratios estimator has been advocated by Siegler and Siegler (1990) and Hayne (1991) among others."

Defining C_i^* and L_i^* as catch and length of time fished, respectively, for a completed fishing episode (as recorded by an access survey) and C_i and L_i as catch and length of time fished up to time interviewed by a survey agent (as recorded by a roving survey), a ratio of means estimator is defined as

$$\hat{R}_{rom} = \frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} L_i^*} \quad \text{or} \quad \frac{\sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} L_i}$$

and a mean of ratios estimator is defined as

$$\hat{R}_{mor} = \frac{1}{n}\sum_{i=1}^{n}\frac{C_i^*}{L_i^*} \quad \text{or} \quad \frac{1}{n}\sum_{i=1}^{n}\frac{C_i}{L_i}.$$

Noting that the sample inclusion is the same regardless of length for any fishing episode in an access survey, whereas the sampling probability is dependent on the length of the time fished in a roving design, Pollock et al. (1997) showed that only the *rom* estimator is unbiased for the access design and that only the *mor* estimator is unbiased for the access design. A somewhat more detailed version with specific focus on roving designs is given by Hoenig et al., (1997). In both accounts catch rate has been assumed constant over the duration of any fishing episode.

Often creel surveys are conducted over large heterogeneous areas or for extended periods of time (e.g. the Georgia Strait Creel Survey). In these situations stratification can be used to improve the precision and facilitate monthly or annual reporting. There may also be administrative advantages. In addition, more elaborate sampling schemes may better suit the ground (*i.e.* access or roving) survey. Because of logistic constraints relating to the deployment of personnel, multistage designs are often employed. A typical example of a multistage design would be to sample anglers within a time block within a landing site. To increase the number of anglers sampled, and hence improve precision, unequal selection probabilities can be used at each stage of sampling (Hayne, 1991).

In addition to access, roving, and aerial surveys, mail and telephone surveys can offer relatively simple and inexpensive off–site methods of obtaining survey information (Pollock et al., 1994). Mail surveys, used primarily for opinion polling, have other advantages. Brown (1991) lists relatively low cost, lack of face–to–face bias, avoidance of time–pressured response, and facility for longer and more complex questions as advantages; while disadvantages included possible need for reminder notices, inability to restate or clarify questions, recall bias, and nonresponse. License files or on–site interviews are often used to construct the sampling frame. However, a weakness is that these may not include the entire target population. The sampling frame for a telephone survey can be constructed by similar means and also by also using random dialing or commercially produced directories. Apart from cost, advantages and disadvantages are generally the complement set of those for mail surveys. In addition, response from telephone surveys is immediate, and automated methods of data entry can be used, making them an attractive choice for large national surveys.

Using a hyphenated convention in which the effort survey methodology is given first followed by the $CPUE$ survey methodology, Pollock et al. (1997) noted that

> "There are four complemented designs which use complete trip interviews to estimate catch (effort $\times$ catch rate); these are mail–access, telephone–access, aerial–access, and roving–access. There are also four complemented designs which use incomplete trip interviews to estimate catch (effort $\times$ catch rate); these are mail–roving, telephone–roving, aerial–roving, and roving–roving. In terms of analysis, aerial–access and roving–access are equivalent designs, and aerial–roving and roving–roving are also equivalent designs. This is because an aerial instantaneous count is treated the same way as any other kind of roving instantaneous or progressive count."

The focus of this thesis is aerial–access and aerial–roving designs.

1.3 The Chapters

A specific set of assumptions has been fundamental for the theoretical development of catch estimation based on $CPUE$. This thesis extends the utility of the technique by considering other sets of assumptions which might better reflect the realities encountered in some applications.

Chapters 2, 3, and 4 are expanded versions of papers submitted for publication and as such, essentially are self–contained units. Consequentially, there is some redundancy of the introductory material. Where possible, consistency in notation has been maintained.

Chapter 2

This chapter is concerned with the assumption of a constant catch rate over the entire duration of any fishing episode. When determining the appropriateness of their estimators (*i.e.* for access versus roving designs), Hoenig et al. (1997) and Pollock et al. (1997) assumed that fishing was a stationary Poisson process. While this assumption may well be suitable for certain types of sport fishing, it may not be a suitable model for live bait fishing where the efficiency of the bait deteriorates over time, or for a gill net fishery where net saturation and fouling can result in a declining catch rate. More appropriate might be a non-homogeneous Poisson process in which the catch rate decreases in a "smooth" fashion over time. Even more appropriate might be a generalization of this, in which the decreases are "irregular" over time and determined by some random process that describes the times at which a catch is made. To illustrate the development of a suitable estimator for use with a declining catch rate, a river–based gill net fishery is used where catch is modeled as a continuous time Markov process.

An estimator for catch rate, $\hat{R}$, is proposed and shown to be unbiased. Next, the variance of $\hat{R}$ is developed, conditional on knowing the parameters that govern the declining catch rate. Estimation of these parameters is then discussed and a method for estimating the unconditional variance of $\hat{R}$ is presented. Issues of overdispersion, model effectiveness and bias are also discussed.

A simulation study is used to examine the performance of the estimators. One scenario of particular interest, and likely to occur in practice, is that of a restriction on the roving survey to daylight hours. Use of a minimum time to interview for sample inclusion as a variance stabilizing device is also examined.

Finally, the proposed methods are applied to a gill net fishery on the Fraser River, British Columbia, Canada.

Chapter 3

When instantaneous counts are used to estimate total effort, it is essential that these counts be taken at times randomly selected over the full fishing day; often this is not practical or economically viable, e.g. use of aircraft. This chapter considers the problem of catch estimation when the times of overflight are restricted to a non-randomly selected schedule. The ground survey for estimating $CPUE$ (*i.e.* access or roving surveys) are assumed to be fully randomized.

Ratio–type estimators using the overflight counts as auxiliary information are developed for use with both access and roving ground surveys and are shown to be unbiased. Their variance estimators are also developed. In each case these results are then extended to incorporate multiple overflights in the same day. The question of whether to combine the overflight counts using ratios of means or using means of ratios is also examined. Relative performance of these estimators is then analyzed using simulated data.

Lastly, data from the gill net fishery on the Fraser River are used to demonstrate the use of the proposed estimators.

Chapter 4

This chapter illustrates how results from the previous chapter can be used when the ground survey is access and has a complex design. In particular, a three stage sampling design is used. For generality, unequal sampling probabilities are used. It is shown that the structure of the sampling design must be built into the estimators and that a catch estimate is actually made using only the data from the randomly implemented access survey. The overflight counts are used to "fine tune" this estimate. This underscores the need for randomization in at least one of the component surveys. Development of the variance formulas is shown to be structured on a decomposition of the variance.

Next, simulations are used to assess the performance of the estimators. Scenarios with differing amounts of variability and sampling rates at each of the levels of the multistage structure are considered.

In conclusion, use of the proposed methods is demonstrated using data from the Georgia

Strait Creel Survey conducted off the west coast of Canada.

Chapter 2

Declining Catch Rates

In general, a $CPUE$ estimator must be based on completed fishing trip information. This is provided directly by "access surveys". In "roving surveys", however, estimators must be model assisted because they obtain catch and length of trip information prior to the its completion. Typically, estimates in these surveys have been developed assuming that the catch rate is constant over time. This chapter extends the problem to that of estimating total catch in the presence of a declining catch rate, due, for example, to gear saturation.

Using a gill net fishery as an example, a mean of ratios type of estimator for the catch rate together with its variance estimator is developed. Their performance is examined using simulations with special attention given to effects of restrictions on the roving survey window. The Fraser River gill net fishery is used to illustrate the use of the proposed estimator and to compare results with those from an estimator based on a constant catch rate.

2.1 Introduction and Motivation

Complemented angler surveys (Pollock et al., 1994) are often used to estimate total catch (C) in large regional surveys. In these designs, two separate and independent surveys are used to estimate the total fishing effort (E) and the catch rate ($R = C/E$) and their product then estimates the total catch.

Methods to estimate the catch rate generally involve some form of angler interviewing. Interviews are categorized in two ways: "Access" interviews conducted after the fishing episodes have been completed, and "roving" interviews conducted at some time during the fishing episode. Pollock et al. (1997) studied the problem of estimation in a sports fishery context and recommended separate estimators for the two interview methods. In a detailed study of estimation in roving surveys, Hoenig et al. (1997) recommended the same estimators. In developing estimators, both papers stated the assumption that,

> "for each angler j, fishing is a stationary Poisson process with parameter λ_j and that the fishing rate parameter does not vary with the angler's starting time or with the length of the fishing trip".

However, this may not be realistic in some fisheries. Gear may loose its efficiency over time because of fouling or saturation resulting in a declining catch rate over time.

A study of non-constant catch rates was motivated by a review of the methodology used for estimating sockeye catch on the Fraser River, Canada, in a gill net fishery. As a net fills, its ability to catch additional fish diminishes resulting in fewer catches over successive time intervals. As well, there are physical limits on the number of fish that a single net can harvest. These were believed to be significant issues for the longer "soak" times observed in the study. Estimates made, assuming a constant catch rate, may have unacceptably high positive bias.

In this chapter total catch is estimated in the presence of catch rate saturation *i.e.* for each episode a decline in the catch efficiency of the gear expressed as a function of the amount already caught. While the ultimate goal is an estimate of total catch, the major focus is on finding an appropriate estimator for the catch rate. Under a sampling design similar to that used by Hoenig et al. (1997) and with catch modelled as a Markov process, a mean of ratios estimator is appropriate and we find its variance estimator. These results are generalizations of those of Hoenig et al. (1997). Performance of these estimators is assessed using simulated data, with special attention being given to the possible effects of restrictions on the window of the roving survey. Finally, a portion of the Fraser River data is used to illustrate the use of the proposed estimator and how catch estimates can differ if catch rate saturation is ignored. The Fraser River study is also used to demonstrate the value of the estimator as a conservative means of dealing with dispersed data.

Note that, because the estimates for catch rate derived from access surveys (*i.e.* based on information from completed fishing episodes) would already account for catch rate saturation, it is only the estimators which are derived from roving designs that are of interest in this chapter.

2.2 Notation

T	Time period for one repetition of the fishing effort pattern. Typically, this will be 24 hours.
N	Number of fishing episodes in the population during one time period T.
n	Number of fishing episodes intercepted by the roving survey
I	Instantaneous count of the active fishing episodes.
C_j^*	Catch from the j^{th} fishing episode when completed, where $j = 1, \ldots, N$.
C_j	Catch from the j^{th} fishing episode at time of interview, where $j = 1, \ldots, N$. Define $C_j = 0$ if the j^{th} episode is not selected in the roving sample.
L_j^*	Length of time fished from the j^{th} fishing episode when completed, where $j = 1, \ldots, N$.
L_j	Length of time fished from the j^{th} fishing episode at time of interview, where $j = 1, \ldots, N$. Define $L_j = 0$ if the j^{th} episode is not selected in the roving sample.
L'	Minimum length of time fished required for sample inclusion.
C^*	Total catch $= \sum_{j=1}^{N} C_j^*$.
E^*	Total effort $= \sum_{j=1}^{N} L_j^*$.
R	Catch rate (i.e. $CPUE) = C^*/E^*$.
λ	Instantaneous, initial catch rate common to all episodes.
C_0	Capacity of net at full saturation.
δ_j	Indicator variable used to denote sample inclusion for the j^{th} episode.

2.3 Survey Design, Assumptions and Model

A gill net fishery provides a good opportunity to examine the problem of total catch estimation in the presence of a declining catch rate, because the declining catch rate is both plausible and easy to comprehend. As a net fills, its ability to catch additional fish diminishes resulting in fewer catches over successive time intervals. In the extreme, no further fish can be caught. Suppose that such a fishery exists and consists of N independent fishers using nets of a common design and that the fishing effort has a pattern that is cyclic with a time period T of one fishing day of 24 hours. A full survey design for estimating catch requires two separate and independent components. The first component should be designed to provide an instantaneous measure of the effort at randomly chosen times (Hoenig et al., 1993, Pollock et al., 1994), while the second should be designed to provide catch and length of time fished information. For example, a helicopter fly-over could be used to count active nets, while a roving boat patrol could be used to collect the time fished and number caught information. In the latter, the number of fish caught, C_i, and the start time of the current episode are recorded for each fisher encountered, from which the length of time fished to interview, L_i, is also obtained.

Assumptions about the process by which episodes are selected to interview are crucial in developing a catch rate estimator when using a roving survey design. Here it is assumed that each episode is sampled independently of every other episode and with probability proportional to the length of the episode. For example, the basic sampling design used by Hoenig et al. (1997): the starting point, time and direction of travel are chosen at random for a pass through the fishing resource by the interviewer. Then, as noted by Hoenig et al. (1997), with $\delta_j = 1$ denoting sample inclusion, the distribution of $L_j \mid L_j^*, \delta_j = 1$ is uniform over $\left[0\,,\, L_j^*\right]$ with $E\left[L_j | \, L_j^*, \delta_j = 1\right] = L_j^*/2$ and $Var\left(L_j | \, L_j^*, \delta = 1\right) = L_j^{*2}/12$. Also, the δ_j are independent Bernoulli random variables with $E\left[\delta_j\right] = L_j^*/T$ and $Var\left(\delta_j\right) = \left(L_j^*/T\right)\left(1 - L_j^*/T\right)$. That is, inclusion of any net in the roving survey is independent of the inclusion of any other net and depends only on the duration of that fishing episode. Further, $n = \sum_{j=1}^{N} \delta_j$ and $E\left[n\right] = \sum_{j=1}^{N} L_j^*/T$. Note that in practice there are many ways in which assumptions could be violated. The implementation of the roving survey may be restricted to a period less that T, *e.g.* boat patrols restricted to daylight hours, while nets are set and picked at all times of the day; or the timing of an episode may depend upon the length of an episode, *e.g.* fishers who will be fishing only for short periods, may tend to fish only in the morning hours. Deviations from the assumptions are apt to lead to bias in the estimates.

Implementation of the roving boat patrol survey will affect its results. In order to determine C_j, the net must be removed from the water, ending the current episode and initiating another, which impacts on the ensuing catch rate. In recognition of this “sampling effect”, n is assumed to be small relative to N and it is assumed that fishing time lost to interview is offset by the higher catch rate resulting from the empty net after interview.

To model the decline in the catch rate, assume that the probability that a fish will encounter a given net is constant over time and location but that the ability of the net to actually catch fish decreases as the net fills to capacity *i.e.* a catch rate saturation model. For example, a simple model for the catch rate of a fishing episode after k fish have been caught is the continuous time Markov process:

$$\lambda_k = \lambda\left(1 - \frac{k}{C_0}\right) \ ; \quad k = 0, 1, ..., C_0 \tag{2.1}$$

where
- λ_k is the catch rate with k fish already caught,
- λ is the catch rate for an empty net ($i.e.\ k = 0$), and
- C_0 is the catch capacity of the net at saturation.

This model is appealing in that the catch rate, λ_k, is dependent on the number of fish already caught and decreases strictly to zero as the net approaches saturation. Standard results (*e.g.* Taylor and Karlin, 1984) can be used to find the *pmf* of $C(l)$, the catch at time l:

$$P(C(l)) = \begin{pmatrix} C_0 \\ C(l) \end{pmatrix} \left(1 - e^{-\frac{\lambda l}{C_0}}\right)^{C(l)} \left(e^{-\frac{\lambda l}{C_0}}\right)^{C_0 - C(l)} \tag{2.2}$$

which is binomial in $\left(1 - e^{-\frac{\lambda l}{C_0}}\right)$ and C_0. The expected catch after elapsed time l is then $E[C(l)] = C_0\left(1 - e^{-\frac{\lambda l}{C_0}}\right)$ and $Var(C(l)) = C_0\left(1 - e^{-\frac{\lambda l}{C_0}}\right)e^{-\frac{\lambda l}{C_0}} = e^{-\frac{\lambda l}{C_0}}E[C(l)]$. Also, the model implies that $E[C(l)]$ is an increasing function of l and $E[C(0)] = 0$. Note that for fishing episode j: $P(C_j | L_j, \delta = 1)$ is binomial in C_0 and $\left(1 - e^{-\frac{\lambda L_j}{C_0}}\right)$; $P\left(C_j^* \middle| L_j^*, \delta = 1\right)$ is binomial in C_0 and $\left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)$; and $P\left(C_j | L_j, C_j^*, L_j^*, \delta = 1\right)$ is binomial in C_j^* and $\left(1 - e^{-\frac{\lambda L_j}{C_0}}\right) \Big/ \left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)$. These results are used extensively in developing the results which follow. (For more detail see Appendix A.)

It is of interest to compare the model proposed in this chapter with the natural generalization of the homogeneous Poisson process based model for catch used by Hoenig et al. (1997). In a non-homogeneous Poisson process, $\lambda(l)$, the catch rate at time l of the j^{th} fishing episode, can be expressed as $\dfrac{\partial C}{\partial l} = \lambda\left(1 - \dfrac{C(l)}{C_0}\right)$ where $C(l)$ is the theoretical value of catch at time l, λ is the initial catch rate, and C_0 is the limit of catch as l becomes infinitely large. Solving the differential equation gives $E[\lambda(l)] = C_0\left(1 - e^{-\frac{\lambda l}{C_0}}\right)$, the same formula for the theoretical catch at time l as the Markov model. (For more detail see Appendix B.)

2.4 Total Effort Estimation

Total effort is to be estimated independently of the catch rate. Appropriate units are fishing hours (or minutes) in order that they match those in the denominator of the expectation of the catch rate estimator.

The physical characteristics of the fishing resource that is being considered, *i.e.* a river in this case, in large part determine the method of measuring effort. By selecting an aerial count of active nets to gauge the effort, it can be argued that the count should be viewed as "instantaneous" (see Pollock et al., 1994). If the instantaneous count, I, is made at a randomly chosen time during the period T, then the expansion $\hat{E} = I \times T$ will give an unbiased estimator for E, the total effort during time period T. Also, $\hat{E}$ is then expressed in units of fishing hours (or minutes) as required.

2.5 The Catch Rate Estimator

As noted earlier, it is natural to look to some form of ratioing of the C_i and L_i from the sampled episodes for an estimator of the catch rate R, but whether it is better to use a ratio of means or a mean of ratios depends on the design of the survey. Since C^* is to be estimated using $\hat{E}^* \times \hat{R}$ from independently conducted surveys, an appropriate estimator for R will be one for which $E\left[\hat{R}\right] = \sum_{j=1}^{N} C_j^* \Big/ \sum_{j=1}^{N} L_j^* = C^* / E^*$.

Since it is the total catch, after all episodes have been completed, that is to be estimated, it is desirable to construct the catch rate estimator using information from completed fishing episodes. But, since the roving surveys deal only with incomplete fishing episodes, completed catch and length of time fished information must be model based. With the distribution of $L_j | \, \delta = 1$ uniform over $\left[0\,,\, L_j^*\right]$, $2L_j$ is an unbiased estimator for L_j^* and therefore an obvious choice for the estimator of the length of time fished for a completed episode. It naturally follows that some approximation of catch at $2L_j$ should be used as the estimator for the catch of a completed episode. To incorporate C_i, the information obtained from the survey, note that

$$\frac{E\left[C\left(2L_i\right)\right]}{E\left[C\left(L_i\right)\right]} = \frac{C_0\left(1-e^{-\frac{2\lambda L_i}{C_0}}\right)}{C_0\left(1-e^{-\frac{\lambda L_i}{C_0}}\right)} = \left(1+e^{-\frac{\lambda L_i}{C_0}}\right).$$

This "nonlinear doubler" for catch then leads to a catch rate estimator which is formed by the ratio of $\left(1+e^{-\frac{\lambda L_i}{C_0}}\right) C_i$ to $2L_i$.

2.5.1 The Ratio of Means Estimator

A ratio of means estimator is the usual ratio estimator in sampling (see Cochran, 1977). Under a roving design and the proposed model, the ratio of means estimator is given by

$$\begin{aligned}\hat{R}_{rom} &= \frac{\sum_{i=1}^{n}\left(1+e^{-\frac{\lambda L_i}{C_0}}\right) C_i}{\sum_{i=1}^{n} 2L_i} \\ &= \frac{\sum_{j=1}^{N}\delta_j\left(1+e^{-\frac{\lambda L_j}{C_0}}\right) C_j}{\sum_{j=1}^{N} 2\delta_j L_j}\end{aligned}$$

where

λ and C_0 are assumed to be known,
δ_j, C_j and L_j are random variables,
$\sum_{i=1}^{n} L_i > 0$,
$P(\delta_j = 1) = \frac{L_j^*}{T}$, and
$L_j|\,\delta_j = 1$ is distributed $Unif\left[0\,,\,L_j^*\right]$.

Assuming that the sample size is large enough that the expectations of the ratio can be approximated with the ratio of expectations and given C_j^* and L_j^*,

$$E\left[\hat{R}_{rom}\right] \approx \frac{E\left[\sum_{j=1}^{N}\delta_j\left(1+e^{-\frac{\lambda L_j}{C_0}}\right) C_j\right]}{E\left[\sum_{j=1}^{N} 2\delta_j L_j\right]}.$$

Evaluating the numerator where $E_{L_j}\left[1-e^{-\frac{2\lambda L_j}{C_0}}\right]$, the expectation of the random variable L_j, is approximated using a first order Taylor expansion about $E\left[L_j\right]=L_j^*/2$:

$$
\begin{aligned}
E_{L_j,C_j,\delta_j}\left[\sum_{j=1}^{N}\delta_j\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)C_j\right] &= \sum_{j=1}^{N}E_{L_j,C_j}\left[0\cdot\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)C_j\middle|\,\delta_j=0\right]P(\delta_j=0)\\
&+ \sum_{j=1}^{N}E_{L_j,C_j}\left[1\cdot\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)C_j\middle|\,\delta_j=1\right]P(\delta_j=1)\\
&= 0+\sum_{j=1}^{N}E_{L_j,C_j}\left[\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)C_j\right]P(\delta_j=1)\\
&= \sum_{j=1}^{N}E_{L_j}\left[\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)E_{C_j}\left[C_j|\,L_j,C_j^*,L_j^*\right]\right]\frac{L_j^*}{T}\\
&= \sum_{j=1}^{N}E_{L_j}\left[\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)C_j^*\left(\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)\right]\frac{L_j^*}{T}\\
&= \sum_{j=1}^{N}E_{L_j}\left[\frac{\left(1-e^{-\frac{2\lambda L_j}{C_0}}\right)}{\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)}C_j^*\frac{\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\right]\frac{L_j^*}{T}\\
&= \sum_{j=1}^{N}E_{L_j}\left[1-e^{-\frac{2\lambda L_j}{C_0}}\right]\frac{C_j^*}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{L_j^*}{T}\\
&\approx \sum_{j=1}^{N}\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)\frac{C_j^*}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{L_j^*}{T}\\
&= \sum_{j=1}^{N}C_j^*L_j^*/\,T.
\end{aligned}
$$

Evaluating the denominator:

$$\begin{aligned}
E_{L_j,C_j,\delta_j}\left[\sum_{j=1}^{N} 2\delta_j L_j\right] &= \sum_{j=1}^{N} E_{L_j}\left[0\cdot L_j|\,\delta_j=0\right]P\left(\delta_j=0\right) \\
&+ \sum_{j=1}^{N} E_{L_j}\left[1\cdot L_j|\,\delta_j=1\right]P\left(\delta_j=1\right) \\
&= 0+\sum_{j=1}^{N} E_{L_j}\left[L_j\right]P\left(\delta_j=1\right) \\
&= \sum_{j=1}^{N} E_{L_j}\left[L_j\right]\frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} \frac{L_j^*}{2}\frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} \left(L_j^*\right)^2 \Big/ 2T.
\end{aligned}$$

Therefore

$$\begin{aligned}
E\left[\hat{R}_{rom}\right] &\approx \frac{\sum_{j=1}^{N} C_j^* L_j^* \Big/ T}{\sum_{j=1}^{N} \left(L_j^*\right)^2 \Big/ 2T} \\
&= \frac{2\sum_{j=1}^{N} C_j^* L_j^*}{\sum_{j=1}^{N} \left(L_j^*\right)^2}
\end{aligned}$$

which is not of the appropriate form and therefore is not recommended for use with the current model.

2.5.2 The Mean of Ratios Estimator

The mean of ratios estimator under the current model and roving design is given by

$$\hat{R}_{mor} = \frac{1}{n}\sum_{i=1}^{n}\left(1+e^{-\frac{\lambda L_i}{C_0}}\right)\frac{C_i}{2L_i} = \frac{\sum\limits_{j=1}^{N}\delta_j\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)\frac{C_j}{2L_j}}{\sum\limits_{j=1}^{N}\delta_j} \tag{2.3}$$

where

λ and C_0 are assumed to be known,
δ_j, C_j and L_j are random variables,
$L_i > 0$ for all i,
$P_r\{\delta_j = 1\} = \frac{L_j^*}{T}$, and
$L_j|\,\delta_j = 1$ is distributed $\mathit{Unif}\left[0\,,\,L_j^*\right]$.

Assuming that the sample size is large enough that the expectations of the ratio can be approximated with the ratio of expectations and given C_j^* and L_j^*,

$$E\left[\hat{R}_{mor}\right] \approx \frac{E\left[\sum\limits_{j=1}^{N}\delta_j\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)\frac{C_j}{2L_j}\right]}{E\left[\sum\limits_{j=1}^{N}\delta_j\right]}.$$

Evaluating the numerator where a first order Taylor expansion about $E[L_j] = L_j^*/2$ is used to approximate $E\left[\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{L_j}\right]$,

$$\begin{aligned}
E_{L_j,C_j,\delta_j}\left[\sum_{j=1}^{N}\delta_j\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)\frac{C_j}{2L_j}\right] &= \sum_{j=1}^{N}E_{L_j,C_j}\left[0\cdot\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)\frac{C_j}{2L_j}\middle|\,\delta_j=0\right]P(\delta_j=0)\\
&+ \sum_{j=1}^{N}E_{L_j,C_j}\left[1\cdot\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)\frac{C_j}{2L_j}\middle|\,\delta_j=1\right]P(\delta_j=1)\\
&= 0+\sum_{j=1}^{N}E_{L_j,C_j}\left[\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)\frac{C_j}{2L_j}\right]P(\delta_j=1)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^{N} E_{L_j}\left[\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)\frac{E_{C_j}\left[C_j \,|\, L_j, C_j^*, L_j^*\right]}{2L_j}\right]\frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} E_{L_j}\left[\frac{\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)}{2L_j}C_j^*\left(\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)\right]\frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} E_{L_j}\left[\frac{C_j^*}{2L_j}\frac{\left(1-e^{-\frac{2\lambda L_j}{C_0}}\right)}{\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)}\frac{\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\right]\frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} E_{L_j}\left[\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{L_j}\right]\frac{C_j^*}{2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{L_j^*}{T} \\
&\approx \sum_{j=1}^{N} \frac{2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}{L_j^*}\frac{C_j^*}{2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} C_j^* \big/ T.
\end{aligned}$$

Evaluating the denominator,

$$\begin{aligned}
E_{\delta_j}\left[\sum_{j=1}^{N}\delta_j\right] &= \sum_{j=1}^{N} E_{\delta_j}\left[\delta_j\right] \\
&= \sum_{j=1}^{N}\left\{0\cdot P\left(\delta_j=0\right)+1\cdot P\left(\delta_j=1\right)\right\} \\
&= \sum_{j=1}^{N} L_j^* \big/ T.
\end{aligned}$$

Therefore

$$E\left[\hat{R}_{mor}\right] \approx \frac{\sum_{j=1}^{N} C_j^* \Big/ T}{\sum_{j=1}^{N} L_j^* \Big/ T} = \frac{\sum_{j=1}^{N} C_j^*}{\sum_{j=1}^{N} L_j^*} = \frac{C^*}{E^*}$$

which is unbiased, to the extent of the approximation, and of the required form and therefore appropriate for use as an estimator in the framework

$$total\ catch = (total\ effort) \times (catch\ rate)$$

under the current model. Unless otherwise specified, $\hat{R}$ and catch rate estimator refer to this mean of ratios estimator.

By modelling catch as a homogeneous Poisson process, Hoenig et al. (1997) obtained similar results. In their model, catch is linear in time fished. Thus, if C_i were the observed catch at time L_i, the expected catch at time $2L_i$ would be $2C_i$. Using this, they found that $\hat{R}_{rom} = \frac{\sum_{i=1}^{n} 2C_i}{\sum_{i=1}^{n} 2L_i} = \frac{\sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} L_i}$, a ratio of means, was an appropriate estimator for catch rate if the survey was an access design while $\hat{R}_{mor} = \frac{1}{n}\sum_{i=1}^{n} \frac{2C_i}{2L_i} = \frac{1}{n}\sum_{i=1}^{n} \frac{C_i}{L_i}$, a mean of ratios, was appropriate if the survey was a roving design. To see the estimator of Hoenig et al. (1997) as a special case of $\hat{R}$, the estimator in the proposed model, let $C_0 \to \infty$, *i.e.* no net saturation. Then $\left(1 + e^{-\frac{\lambda L_i}{C_0}}\right) \frac{C_i}{2L_i} \to \frac{2C_i}{2L_i}$. Therefore, as $C_0 \to \infty$, $\hat{R} \to \frac{1}{n}\sum_{i=1}^{n} \frac{2C_i}{2L_i}$ showing that $\hat{R}$ reduces to the mean of ratios estimator used by Hoenig et al. (1997). Also note that in the proposed model, $\lambda_k = \lambda\left(1 - \frac{k}{C_0}\right)$ where λ is the initial catch rate and also the constant catch rate used by Hoenig et al. (1997). Therefore as $C_0 \to \infty$, λ_k approaches λ and both models employ the same catch rate. Further details are given in appendix B.

2.6 Variance of $\hat{R}$ (λ and C_0 known)

Contributing to the variance of $\hat{R}$ are uncertainties in C_i, L_i and n as well as variability associated with the estimation of λ and C_0. Also, since $Var\left(C_j\right)$ is a function of C_j^*, the variability in C_j^* is also a contributor. To find a form for the variance of $\hat{R}$, λ and C_0 are initially assumed to be known.

2.6.1 The Form of $Var\left(\hat{R} \mid \lambda, C_0\right)$

The appropriate estimator for catch rate was shown to be $\hat{R} = \frac{1}{n}\sum_{i=1}^{n}\left(1+e^{-\frac{\lambda L_i}{C_0}}\right)\frac{C_i}{2L_i}$ which, by writing $Z_i = \left(1+e^{-\frac{\lambda L_i}{C_0}}\right)\frac{C_i}{2L_i}$, notationally simplifies to $\hat{R} = \frac{1}{n}\sum_{i=1}^{n} Z_i$, where now n and Z_i are random variables. Then, assuming λ and C_0 are known and conditioning on s, the sample selected (*i.e.* a particular set of intercept times),

$$Var\left(\hat{R} \mid \lambda, C_0\right) = E_s\left[Var\left(\bar{Z}\middle| s\right)\right] + Var_s\left(E\left[\bar{Z}\middle| s\right]\right)\cdot \tag{2.4}$$

Expansion of the first term using a Taylor series approximation gives

$$\begin{aligned} E_s\left[Var\left(\bar{Z} \mid s\right)\right] &= E_s\left[\frac{1}{n^2}Var\left(\sum_{j=1}^{N}\delta_j Z_j\right) \mid s\right] \\ &\approx \frac{\sum_{j=1}^{N} E_s\left[\delta_j{}^2\right] Var\left(Z_j\middle| s\right)}{\left(E_s[n]\right)^2} \\ &= \frac{\sum_{j=1}^{N} \frac{L_j^*}{T} Var\left(Z_j\middle| s\right)}{\left(\sum_{j=1}^{N}\frac{L_j^*}{T}\right)^2} \end{aligned}$$

since δ_j is Bernoulli in L_j^*/T. Then for any j, given that $\delta_j = 1$,

$$\begin{aligned} Var\left(Z_j\middle| C_j^*, L_j^*\right) &= E_{L_j}\left[Var\left(Z_j\middle| L_j, C_j^*, L_j^*\right)\right] + Var_{L_j}\left(E\left[Z_j\middle| L_j, C_j^*, L_j^*\right]\right) \\ &= E\left[\left(\frac{1+e^{-\frac{\lambda L_j}{C_0}}}{2L_j}\right)^2 C_j^*\left(\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)\left(1-\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)\right] \\ &\quad +Var\left(\left(\frac{1+e^{-\frac{\lambda L_j}{C_0}}}{2L_j}\right)C_j^*\left(\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)\right) \end{aligned}$$

$$
\begin{aligned}
= \quad & E\left[\frac{C_j^*}{(2L_j)^2}\left(\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j}{C_0}}}\right)^2\left(\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)\right] \\
& -E\left[\frac{C_j^*}{(2L_j)^2}\left(\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j}{C_0}}}\right)^2\left(\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)^2\right] \\
& +Var\left(\frac{C_j^*}{2L_j}\left(\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j}{C_0}}}\right)\left(\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)\right) \\
= \quad & E\left[\frac{C_j^*}{(2L_j)^2}\cdot\frac{\left(1-e^{-\frac{2\lambda L_j}{C_0}}\right)\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)}{\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)^2}\cdot\frac{\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\right] \\
& -E\left[\frac{C_j^*}{(2L_j)^2}\cdot\frac{1}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)^2}\cdot\frac{\left(1-e^{-\frac{2\lambda L_j}{C_0}}\right)^2}{1}\right] \\
& +Var\left(\frac{C_j^*}{2L_j}\cdot\frac{1}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\cdot\frac{\left(1-e^{-\frac{2\lambda L_j}{C_0}}\right)}{1}\right) \\
= \quad & \frac{C_j^*}{4\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\cdot E\left[\frac{\left(1-e^{-\frac{2\lambda L_j}{C_0}}\right)\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)}{L_j^2}\right] \\
& -\frac{C_j^*}{4\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)^2}\cdot E\left[\left(\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{L_j}\right)^2\right] \\
& +\frac{\left(C_j^*\right)^2}{4\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)^2}\cdot Var\left(\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{L_j}\right).
\end{aligned}
$$

Note: Expressed this way, it is clear that $Var\left(\hat{R}\right)$ does not exist unless L_j is bounded away from zero. It is also clear that the greater this bound, the more stable is $Var\left(\hat{R}\right)$. In theory, L_j is a continuous random variable and can approach zero, however, in practice, whether as part of sampling protocol or by use of discrete time units, a minimum length of time to interview, L', will exist. Accordingly, $E\left[L_j\right] = \left(L_j^* + L'\right)/2$ and $Var\left(L_j\right) = \left(L_j^* - L'\right)^2/12$ and the probability of sample inclusion for any episode is $\left(L_j^* - L'\right)/T$. In the development of expressions for both the expectation and variance of Z_j, and hence $\hat{R}$, L' will be taken to be sufficiently small that its effects are negligible compared to the order of the approximations already made and therefore dropped from the notation allowing for some cancelations and considerable simplification. In the simulations which follow, however, L' is carried. A minimum time to interview was also used by Hoenig et al. (1997) and Pollock et al. (1997) in their work with a sports fishery under the constant catch rate assumption.

Now, with L' sufficiently small and using Taylor series approximation about $E\left[L_j\right] = L_j^*/2$,

$$
\begin{aligned}
Var\left(Z_j \middle| C_j^*, L_j^*\right) \approx\ & \frac{C_j^*}{4\left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)} \cdot \frac{\left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)\left(1 + e^{-\frac{\lambda L_j^*/2}{C_0}}\right)}{\left(L_j^*/2\right)^2} \\
& - \frac{C_j^*}{4\left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)^2} \cdot \left(\frac{\left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)}{L_j^*/2}\right)^2 \\
& + \frac{\left(C_j^*\right)^2}{4\left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)^2} \cdot \left(\frac{\frac{\lambda L_j^*}{C_0} e^{-\frac{\lambda L_j^*}{C_0}} - \left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)}{\left(L_j^*/2\right)^2}\right)^2 \cdot \frac{L_j^{*2}}{12} \\
=\ & \frac{C_j^*}{\left(L_j^*\right)^2} \cdot \left(1 + e^{-\frac{\lambda L_j^*/2}{C_0}}\right) - \frac{C_j^*}{\left(L_j^*\right)^2} \\
& + \frac{\left(C_j^*\right)^2}{3\left(L_j^*\right)^2\left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)^2}\left\{\frac{\lambda L_j^*}{C_0} e^{-\frac{\lambda L_j^*}{C_0}} - \left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)\right\}^2.
\end{aligned}
$$

To express this variance as a function of L_j^* only, first note that, given $\delta_j = 1$,

$$\begin{aligned}
E\left[\left. Z_j \right| C_j^*, L_j^*\right] &= E_{L_j}\left[E\left[\left. Z_j \right| L_j, C_j^*, L_j^*\right]\right] \\
&= E_{L_j}\left[\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)\frac{C_j^*}{2L_j}\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right] \\
&= E_{L_j}\left[\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j}{C_0}}}\frac{C_j^*}{2L_j}\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right] \\
&= E_{L_j}\left[\frac{C_j^*}{2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{\left(1-e^{-\frac{2\lambda L_j}{C_0}}\right)}{L_j}\right]
\end{aligned}$$

and using a first order Taylor approximation about $E\left[L_j\right] = L_j^*/2$, $E\left[\left. Z_j \right| C_j^*, L_j^*\right] \approx C_j^*/L_j^*$. Therefore, as a function of L_j^*,

$$\begin{aligned}
E\left[Z_j\right] &\approx \frac{1}{L_j^*}E\left[\left. C_j^* \right| L_j^*\right] \\
&= \frac{1}{L_j^*}C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right).
\end{aligned}$$

Now, with $\left. C_j^* \right| L_j^*$ distributed binomial in C_0 and $\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)$, as a function of L_j^* only

$$\begin{aligned}
Var\left(Z_j\right) &\approx E_{C_j^*}\left[Var\left(\left. Z_j \right| C_j^*, L_j^*\right)\right] + Var_{C_j^*}\left(E\left[\left. Z_j \right| C_j^*, L_j^*\right]\right) \\
&= \frac{C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}{\left(L_j^*\right)^2}\left(1+e^{-\frac{\lambda L_j}{C_0}}\right) \\
&\quad -\frac{C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}{\left(L_j^*\right)^2}
\end{aligned}$$

$$+\frac{C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)\left(e^{-\frac{\lambda L_j^*}{C_0}}\right)}{3\left(L_j^*\right)^2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)^2}\left\{\frac{\lambda L_j^*}{C_0}e^{-\frac{\lambda L_j^*}{C_0}}-\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)\right\}^2$$

$$+\frac{(C_0)^2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)^2}{3\left(L_j^*\right)^2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)^2}\left\{\frac{\lambda L_j^*}{C_0}e^{-\frac{\lambda L_j^*}{C_0}}-\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)\right\}^2$$

$$+\frac{C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)\left(e^{-\frac{\lambda L_j^*}{C_0}}\right)}{\left(L_j^*\right)^2}. \tag{2.5}$$

It then follows that for the first term in equation 2.4, as a function of L_j^*,

$$E_s\left[Var\left(\bar{Z}\middle|\, s\right)\right] = \frac{1}{\left(\sum_{j=1}^{N}\frac{L_j^*}{T}\right)^2}\left\{\sum_{j=1}^{N}\left(V_j+W_j\right)\right\} \tag{2.6}$$

where

$$V_j = \frac{C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}{TL_j^*}\left\{e^{\frac{-\lambda L_j^*/2}{C_0}}+e^{-\frac{\lambda L_j^*}{C_0}}\right\}$$

and

$$W_j = \frac{1}{3TL_j^*}\left\{1+\frac{e^{-\frac{\lambda L_j^*}{C_0}}}{C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\right\}\left\{\lambda L_j^* e^{-\frac{\lambda L_j^*}{C_0}}-C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)\right\}^2.$$

Next,to express the second term in equation 2.4 as a function of L_j^* only first recall that $E\left[Z_j|\, C_j^*, L_j^*\right] = C_j^*/L_j^*$. It is then possible to write

$$\begin{aligned} Var_s\left(E\left[\bar{Z}|\, s\right]\right) &= Var_s\left(\frac{1}{n}E_s\left[\sum_{j=1}^{N}\delta_j Z_j\right]|\, s\right) \\ &\approx Var_s\left(\frac{1}{n}\sum_{j=1}^{N}\delta_j\frac{C_j^*}{L_j^*}\right) \\ &= Var_s\left(\frac{\sum_{j=1}^{N}\delta_j\frac{C_j^*}{L_j^*}}{\sum_{j=1}^{N}\delta_j}\right). \end{aligned}$$

Finally, taking Taylor series approximations about each $E\left[\delta_j\right] = L_j^*/T$,

$$\begin{aligned} Var_s\left(E\left[\bar{Z}|\, s\right]\right) &\approx \sum_{j=1}^{N} Var\left(\delta_j\right)\left[\frac{\frac{C_j^*}{L_j^*}}{\sum_{j=1}^{N}\frac{L_j^*}{T}} - \frac{\sum_{j=1}^{N}\frac{L_j^*}{T}\cdot\frac{C_j^*}{L_j^*}}{\left(\sum_{j=1}^{N}\frac{L_j^*}{T}\right)^2}\right]^2 \\ &= \frac{1}{\left(\sum_{j=1}^{N}\frac{L_j^*}{T}\right)^2}\sum_{j=1}^{N} Var\left(\delta_j\right)\left[\frac{C_j^*}{L_j^*} - \frac{\sum_{j=1}^{N}C_j^*}{\sum_{j=1}^{N}L_j^*}\right]^2 \\ &= \frac{1}{\left(\sum_{j=1}^{N}\frac{L_j^*}{T}\right)^2}\sum_{j=1}^{N}\left(\frac{L_j^*}{T}\right)\left(1-\frac{L_j^*}{T}\right)\left[\frac{C_j^*}{L_j^*} - \frac{\sum_{j=1}^{N}C_j^*}{\sum_{j=1}^{N}L_j^*}\right]^2 \end{aligned}$$

and as a function of L_j^* only where C_j^* is replaced by $E\left[C_j^*\middle|\, L_j^*, \delta = 1\right]$,

$$Var_s\left(E\left[\bar{Z}|\, s\right]\right) \approx$$
$$\frac{1}{\left(\sum_{j=1}^{N}\frac{L_j^*}{T}\right)^2}\sum_{j=1}^{N}\left(\frac{L_j^*}{T}\right)\left(1-\frac{L_j^*}{T}\right)\left[\frac{C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}{L_j^*} - \frac{\sum_{j=1}^{N}C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}{\sum_{j=1}^{N}L_j^*}\right]^2. \quad (2.7)$$

$Var\left(\hat{R} \mid \lambda, C_0\right)$ is now found by combining equations 2.6 and 2.7.

In section 2.5.2 it was shown that the estimator for catch per effort developed by Hoenig et al. (1997) was the same as that proposed in this chapter under the limiting condition $C_0 \to \infty$. It was also shown that when $C_0 \to \infty$, λ is the same constant catch rate for both models. In addition, developing the variance formula for the estimator of catch per effort for either model will start with equation 2.4 but with Z_i equal to $\left(1+e^{-\frac{\lambda L_i}{C_0}}\right)\dfrac{C_i}{2L_i}$ under the proposed model or equal to C_i/L_i under the model used by Hoenig et al. (1997). Therefore, to show that the limiting value of $Var\left(\hat{R} \mid \lambda, C_0\right)$ for $C_0 \to \infty$ under the proposed model approaches $Var\left(\hat{R}_{Poiss}\right)$ under the Hoenig et al. (1997) model, it is sufficient to show that $\underset{C_0\to\infty}{lim} E\left[Z\right] = E\left[Z_{Poiss}\right]$ and $\underset{C_0\to\infty}{lim} Var\left(Z\right) = Var\left(Z_{Poiss}\right)$ for any episode.

To do this, first recall that for the proposed model

$$E\left[Z_j\right] = \frac{1}{L_j^*}E\left[C_j^* \middle| L_j^*\right]$$

and rewrite equation 2.5 as

$$\begin{aligned}
Var\left(Z_j\right) \approx\ & \frac{E\left[C_j^* \middle| L_j^*\right]}{\left(L_j^*\right)^2}\left(1+e^{\frac{-\lambda L_j^*/2}{C_0}}\right) \\
& -\frac{E\left[C_j^* \middle| L_j^*\right]}{\left(L_j^*\right)^2} \\
& +\frac{E\left[C_j^* \middle| L_j^*\right]\left(e^{-\frac{\lambda L_j^*}{C_0}}\right)}{3\left(L_j^*\right)^2}\left\{\frac{\frac{\lambda L_j^*}{C_0}e^{-\frac{\lambda L_j^*}{C_0}} - \left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\right\}^2 \\
& +\frac{\left(E\left[C_j^* \middle| L_j^*\right]\right)^2}{3\left(L_j^*\right)^2}\left\{\frac{\frac{\lambda L_j^*}{C_0}e^{-\frac{\lambda L_j^*}{C_0}} - \left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\right\}^2 \\
& +\frac{E\left[C_j^* \middle| L_j^*\right]\left(e^{-\frac{\lambda L_j^*}{C_0}}\right)}{\left(L_j^*\right)^2}.
\end{aligned}$$

Letting $C_0 \to \infty$ at this point results in indeterminant forms for the third and fourth terms but, by applying l'Hopital's rule,

$$\lim_{C_0\to\infty} \frac{\frac{d}{dC_0}\left\{\frac{\lambda L_j^*}{C_0}e^{-\frac{\lambda L_j^*}{C_0}} - \left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)\right\}^2}{\frac{d}{dC_0}\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)^2}$$

$$= \lim_{C_0\to\infty} \frac{-\lambda L_j^*\left\{\frac{\lambda L_j^* e^{-\frac{\lambda L_j^*}{C_0}}}{C_0} - \left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)\right\}}{C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}$$

$$= \lim_{C_0\to\infty} \frac{-\lambda L_j^*\left\{\frac{\lambda L_j^* e^{-\frac{\lambda L_j^*}{C_0}}}{C_0} - \left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)\right\}}{E\left[C_j^*\middle| L_j^*\right]}$$

$$= 0$$

since $E\left[C_j^*\middle| L_j^*\right]$ is finite because $L_j^* < T$ is finite. Therefore,

$$\lim_{C_0\to\infty} E\left[Z_j\right] = \frac{\lim_{C_0\to\infty} E\left[C_j^*\middle| L_j^*\right]}{L_j^*}$$

and

$$\lim_{C_0\to\infty} Var\left(Z_j\right) = \frac{2\lim_{C_0\to\infty} E\left[C_j^*\middle| L_j^*\right]}{\left(L_j^*\right)^2}.$$

Referring to appendix B, the limiting distribution of $C_j^*\big| L_j^*, \delta = 1$ as $C_0 \to \infty$ is Poisson with parameter $\mu = \lambda L_j^*$ where λ is now the constant catch rate over all L_j^* as used by Hoenig et al. (1997). Thus,

$$\lim_{C_0\to\infty} E\left[\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)\frac{C_j}{2L_j}\right] = \frac{\lambda L_j^*}{L_j^*} = \lambda$$

and

$$\lim_{C_0\to\infty} Var\left(\left(1+e^{-\frac{\lambda L_j}{C_0}}\right)\frac{C_j}{2L_j}\right) = \frac{2\lambda L_j^*}{\left(L_j^*\right)^2}.$$

As noted previously, constructing $Var\left(\hat{R}_{Poiss}\right)$ for the Hoenig et al. (1997) model starts with equation 2.4 but now using $Z_j = C_j/L_j$. Also, since under their model, $P(C_j \mid L_j, \delta_j = 1)$ is Poisson with parameter λL_j, it follows that $P(C_j \mid L_j, C_j^*, L_j^*, \delta_j = 1)$ is binomial with parameters C_j^* and L_j/L_j^*. Hence

$$\begin{aligned} E\left[\left.\frac{C_j}{L_j}\right| C_j^*, L_j^*\right] &= E_{L_j}\left[E\left[\left.\frac{C_j}{L_j}\right| L_j, C_j^*, L_j^*\right]\right] \\ &= E_{L_j}\left[\frac{1}{L_j}E\left[C_j | L_j, C_j^*, L_j^*\right]\right] \\ &= E_{L_j}\left[\frac{1}{L_j}C_j^*\frac{L_j}{L_j^*}\right] \\ &= E_{L_j}\left[\frac{C_j^*}{L_j^*}\right] \\ &= \frac{C_j^*}{L_j^*}, \end{aligned}$$

and as a function of L_j^*,

$$\begin{aligned} E\left[\frac{C_j}{L_j}\right] &= E\left[\left.\frac{C_j^*}{L_j^*}\right| L_j^*\right] \\ &= \frac{\lambda L_j^*}{L_j^*} \\ &= \lambda \end{aligned}$$

which is equal to $\underset{C_0\to\infty}{lim} E[Z]$ under the proposed model. It also follows that

$$\begin{aligned} Var\left(\left.\frac{C_j}{L_j}\right| C_j^*, L_j^*\right) &= E_{L_j}\left[Var\left(\left.\frac{C_j}{L_j}\right| L_j, C_j^*, L_j^*\right)\right] + Var_{L_j}\left(E\left[\left.\frac{C_j}{L_j}\right| L_j, C_j^*, L_j^*\right]\right) \\ &= E_{L_j}\left[\frac{1}{L_j^2}C_j^*\left(\frac{L_j}{L_j^*}\right)\left(1-\frac{L_j}{L_j^*}\right)\right] + Var\left(\frac{1}{L_j}C_j^*\frac{L_j}{L_j^*}\right) \\ &= \frac{C_j^*}{L_j^*}E\left[\frac{1}{L_j}\right] - \frac{C_j^*}{\left(L_j^*\right)^2} + 0 \\ &\approx \frac{2C_j^*}{\left(L_j^*\right)^2} - \frac{C_j^*}{\left(L_j^*\right)^2} \\ &= \frac{C_j^*}{\left(L_j^*\right)^2} \end{aligned}$$

with the use of first order Taylor approximations. Therefore, as a function of L_j^*,

$$\begin{aligned} Var\left(\frac{C_j}{L_j}\right) &\approx E_{C_j^*}\left[Var\left(\frac{C_j}{L_j}\middle| C_j^*, L_j^*\right)\right] + Var_{C_j^*}\left(E\left[\frac{C_j}{L_j}\middle| C_j^*, L_j^*\right]\right) \\ &= \frac{1}{\left(L_j^*\right)^2} E_{C_j^*}\left[C_j^* \middle| L_j^*\right] + \frac{1}{\left(L_j^*\right)^2} Var_{C_j^*}\left(C_j^* \middle| L_j^*\right) \\ &= \frac{1}{\left(L_j^*\right)^2}\lambda L_j^* + \frac{1}{\left(L_j^*\right)^2}\lambda L_j^* \\ &= \frac{2\lambda L_j^*}{\left(L_j^*\right)^2}, \end{aligned}$$

again equal to the limiting result under the proposed model. That is, the expectations and variances in the Hoenig et al. (1997) model are equal to those in the limiting case of the proposed model. Therefore, since variance construction for both models proceeds from equation 2.4 with identical inputs (in the limiting case), it follows that $\lim_{C_0\to\infty} Var(\hat{R} \mid \lambda, C_0) = Var\left(\hat{R}_{Poiss}\right)$.

2.6.2 The Asymptotic Variance of $\hat{R}$

For an unbiased estimator formed by some function of the observations divided by the sample size, expressing a bound on its variance as $\frac{1}{n^2}$ times some finite quantity is a useful way to show consistency. For $\hat{R} = \frac{1}{n}\sum_{i=1}^{n}\left(1 + e^{-\frac{\lambda L_i}{C_0}}\right)\frac{C_i}{2L_i}$ such an approach is not possible since n is a random variable and there is no guarantee that it will not persist in assuming small values. To overcome this difficulty, it is necessary to find a surrogate for n.

To do this, first note that with L_j uniformly distributed over $[0\,,\, L_j^*]$ but bounded away from zero by ϵ,

$$E_{L_j}\left[\frac{1}{L_j^2}\right] = \frac{1}{L_j^* - \epsilon}\int_{x=1/L_j^*}^{1/\epsilon} dx = \frac{1}{L_j^* - \epsilon}\left(\frac{1}{\epsilon} - \frac{1}{L_j^*}\right)$$

is bounded. Then, with $Z_i = \left(1 + e^{-\frac{\lambda L_i}{C_0}}\right)\dfrac{C_i}{2L_i} > 0$ and C_0, the net capacity finite, $0 \leq$

$Z_j \leq \frac{2C_j}{2L_j} \leq \frac{2C_j^*}{2L_j} \leq \infty$ making $E\left[\frac{C_j^*}{L_j}\right]$ finite. It now follows that

$$
\begin{aligned}
Var\left(\hat{R} \mid \lambda, C_0\right) &= E_s\left[Var\left(\bar{Z}\middle| s\right)\right] + Var_s\left(E\left[\bar{Z}\middle| s\right]\right) \\
&\approx \frac{\sum_{j=1}^{N} \frac{L_j^*}{T} Var\left(Z_j\middle| s\right)}{(E[n])^2} + \frac{\sum_{j=1}^{N}\left(\frac{L_j^*}{T}\right)\left(1-\frac{L_j^*}{T}\right)\left[\frac{C_j^*}{L_j^*} - \frac{\sum_{j=1}^{N} C_j^*}{\sum_{j=1}^{N} L_j^*}\right]^2}{(E[n])^2} \\
&\leq \frac{1}{(E[n])^2}\sum_{j=1}^{N}\left\{E\left[\left(\frac{C_j^*}{L_j}\right)^2\right] + \left(\frac{C_j^*}{L_j^*}\right)^2\right\}
\end{aligned}
$$

since $L^*/T \leq 1,\ Var(Z) \leq Var(C_j^*/L_j) \leq E[(C_j^*/L_j)^2]$ and $\sum_{j=1}^{N} C_j^* \Big/ \sum_{j=1}^{N} L_j^* \geq 0$. Therefore, with $E[n] = \sum_{j=1}^{N} \frac{L_j^*}{T} = E^*/T$, for some $M > E\left[\left(\frac{C_j^*}{L_j}\right)^2\right] + \left(\frac{C_j^*}{L_j^*}\right)^2$,

$$
\begin{aligned}
Var\left(\hat{R} \mid \lambda, C_0\right) &\leq \frac{NM}{(E^*/T)^2} \\
&\leq \frac{TM}{E^*\left(\sum_{j=1}^{N} L_j^*/N\right)}
\end{aligned}
$$

which approaches 0 as E^* approaches ∞ because T, M and $\sum_{j=1}^{N} L_j^*/N$ are bounded.

With the relationship between n, E^*, and N being $E[n] = E\left[\sum_{j=1}^{N} \delta_j\right] = \sum_{j=1}^{N} \frac{L_j^*}{T} = \frac{E^*}{T}$, provided that nets are added in such a way that the average episode length does not go to 0 and provided that each episode is correlated with only a finite number of other episodes, $\lim_{n\to\infty} Var\left(\hat{R}\right) = 0$ iff $\lim_{N\to\infty} Var\left(\hat{R}\right) = 0$ iff $\lim_{E^*\to\infty} Var\left(\hat{R}\right) = 0$ making $\hat{R}$, to its approximate unbiasedness, consistent. Interpretation of the first proviso is straightforward. The latter demands that as effort is increased, it must be disbursed rather than concentrated at specific locations or times.

2.6.3 A Sample Estimator for $Var\left(\hat{R} \mid \lambda, C_0\right)$

An estimator for $Var\left(\hat{R} \mid \lambda, C_0\right)$ follows naturally from equations 2.6 and 2.7. An estimator for equation 2.6, the variance within a sample due to possible differences in time of intercept, is formed by using $\sum_{i=1}^{n} \frac{L_i^*}{T} Var\left(Z_i | s\right) \Big/ \frac{L_i^*}{T}$ to approximate $\sum_{j=1}^{N} \frac{L_j^*}{T} Var\left(Z_j | s\right)$ while an estimator for equation 2.7, the variance between samples, is formed by using the sample ratio to estimate the population ratio. These estimators should account for L', the minimum time to interview. This then gives, again assuming λ and C_0 known,

$$\widehat{Var}\left(\hat{R} \mid \lambda, C_0\right) = \widehat{term}1 + \widehat{term}2 \tag{2.8}$$

for which, using $\hat{L}_i^* = 2L_i - L'$,

$$\begin{aligned}
\widehat{term}1 &= \frac{1}{n^2}\sum_{i=1}^{n} \frac{C_0}{4} \cdot \frac{\left(1-e^{-\frac{2\lambda L_i}{C_0}}\right)\left(1+e^{-\frac{\lambda L_i}{C_0}}\right)}{L_i^2} \\
&\quad - \frac{C_0}{4\left(1-e^{-\frac{\lambda \hat{L}_i^*}{C_0}}\right)} \cdot \frac{\left(1-e^{-\frac{2\lambda L_i}{C_0}}\right)^2}{L_i^2} \\
&\quad + \frac{C_0 e^{-\frac{\lambda \hat{L}_i^*}{C_0}} + C_0^2\left(1-e^{-\frac{\lambda \hat{L}_i^*}{C_0}}\right)}{4\left(1-e^{-\frac{\lambda \hat{L}_i^*}{C_0}}\right)} \cdot \left(\frac{\frac{2\lambda L_i}{C_0} e^{-\frac{2\lambda L_i}{C_0}} - \left(1-e^{-\frac{2\lambda L_i}{C_0}}\right)}{L_i^2}\right)^2 \cdot \frac{(2L_i - L')^2}{12} \\
&\quad + \frac{C_0\left(1-e^{-\frac{\lambda \hat{L}_i^*}{C_0}}\right) e^{-\frac{\lambda \hat{L}_i^*}{C_0}}}{(2L_i)^2} \cdot \left(\frac{1-e^{-\frac{2\lambda L_i}{C_0}}}{1-e^{-\frac{\lambda \hat{L}_i^*}{C_0}}}\right)^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \left\{ \frac{C_0\left(1-e^{-\frac{2\lambda L_i}{C_0}}\right)}{4L_i^2}\left(e^{-\frac{\lambda L_i}{C_0}} + e^{-\frac{2\lambda L_i}{C_0}}\right) \right. \\
&\quad \left. + \left[1 + \frac{e^{-\frac{\lambda \hat{L}_i^*}{C_0}}}{C_0\left(1-e^{-\frac{\lambda \hat{L}_i^*}{C_0}}\right)}\right]\left[\frac{2\lambda L_i e^{-\frac{2\lambda L_i}{C_0}} - C_0\left(1-e^{-\frac{2\lambda L_i}{C_0}}\right)}{2{L_i}^2}\right]^2 \frac{(2L_i - L')^2}{12} \right\}
\end{aligned}$$

and

$$\widehat{term}2 = \frac{1}{\left(\sum_{i=1}^{n} \frac{\hat{L}_i^*}{T}\right)^2} \sum_{i=1}^{n} \left(\frac{\hat{L}_i^*}{T}\right)\left(1-\frac{\hat{L}_i^*}{T}\right)\left(Z_i - \hat{R}\right)^2$$

where again $Z_i = \left(1 + e^{-\frac{\lambda L_i}{C_0}}\right) \dfrac{C_i}{2L_i}$ and $\hat{R} = \frac{1}{n} \sum_{i=1}^{n} \left(1 + e^{-\frac{\lambda L_i}{C_0}}\right) \frac{C_i}{2L_i}$.

2.7 Estimation of λ and C_0

Previous results have been developed assuming λ, the initial catch rate, and C_0, the net capacity, were known. In practice it is likely that these parameters will be unknown. This section discusses their estimation and the effect that this has on the estimators for R and $Var\left(\hat{R}\right)$.

Values for the parameters λ and C_0 can be found by using the roving survey data to fit the model $E\left[C_i\right] = C_0\left(1 - e^{-\frac{\lambda L_i}{C_0}}\right)$. The sole motivation for modelling the catch is to provide usable estimates of catch that might have occurred had the fishing episodes interrupted by the roving interview been allowed to continue to completion. These estimates are then to be used in constructing $\hat{R}$. Prediction, therefore, is the prime concern and a best fit in the least squares sense has appeal over a maximum likelihood technique. This is particularly true for the estimation of C_0 where, for small sample sizes, maximum likelihood estimation could produce unrealisticly low estimates. The interpretation of λ is still the initial catch rate, however, the interpretability of C_0 as a net capacity becomes less clear. C_0 now as an average capacity of similar nets, however, is an acceptable interpretation since the net capacity of an individual episode is determined by its catch pattern. Note that if a catch can be made by a vacant mesh opening only if all surrounding mesh openings are also vacant, then a net with one vacant mesh opening between catches is at capacity as is an identical net with two vacant mesh openings between catches; yet the total catch of the first net is approximately double that of the second. This variability in maximum capacity can also contribute to overdispersion.

Standard output from statistical software packages will provide approximately unbiased estimates for λ and C_0 together with an asymptotic variance-covariance matrix. Direct substitution of $\tilde{\lambda}_{wls}$ and $\tilde{C}_{0wls}$ into the formulae for $\hat{R}$ and $Var\left(\hat{R} \mid \lambda, C_0\right)$ then yield first order approximations. Note, however, that the distribution of C_i is binomial at each L_i and that the variance of C_i is dependent upon its mean. As such, variance homogeneity does not hold and weighted least squares might be considered using the weightings $w_i = \frac{1}{\sigma_i^2} =$

$\left[C_0 e^{-\frac{\lambda L_i}{C_0}}\left(1-e^{-\frac{\lambda L_i}{C_0}}\right)\right]^{-1}$. That is, the weighted least squares estimates for λ and C_0 in

$$\min_{\lambda, C_0} \mathcal{S}\left(\lambda, C_0\right) = \sum\nolimits_{i=1}^{n} \frac{\left[C_i - \tilde{C}_0\left(1-e^{-\frac{\tilde{\lambda} L_i}{\tilde{C}_0}}\right)\right]^2}{\tilde{C}_0 e^{-\frac{\tilde{\lambda} L_i}{\tilde{C}_0}}\left(1-e^{-\frac{\tilde{\lambda} L_i}{\tilde{C}_0}}\right)}$$

for which a solution is found using iterated weighted least squares.

While there are advantages that can strongly support a least squares approach to the estimation of λ and C_0, it is worth noting that, with the approximate consistency of $\hat{R}$ in the restricted case where λ and C_0 are known, consistent estimators of λ and C_0 would ensure the consistency of $\hat{R}$ in the general case where λ and C_0 are not known. Consistency would then be transferred to $\hat{C} = \hat{E}^* \times \hat{R}$, the estimate for total catch since $\hat{E}^*$ from the overflight survey is also consistent.

Maximum likelihood estimates of λ and C_0 would thus ensure consistency throughout. However, Green(1984) showed that, for exponential families, the estimators that are produced through iterated weighted least squares lead to those obtained using maximum likelihood techniques. Alternatively, if maximum likelihood estimation is preferred, a suggested method is to exploit the integer-valued property of C_0 (Dahiya, 1981). Required is a V such that $\mathcal{L}\left(V\right) = \mathcal{L}\left(V-1\right)$ where $\mathcal{L}$ is the likelihood function. At this point it is helpful to reparameterize as $\hat{\lambda}' = \lambda / C_0$. Now, since C_i is binomial in $\left(1-e^{-\hat{\lambda}' L_i}\right)$, V is such that

$$\frac{\mathcal{L}\left(V\right)}{\mathcal{L}\left(V-1\right)} = \frac{\prod_{i=1}^{n}\begin{pmatrix} V \\ C_i \end{pmatrix} e^{-\hat{\lambda}' L_i (V - C_i)}\left(1-e^{-\hat{\lambda}' L_i}\right)^{C_i}}{\prod_{i=1}^{n}\begin{pmatrix} V-1 \\ C_i \end{pmatrix} e^{-\hat{\lambda}' L_i (V - C_i - 1)}\left(1-e^{-\hat{\lambda}' L_i}\right)^{C_i}} = 1.$$

This reduces to

$$n \, log\left(V\right) - \sum_{i=1}^{n}\left\{log\left(V - C_i\right) - \hat{\lambda}' L_i\right\} = 0. \tag{2.9}$$

Since a linearization of the model for catch, $C_i = \left(1-e^{-\hat{\lambda}' L_i}\right) C_0$, is given by $log\left(\frac{C_0 - C_i}{C_0}\right)$ $= -L_i \lambda'$, an mle for λ' is

$$\hat{\lambda}' = -\sum_{i=1}^{n}\left\{log\left(\frac{C_0 - C_i}{C_0}\right)\right\}\left\{L_i\right\} \Big/ \sum_{i=1}^{n} L_i^2 \, . \tag{2.10}$$

Substituting V for C_0, $\hat{\lambda}'$ as a function of V can now be used in equation 2.9 above which can then be solved for V. $\hat{C}_0$ is then taken to be the integer that lies in the interval $[V-1\,,\,V)$. The *mle* for λ' can then be found from equation 2.10 which, by the generative property of *mle*'s, can then be used to find the *mle* $\hat{\lambda}$.

2.7.1 Estimation of $\hat{R}$

Previously it was shown that conditionally, $E_{L,C,\delta}\left[\hat{R}\,\middle|\,\lambda\,,\,C_0\right] \approx R = \sum_{j=1}^{N} C_j^* \Big/ \sum_{j=1}^{N} L_j^*$ which involves neither λ nor C_0. The unconditional expectation using unbiased estimates of λ and C_0 is

$$E\left[\hat{R}\right] = E_{\tilde{\lambda},\tilde{C}_0}\left[E_{L,C,\delta}\left[\hat{R}\,\middle|\,\tilde{\lambda}\,,\,\tilde{C}_0\right]\right] \approx E_{L,C,\delta}\left[\hat{R}\,\middle|\,E[\tilde{\lambda}]\,,\,E[\tilde{C}_0]\right] \approx R.$$

Hence $\hat{R} = \frac{1}{n}\sum_{i=1}^{n}\left(1+e^{-\frac{\tilde{\lambda}L_i}{\tilde{C}_0}}\right)\frac{C_i}{2L_i}$ is taken as an unconditioned and approximately unbiased estimator for R. As such, no adjustment to $\hat{R}$ is deemed necessary when least squares estimates for λ and C_0 are substituted directly into the formula.

2.7.2 Estimation of $Var\left(\hat{R}\right)$

The unconditional variance of $\hat{R}$ is given by

$$Var\left(\hat{R}\right) = E_{\tilde{\lambda},\tilde{C}_0}\left[Var\left(\hat{R}\,\middle|\,\tilde{\lambda}\,,\,\tilde{C}_0\right)\right] + Var_{\tilde{\lambda},\tilde{C}_0}\left(E\left[\hat{R}\,\middle|\,\tilde{\lambda}\,,\,\tilde{C}_0\right]\right)$$

where the term $Var_{\tilde{\lambda},\tilde{C}_0}\left(E\left[\hat{R}\,\middle|\,\tilde{\lambda}\,,\,\tilde{C}_0\right]\right)$ can be taken to be zero because, as argued above, the term $E\left[\hat{R}\,\middle|\,\tilde{\lambda}\,,\,\tilde{C}_0\right]$ should involve neither λ nor C_0. To estimate the remaining term, note that in general, for any function $g\left(\tilde{\lambda}\,,\,\tilde{C}_0\right)$, approximate expectations are given by

$$\begin{aligned}
\text{First Order:} \quad & E_{\tilde{\lambda},\tilde{C}_0}\left[g\left(\tilde{\lambda}\,,\,\tilde{C}_0\right)\right] &=& \; g\left(\mu_1\,,\,\mu_2\right) \\
\text{Second Order:} \quad & E_{\tilde{\lambda},\tilde{C}_0}\left[g\left(\tilde{\lambda}\,,\,\tilde{C}_0\right)\right] &=& \; g\left(\mu_1\,,\,\mu_2\right) + \frac{1}{2}E\left[\boldsymbol{\eta}'\mathbf{A}\boldsymbol{\eta}\right] \\
& &=& \; g\left(\mu_1\,,\,\mu_2\right) + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}\mathbf{A}\right)
\end{aligned}$$

where

$$\boldsymbol{\mu} = (\lambda\,, C_0)';$$
$$\boldsymbol{\eta} = \left(\tilde{\lambda} - \lambda\,,\, \tilde{C}_0 - C_0\right)';$$
$$\boldsymbol{\Sigma} = \text{variance-covariance matrix of } \tilde{\lambda} \text{ and } \tilde{C}_0; \text{ and}$$
$$\mathbf{A} = \begin{bmatrix} \dfrac{\partial^2 g}{\partial \tilde{\lambda}^2} & \dfrac{\partial^2 g}{\partial \tilde{\lambda}\,\partial \tilde{C}_0} \\ \dfrac{\partial^2 g}{\partial \tilde{C}_0\,\partial \tilde{\lambda}} & \dfrac{\partial^2 g}{\partial \tilde{C}_0^2} \end{bmatrix} \text{ evaluated at } \boldsymbol{\mu}.$$

Then, for the sample estimator, where $g\left(\tilde{\lambda}\,,\,\tilde{C}_0\right) = \widehat{Var}\left(\hat{R}\,\middle|\,\tilde{\lambda}\,,\,\tilde{C}_0\right)$,

$$\begin{aligned}
\widehat{Var}\left(\hat{R}\right) &= \hat{E}_{\tilde{\lambda},\tilde{C}_0}\left[\widehat{Var}\left(\hat{R}\,\middle|\,\tilde{\lambda}\,,\,\tilde{C}_0\right)\right] \\
&= \hat{E}_{\tilde{\lambda},\tilde{C}_0}\left[\frac{1}{n^2}\sum_{i=1}^{n}\left\{\frac{\tilde{C}_0\left(1-e^{-\frac{2\tilde{\lambda}L_i}{\tilde{C}_0}}\right)}{4L_i^2}\left(e^{-\frac{\tilde{\lambda}L_i}{\tilde{C}_0}}+e^{-\frac{2\tilde{\lambda}L_i}{\tilde{C}_0}}\right)\right.\right. \\
&\quad + \left[1+\frac{e^{-\frac{\tilde{\lambda}\hat{L}_i^*}{\tilde{C}_0}}}{\tilde{C}_0\left(1-e^{-\frac{\tilde{\lambda}\hat{L}_i^*}{\tilde{C}_0}}\right)}\right]\left[\frac{2\tilde{\lambda}L_i e^{-\frac{2\tilde{\lambda}L_i}{\tilde{C}_0}}-\tilde{C}_0\left(1-e^{-\frac{2\tilde{\lambda}L_i}{\tilde{C}_0}}\right)}{2{L_i}^2}\right]^2\left.\frac{(2L_i-L')^2}{12}\right\} \\
&\quad \left.+\frac{1}{\left(\sum_{i=1}^{n}\frac{\hat{L}_i^*}{T}\right)^2}\sum_{i=1}^{n}\left(\frac{\hat{L}_i^*}{T}\right)\left(1-\frac{\hat{L}_i^*}{T}\right)\left(Z_i-\hat{R}\right)^2\right]
\end{aligned} \tag{2.11}$$

where unknown parameters are replaced by their sample estimates. One method of evaluating the $\mathbf{A}$ matrix of partial derivatives of $g\left(\tilde{\lambda}\,,\,\tilde{C}_0\right)$ is the use of finite differences and the definition of a derivative. That is, for the second partial of $g\left(\tilde{\lambda}\,,\,\tilde{C}_0\right)$ with respect to $\tilde{\lambda}$

$$\frac{\partial^2}{\partial\tilde{\lambda}^2}\,g\left(\tilde{\lambda}\,,\,\tilde{C}_0\right) = \frac{\left(g\left(\tilde{\lambda}+\Delta\,,\,\tilde{C}_0\right)-g\left(\tilde{\lambda}\,,\,\tilde{C}_0\right)\right)-\left(g\left(\tilde{\lambda}\,,\,\tilde{C}_0\right)-g\left(\tilde{\lambda}-\Delta\,,\,\tilde{C}_0\right)\right)}{\Delta^2}$$

for some suitable small Δ. A similar procedure is used to approximate the second partial of g with respect to $\tilde{C}_0$. For the partial with respect to $\tilde{\lambda}$ and $\tilde{C}_0$

$$\begin{aligned}
\frac{\partial^2}{\partial\tilde{\lambda}\partial\tilde{C}_0}\,g\left(\tilde{\lambda}\,,\,\tilde{C}_0\right) &= \left[\frac{\left(g\left(\tilde{\lambda}+\Delta\,,\,\tilde{C}_0+\Delta\right)-g\left(\tilde{\lambda}+\Delta\,,\,\tilde{C}_0-\Delta\right)\right)}{2\Delta}\right. \\
&\quad \left.-\frac{\left(g\left(\tilde{\lambda}-\Delta\,,\,\tilde{C}_0+\Delta\right)-g\left(\tilde{\lambda}-\Delta\,,\,\tilde{C}_0-\Delta\right)\right)}{2\Delta}\right]\Big/\,2\Delta.
\end{aligned}$$

Note that it is the $\frac{1}{2}E\left[\boldsymbol{\eta}'\mathbf{A}\boldsymbol{\eta}\right] = \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}\mathbf{A}\right)$ term that accounts for the additional variability due the estimation of λ and C_0.

2.8 Overdispersion

It was noted earlier that overdispersion is a real possibility in applications. Based on the Pearson χ^2 goodness of fit statistic, a test statistic for the presence of overdispersion is based on

$$\psi = \sum_{i=1}^{n} \frac{(Residual)_i{}^2}{\tilde{C}_0\, e^{-\frac{\tilde{\lambda} L_i}{\tilde{C}_0}} \left(1 - e^{-\frac{\tilde{\lambda} L_i}{\tilde{C}_0}}\right)} \tag{2.12}$$

where, under the assumption that catch at L_i is binomially distributed and since two parameters have been estimated, ψ is asymptotically distributed as chi-square with $n-2$ degrees of freedom.

It also follows that an unbiased estimator for the scale parameter, ϕ^2, is $\hat{\phi}^2 = \psi/(n-2)$ since $E\left[\psi\right] = \phi^2(n-2)$. However, ψ is apt to be unstable because, in theory, variance for catch becomes small as catch nears $\tilde{C}_0$ while in practice, larger deviations are possible. A more stable form based on approximating the expectation of a ratio with the ratio of the expectations is then

$$\hat{\phi}^2 = \sum_{i=1}^{n} (Residual)_i{}^2 \Bigg/ \sum_{i=1}^{n} \tilde{C}_0\, e^{-\frac{\tilde{\lambda} L_i}{\tilde{C}_0}} \left(1 - e^{-\frac{\tilde{\lambda} L_i}{\tilde{C}_0}}\right) \tag{2.13}$$

and the variance estimator accounting for overdispersion is

$$\widehat{Var}^*\left(\hat{R}\right) = \hat{\phi}^2 \widehat{Var}\left(\hat{R}\right) \cdot \tag{2.14}$$

2.9 Variance of Total Catch

Total effort is estimated as $\hat{E}^* = I \times T$ where I are counts recorded by an overflight survey. If the count is "instantaneous" and if the time of overflight is chosen randomly, $\hat{E}^*$ is an unbiased estimate for the total effort during T (Robson, 1960 , Pollock et al., 1994). With independent and approximately unbiased estimators for total effort and catch rate, the result by Goodman (1960) is used for the product $Var\left(\hat{C}^*\right) = Var\left(\hat{E}^* \times \hat{R}\right)$ to give

$$\widehat{Var}\left(\hat{C}^*\right) = \left(\hat{R}\right)^2 \times \widehat{Var}\left(\hat{E}^*\right) + \left(\hat{E}^*\right)^2 \times \widehat{Var}\left(\hat{R}\right) - \widehat{Var}\left(\hat{E}^*\right) \times \widehat{Var}\left(\hat{R}\right) \cdot \tag{2.15}$$

2.10 Measuring Model Effectiveness

Use of the proposed model is appropriate if the rate of catch declines as a function of catch. Suppose that catch rate can be modelled as a continuous time Markov process where expected catch at time L_j is given by $C_0\left(1-e^{-\frac{\lambda L_i}{C_0}}\right)$. If $\hat{R}=\frac{1}{n}\sum_{i=1}^{n}\left(1+e^{-\frac{\lambda L_i}{C_0}}\right)\frac{C_i}{2L_i}$, *i.e.* the approximately unbiased estimator of catch per effort developed under the proposed model, and $\hat{R}_{\lambda^*}=\frac{1}{n}\sum_{i=1}^{n}\frac{C_i}{L_i}=\frac{1}{n}\sum_{i=1}^{n}\frac{2C_i}{2L_i}$ where λ^* is some constant catch rate, *i.e.* the estimator of catch per effort assuming a constant catch rate, then one way to measure the effectiveness of using $\hat{R}$ would be to examine the difference between $\hat{R}_{\lambda^*}$ and $\hat{R}$. (Note that if the catch rate does in fact decrease with catch, $\hat{R}_{\lambda^*}$ will tend to overestimate R.) A sample estimate of this difference can be found from

$$\begin{aligned}\hat{R}_{\lambda^*}-\hat{R} &= \frac{1}{n}\sum_{i=1}^{n}\frac{2C_i-\left(1+e^{-\frac{\lambda L_i}{C_0}}\right)C_i}{2L_i}\\ &= \frac{1}{n}\sum_{i=1}^{n}\left(1-e^{-\frac{\lambda L_i}{C_0}}\right)\frac{C_i}{2L_i}.\end{aligned}$$

An approximate expected value of such a measure is

$$E\left[\hat{R}_{\lambda^*}-\hat{R}\right]\approx\frac{E\left[\sum_{j=1}^{N}\delta_j\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)C_j\right]}{E\left[\sum_{j=1}^{N}\delta_j\right]E\left[\sum_{j=1}^{N}2\delta_jL_j\right]}$$

since $E[n]=E\left[\sum_{j=1}^{N}\delta_j\right]$.

Evaluating the numerator, where $E_{L_j}\left[1-e^{-\frac{2\lambda L_j}{C_0}}\right]$ is found using a first order Taylor approximation, gives

$$\begin{aligned}E_{L_j,C_j,\delta_j}\left[\sum_{j=1}^{N}\delta_j\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)C_j\right] &= \sum_{j=1}^{N}E_{L_j}\left[\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)E_{C_j}\left[C_j|\,L_j,C_j^*,L_j^*\right]\right]\frac{L_j^*}{T}\\ &= \sum_{j=1}^{N}E_{L_j}\left[\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)C_j^*\left(\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)\right]\frac{L_j^*}{T}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{N} E_{L_j}\left[\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)^2\right]\frac{C_j^*}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{L_j^*}{T} \\
&\approx \sum_{j=1}^{N}\frac{\left(1-e^{-\frac{\lambda L_j^*/2}{C_0}}\right)}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{C_j^* L_j^*}{T} \\
&= \sum_{j=1}^{N}\frac{C_0\left(1-e^{-\frac{\lambda L_j^*/2}{C_0}}\right)}{C_0\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{C_j^* L_j^*}{T} \\
&= \sum_{j=1}^{N}\frac{E\left[C(L_j^*/2)\right]}{E\left[C(L_j^*)\right]}\frac{C_j^* L_j^*}{T}
\end{aligned}
$$

which is seen to involve a downward adjustment of C_j^*, the catch at completion of a fishing episode, by the ratio of the expected catch at half episode over expected catch at full episode.

Components in the denominator are the same as those found in the denominators of the mean of ratios estimator considered in section 2.5.2 and of the ratio of means estimator considered in section 2.5.1 where it was shown that

$$E_{\delta_j}\left[\sum_{j=1}^{N}\delta_j\right] = \sum_{j=1}^{N} L_j^* \Big/ T$$

and

$$E_{L_j,C_j,\delta_j}\left[\sum_{j=1}^{N}2\delta_j L_j\right] = \sum_{j=1}^{N}\left(L_j^*\right)^2 \Big/ 2T.$$

Therefore,

$$E\left[\hat{R}_{\lambda^*}-\hat{R}\right] \approx \frac{\sum_{j=1}^{N} 2\frac{E\left[C(L_j^*/2)\right]}{E\left[C(L_j^*)\right]}C_j^* L_j^*}{\left\{\sum_{j=1}^{N} L_j^* \Big/ T\right\}\left\{\sum_{j=1}^{N}\left(L_j^*\right)^2\right\}}$$

which has no easy interpretation.

More meaningful might be a factor, F_{λ,λ^*}, that could be used to adjust a total catch that had been estimated incorrectly based on $\hat{R}_{\lambda^*} = \sum_{i=1}^{n} 2C_i \Big/ \sum_{i=1}^{n} 2L_i$. A possible sample value for such a factor is given by

$$\hat{F}_{\lambda,\lambda^*} = \frac{\sum_{i=1}^{n} \left(1 + e^{-\frac{\lambda L_i}{C_0}}\right) C_i/2L_i}{\sum_{i=1}^{n} 2C_i/2L_i}. \tag{2.16}$$

Note that since this quantity deals with end of episode, albeit estimated, the more usual ratio of means is now used.

The expected value of this measure is approximated as

$$E\left[\hat{F}_{\lambda,\lambda^*}\right] \approx \frac{E\left[\sum_{j=1}^{N} \delta_j \left(1 + e^{-\frac{\lambda L_j}{C_0}}\right) \frac{C_j}{2L_j}\right]}{E\left[\sum_{j=1}^{N} \delta_j \frac{C_j}{L_j}\right]}.$$

Evaluation of the numerator is the same as that for the numerator in the mean of ratios estimator considered in section 2.5.2 where it was shown that

$$E_{L_j,C_j,\delta_j}\left[\sum_{j=1}^{N} \delta_j \left(1 + e^{-\frac{\lambda L_j}{C_0}}\right) \frac{C_j}{2L_j}\right] \approx \sum_{j=1}^{N} C_j^* / T.$$

Evaluating the denominator, again using a first order Taylor approximation for expectations of L_j, gives

$$\begin{aligned}
E_{L_j,C_j,\delta_j}\left[\sum_{j=1}^{N} \delta_j \frac{C_j}{L_j}\right] &= \sum_{j=1}^{N} E_{L_j}\left[\frac{1}{L_j} E_{C_j}\left[C_j \middle| L_j, C_j^*, L_j^*\right]\right] \frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} E_{L_j}\left[\frac{1}{L_j} C_j^* \left(\frac{1 - e^{-\frac{\lambda L_j}{C_0}}}{1 - e^{-\frac{\lambda L_j^*}{C_0}}}\right)\right] \frac{L_j^*}{T} \\
&\approx \sum_{j=1}^{N} \frac{2}{L_j^*} C_j^* \left(\frac{1 - e^{-\frac{\lambda L_j^*/2}{C_0}}}{1 - e^{-\frac{\lambda L_j^*}{C_0}}}\right) \frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} \frac{2C_0\left(1 - e^{-\frac{\lambda L_j^*/2}{C_0}}\right)}{C_0\left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)} \frac{C_j^*}{T} \\
&= \sum_{j=1}^{N} \frac{2E[C(L_j^*/2)]}{E[C(L_j^*)]} \frac{C_j^*}{T}.
\end{aligned}$$

The expected value of $\hat{F}_{\lambda,\lambda^*}$ is then approximated as

$$E\left[\hat{F}_{\lambda,\lambda^*}\right] \approx \frac{\sum_{j=1}^{N} C_j^*}{\sum_{j=1}^{N} \frac{2E[C(L_j^*/2)]}{E[C(L_j^*)]} C_j^*}$$

which has the interpretation of total catch divided by total catch increased by a factor equal to the ratio of strict doubling to nonlinear doubling. In this sense, after an erroneous use of $\hat{R}$, $\hat{F}_{\lambda,\lambda^*}$ would be the natural choice for an adjustment factor.

Also note that as $C_0 \to \infty$, both $\hat{F}_{\lambda,\lambda^*}$ and $E\left[\hat{F}_{\lambda,\lambda^*}\right] \to 1$; again consistent with the proposed model as a generalization of the Hoenig et al. (1997) model.

2.11 Bias in $\hat{R}$

Previously it was shown that $\hat{R} = \frac{1}{n}\sum_{i=1}^{n}\left(1 + e^{-\frac{\lambda L_i}{C_0}}\right)\frac{C_i}{2L_i}$ was an approximately unbiased estimator for R, however, a slight positive bias has been masked since $E\left[1/n\right] > 1/E\left[n\right]$ It was also noted that some minimum time to interview, L', must be chosen in order to stabilize $Var\left(\hat{R}\right)$. In doing so, greater stability is realized with larger values of L' but this is also accompanied by increasing bias in $\hat{R}$. It is therefore useful when determining L' to have this portion of the bias expressed as a function of L'. Therefore, and again ignoring the bias introduced by approximating the expectation of a ratio with the ratio of expectations and proceeding as in section 2.5.2,

$$\begin{aligned} E\left[\hat{R}\right] &= \frac{E\left[\sum_{j=1}^{N} \delta_j \left(1 + e^{-\frac{\lambda L_j}{C_0}}\right)\frac{C_j}{2L_j}\right]}{E\left[\sum_{j=1}^{N} \delta_j\right]} \\ &= \frac{\sum_{j=1}^{N} E_{L_j}\left[\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{L_j}\right]\frac{C_j^*}{2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{L_j^*-L'}{T}}{\sum_{j=1}^{N}\frac{L_j^*}{T}} \end{aligned}$$

$$= \frac{\sum_{j=1}^{N} E_{L_j}\left[\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{L_j}\right]\frac{C_j^*}{2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{L_j^*}{T} - \sum_{j=1}^{N} E_{L_j}\left[\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{L_j}\right]\frac{C_j^*}{2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{L'}{T}}{\sum_{j=1}^{N}\frac{L_j^*}{T}}$$

$$\approx \frac{\sum_{j=1}^{N} C_j^* \Big/ T}{\sum_{j=1}^{N} L_j^* \Big/ T} - \frac{\sum_{j=1}^{N} E_{L_j}\left[\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{L_j}\right]\frac{C_j^*}{2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{L'}{T}}{\sum_{j=1}^{N}\frac{L_j^*}{T}}$$

$$= R - \frac{\sum_{j=1}^{N} E_{L_j}\left[\frac{1-e^{-\frac{2\lambda L_j}{C_0}}}{L_j}\right]\frac{C_j^*}{2\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)}\frac{L'}{T}}{\sum_{j=1}^{N}\frac{L_j^*}{T}}$$

showing that, with the inclusion of a L', $\hat{R}$ does tend to overestimate R and using first order Taylor approximations,

$$\hat{B}_{L'} = \frac{1}{n}\sum_{i=1}^{n} \frac{C_0\left(1-e^{-\frac{\hat{\lambda}(2L_i+L')}{\hat{C}_0}}\right)}{(2L_i+L')}\frac{L_i'}{T}.$$

2.12 Simulation Results

As shown in Table 2.1, a test population of 225 fishing episodes was generated to assess estimators for R and its variance. To ensure that the distribution of effort resembles that of an actual gill net fishery, episode lengths were set to multiples of 30 minutes and ranged between 0.5 and 24.0 hours with the greatest repetition of lengths occurring between 2.5 and 6.0 hours (Figure 2.1). This pattern resembles that of the Fraser River fisheries except for a noticeable spike which occurs at 24 hours (see Palermo and Ennever, 1997). Each episode was assigned "catches" over its duration according to the continuous time Markov process given by Equation 2.1. Assuming that no more than one fish would be caught during a one minute interval, an initial catch rate was taken to be 12 fish/hour and the net

Table 2.1: Defining characteristics of the simulated population.

Start of the fishing day	00:00
End of the fishing day	24:00
Total number of fishing episodes	225
Net capacity (C_0)	75
Initial catch rate (as catch/hr)	12
Initial catch rate (as λ)	0.2003
Total catch	10,533
Total effort (in minutes)	111,060
Average of the episode catch rates	0.1154
Total catch / Total effort (CPUE)	0.0948

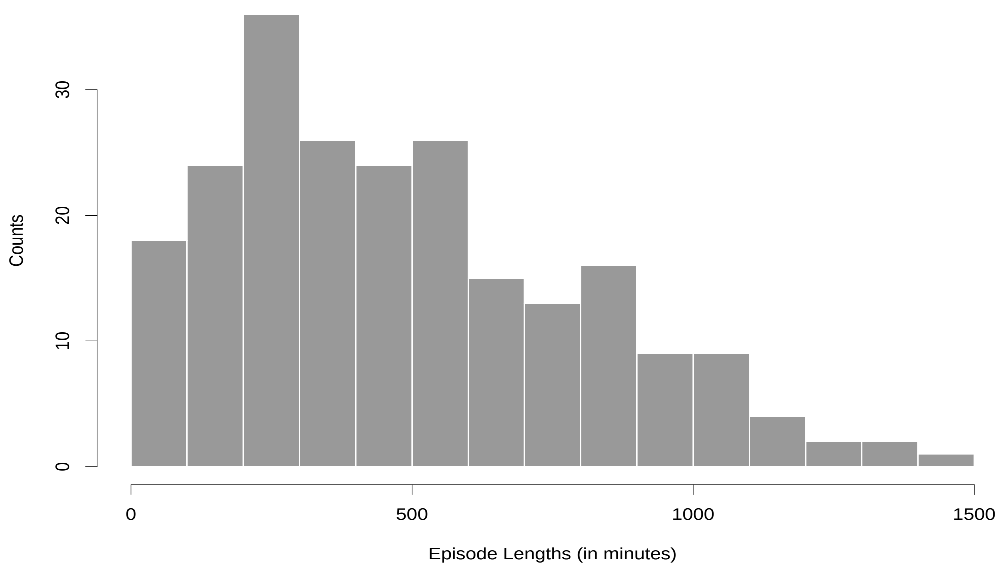

Figure 2.1: Fishing episodes by length for the simulated population (total count = 225).

capacity, C_0, to be 75 fish which results in an initial catch rate of $\lambda = 0.2003$/minute. A cumulative geometric distribution was used to simulate the time between the k^{th} and the $(k+1)^{st}$ catch (Figure 2.2). Episodes were placed either uniformly over the fishing day or

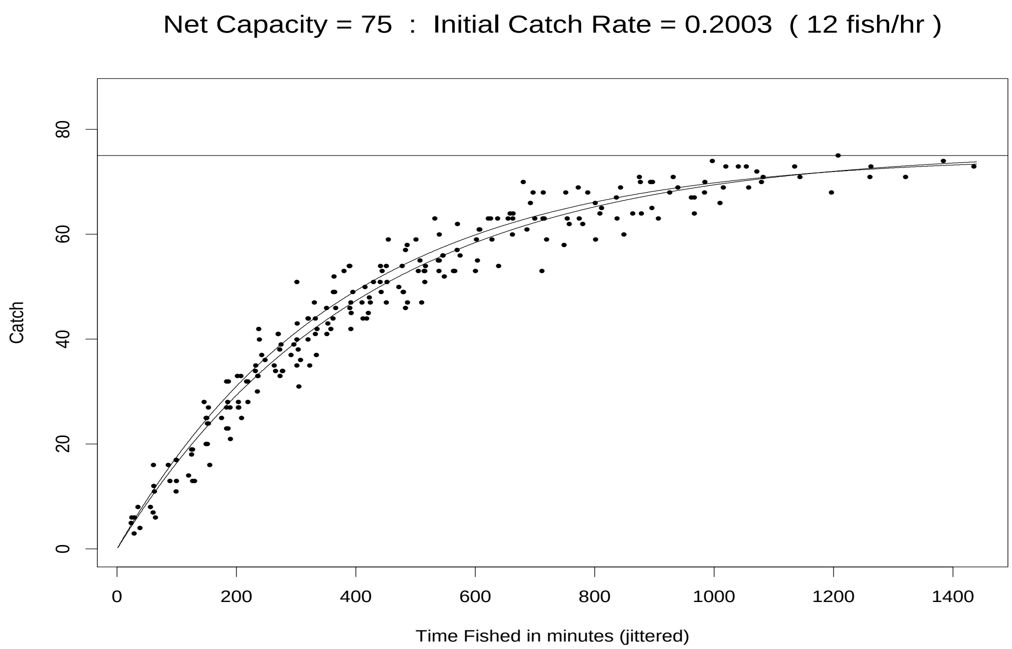

Figure 2.2: Catch vs. Time Fished for completed episodes of the simulated population. Also shown are plots of $C_0\left(1-e^{-\frac{\lambda t}{C_0}}\right)$, the expected catch at time t for true (upper line) and least square estimates (lower line) of λ and C_0.

using a fishing time preference curve having a peak at 7:30 a.m. Windows for the roving surveys were defined as the full fishing day or the daylight between 6:00 a.m. and 6:00 p.m.

Results of the simulations are given in Table 2.2. In each simulation, 500 roving surveys were generated within the windows indicated under each fishing time preference scenario. The positive bias in $\hat{R}$ in part reflects the bias caused by assuming the expectation of a ratio equal the ratio of the expectations as done in the Taylor approximation. Note that while seemingly small, $B/Std(\hat{R}) > 0.2$ and therefore should not be viewed as negligible (Cochran, 1977). While variance estimation is made difficult by the number of random components and, in general, results in an overestimation, the magnitude of the bias leads to confidence intervals with coverage less than the nominal level.

Restricting the window on the roving survey induces further bias (here offsetting) and a noticeable decrease in the variance estimates. A negative bias stems from the fact that episodes which are not fully contained within the window are intercepted, on average, at points other than mid-episode. Interceptions at points low on the "doubling" curve overestimate catch at $2L_i$ while those at points high on the curve underestimate. With use of the window, overall there are greater losses from "missed doublings" on the steep (*i.e.* low) end of the curve than from "missed doublings" at the other (*i.e.* high) end. For variance estimation, a roving survey must provide for possible inclusion of all episodes by making a complete pass through the entire fishing resource over the full length of the fishing day (Robson, 1961). With use of a window this does not happen. A "quick fix" would be to assume that the window was a randomly chosen interval and expand the results by the ratio of the length of the fishing day to the width of the window (Hoenig et al., 1993).

When comparing the effects of a preference in fishing times, note that a preference scenario tends to have a greater concentration of shorter episodes near the peak time. With a full roving survey this should have little effect on $\hat{R}$ and its variance, however, with a restricted survey window that contains this peak time, the sample will contain a greater concentration of shorter episodes, *i.e.* episodes that "double" low on the curve, tending to inflate estimates of R. At the same time, with the greater number of episodes eligible for sample inclusion, *i.e.* in effect more closely resembling a complete pass over the full length of the fishing day, the variance estimator is apt to be less affected by the window restriction.

Hoenig et al. (1997) suggested that some minimum time for intercepted episodes should be used (they suggested 30 minutes). Here a 15 minute minimum was used. Table 2.3 shows that $L' = 15$ seems to be a reasonable choice (for this particular population) since both theoretical and estimated variances appear to be stabilized and contribution to bias is minimal. Note that $B_{L'}$ has a negative effect on $\hat{R}$ and is increasing with L', however, it should not be used as a device to offset the inherent positive bias in $\hat{R}$.

Table 2.2: Effects of varying the roving survey window for two fishing time preference scenarios on the performance of $\hat{R}$, estimators of its variance and the confidence interval coverage. (Simulated population with 500 replicated samples.)

Fishing Time Preference:	Uniform[1]		Uni-Modal[1]	
Roving Window:	Full Day[2]	Daylight[2]	Full Day[2]	Daylight[2]
R	0.0948	0.0948	0.0948	0.0948
Mean($\hat{\mathrm{R}}$)	0.1003	0.0978	0.1006	0.0981
Std($\hat{\mathrm{R}}$)	0.0048	0.0034	0.0047	0.0034
Mean($\hat{\mathrm{B}} = \hat{\mathrm{R}}$-R)	0.0055	0.0030	0.0058	0.0033
Var($\hat{\mathrm{R}}$) Over simulations	2.28×10^{-5}	1.14×10^{-5}	2.19×10^{-5}	1.14×10^{-5}
$\mathrm{Var}(\hat{\mathrm{R}}\mid\lambda, \mathrm{C}_0)$ Eqn 2.4	2.08×10^{-5}	2.08×10^{-5}	2.08×10^{-5}	2.08×10^{-5}
$\mathrm{Mean}\left(\widehat{\mathrm{Var}}(\hat{\mathrm{R}})\right)$ Eqn 2.11	2.80×10^{-5}	1.90×10^{-5}	2.82×10^{-5}	1.83×10^{-5}
Confidence Intervals:				
Nominal coverage	Actual percent coverage			
95 %	88	96	86	95
90 %	76	90	76	89
85 %	67	84	66	82
80 %	61	78	58	76

[1] Uniform: no preferred time; Unimodal: preferred time of 7:30 a.m.
[2] Full Day is 24 hr. period; Daylight hours taken to be 6:00 a.m. to 6:00 p.m.

Table 2.3: Effects of varying the minimum time required for sample inclusion on bias and variance. (Simulated population.)

	Minimum Time Required		
	5 min.	15 min.	25 min.
R	0.0948	0.0948	0.0948
Mean($\hat{R}$)	0.1017	0.1003	0.0987
Mean($\hat{B} = \hat{R}$-R)	0.0069	0.0055	0.0039
Mean($\hat{B}_{L'}$)	0.0004	0.0010	0.0017
Var($\hat{R}$) Over simulations	2.70×10^{-5}	2.28×10^{-5}	2.12×10^{-5}
Mean$\left(\widehat{\mathrm{Var}}(\hat{R})\right)$ Eqn 2.11	3.30×10^{-5}	2.80×10^{-5}	2.60×10^{-5}

2.13 The Fraser River Study

The Fraser River is a major sockeye river in British Columbia, Canada. Along with a large ocean fishery, there are substantial in-river terminal fisheries. The Department of Fisheries and Oceans, Canada, is responsible for the monitoring of this fishery.

In 1995, in-river catch estimates were formed using combinations of helicopter overflights to estimate effort and access or roving surveys (depending upon location) to estimate the catch rate. During the roving surveys, boat patrols approached active fishers who were asked to pull in their nets to check that it met regulations and to count catch. A selected subset of this data consisting of two adjoining stretches of river over two consecutive weeks is used to illustrate the use of the proposed estimators and to compare these results with those had a constant catch rate been employed.

For practical reasons, overflight and roving surveys were restricted to daylight hours, but nets could be set any time in a 24–hour period. This practice violates the assumptions of inclusion based on length of time fished and mid-episode interception. Nonsampling errors that could impact on the results were also present. For example, retrieving a net in mid-episode is an imposition, and poor cooperation on the part of some fishers is almost assured. Also, it was widely known that policy and quotas would be based on the results of the study. No attempt was made to quantify these effects or question the integrity of the protocol regarding enforcement or instances of no information from unattended nets. Also, the raw data files contained "information" that exceeded reasonable physical limitations. For example, it contained fishing episodes longer than 24 hours or less than half an hour, or episodes with catch in excess of 33 fish (*i.e.* approximately 200 pounds for any net assuming 6 lbs./fish). These peculiar data were excluded. This screening did little to change the overall pattern of the plots (not shown) of fish caught versus time fished, for which there was considerable scatter and variability by week and area (Figure 2.3).

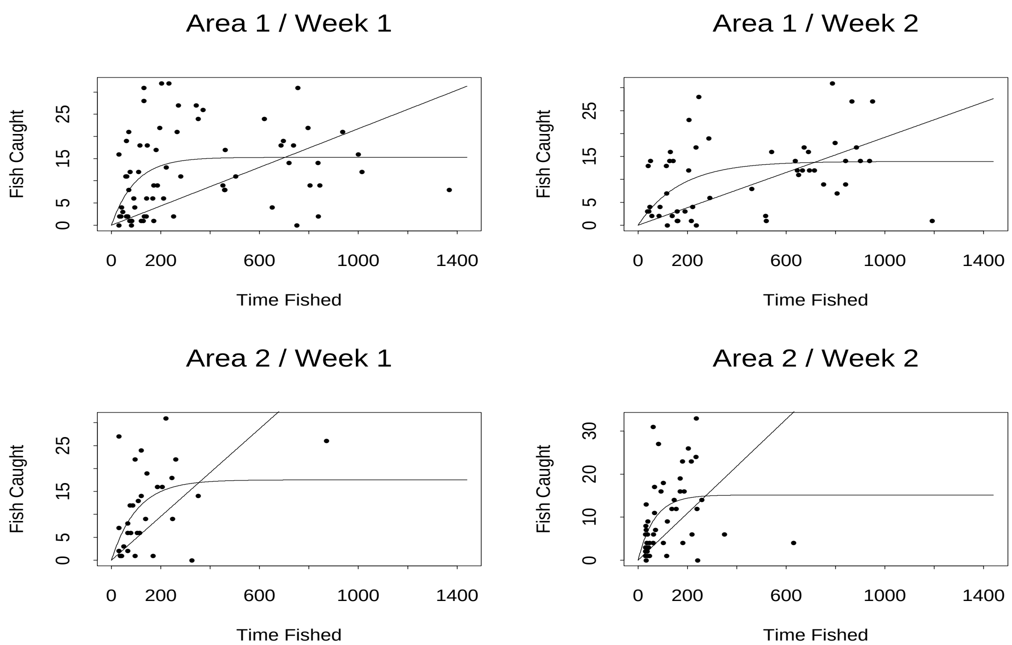

Figure 2.3: Edited data for selected study areas located on the Fraser River, Canada. Lines show the expected catch using the constant catch rate model (straight lines) and the proposed saturation model (curved lines).

Estimates were made for each week and stretch of river. Table 2.4 summarizes these results. The overdispersion suggested by plots (not shown) was confirmed, e.g. for area 1 week 1, the test for overdispersion gave p-value < 0.0001 with $\hat{\phi}^2 = 40.7$. Significant values were also obtained for the other week and areas. To underscore the effects of this large overdispersion, results are presented with and without inclusion of the $\hat{\phi}^2$ factor. Note that the estimates for catch using the Poisson model are considerably larger than those using the proposed model (35-50 %) showing that careful consideration should be given to choice of model. Yet, with such large variability in the data, $\hat{R}_{Poiss}$ remains within the confidence region of $\hat{R}$, albeit at the edge.

Surveys restricted to daylight hours might be partially rationalized with claims of obtaining a sample representative of catch per effort. While a reasonable estimate of R might be obtained, variance estimates are apt to be inflated because doubling L_i to estimate time of a completed episode, results in extreme values of Z_i. It should also be noted that, with both overflight and roving surveys restricted to daylight hours, measurement of night time fishing practices was inadequate. If it can be reasoned that night fishing consists mostly of fewer but more lengthy episodes, then it can be argued that, with saturation, night catch rates will be less than that of daytime catch rates, resulting in an overestimation of $\hat{C}$. Also, overflight counts would tend to be high as a measure typical of activity for the full fishing day, again contributing to an overestimation. With its "full" doubling of observed catch, errors caused by these factors would be more pronounced under the Poisson model.

Table 2.4: Summary of results from the Fraser River study.

		Area 1		Area 2	
		Week 1	Week 2	Week 1	Week 2
$\hat{E}^1$ (net-hours)	Estimate	2,608	2,692	1,036	968
	$\widehat{\text{SE}}$	84	457	123	273
$\hat{R}$ (catch/minute)	Estimate	0.0463	0.0334	0.0815	0.0776
	$\widehat{\text{SE}}$	0.0045	0.0034	0.0225	0.0202
	$\widehat{\text{SE}}_\phi^2$	0.0289	0.0191	0.1017	0.0986
$\hat{R}_{Poiss}$ (catch/minute)	Estimate	0.0690	0.0459	0.1109	0.1137
$\hat{C} = \hat{R} \times \hat{E}$	Estimate	7,248	5,397	5,065	4,508
	$\widehat{\text{SE}}$	747	1,071	1,531	1,761
	$\widehat{\text{SE}}_\phi^2$	4,534	3,255	6,396	6,087
$\hat{C} = \hat{R}_{Poiss} \times \hat{E}$	Estimate	10,804	7,420	6,896	6,606
		95% Confidence intervals for R			
Without overdispersion	Lower bound	0.0374	0.0268	0.0374	0.0380
	Upper bound	0.0552	0.0400	0.1256	0.1172
With overdispersion	Lower bound	-0.0104	-0.0039	-0.1179	-0.1157
	Upper bound	0.1030	0.0708	0.2809	0.2709

[1] Derived from the overflight survey.
[2] ϕ denotes overdispersion.

2.14 Conclusions and Discussion

In certain applications the proposed model should be a useful alternative to the more usual model based on a constant catch rate. Use, however, involves a more complex model for catch rate and the need to estimate λ and C_0, which can lead to an overall loss in precision. Use also requires more assumptions than the stationary Poisson process method of Hoenig et al. (1997), for example, that n be small relative to N. Nevertheless, as a generalization of the constant catch rate model, it should serve to expand the scope and increase the utility of the complemented survey approach to catch estimation.

Shortcomings inherent in using a roving survey design, as compared with an access design, are also apparent. Most notably, there are the difficulties associated with designing a proper sampling scheme, the random sample size which makes planning difficult, and the less stable mean of ratios form of the catch rate estimator. These features can result in variance estimates which are unacceptably large and, under a sampling scheme where inclusion is dependent upon L_j^*, exclude the option of increasing sample size as a means of remedy. In addition, unbiasedness in $\hat{R}$ is only approximate. In the simulation results the bias was not inconsequential and it is doubtful that real data would fare better.

Restrictions in the roving window are a real possibility, especially when considering factors such as darkness and helicopter availability. As pointed out in the Section 2.12, estimators of variance tend to be more affected than those of R. In general, underestimations are expected. While the biasing effects on $\hat{R}$ caused by episodes overlapping the window at one end might well offset those at the other, to produce usable estimates of R, the window must still capture a typical mix of episodes - this is unlikely if catch rates or length of episode varies by time of day. Similar problems exist when episodes overlap the fishing day and these are the subject of further research.

2.15 Acknowledgments

Data from The Georgia Strait Creel Surveys produced by Department of Fisheries and Oceans, 3225 Stephenson Point Road, Nanaimo, B.C., V9T 1K3. The research was supported by Research Grants from the Natural Science and Engineering Research Council (NSERC) of Canada.

Chapter 3

Restricted Randomization in the Effort Survey

The method of estimating total effort based on instantaneous counts requires that sample times are random over the full fishing period. In practice this may not always be possible and, as a result, the usual estimator may be severely biased. Such a restriction in randomization is likely when aircraft are used to make "instantaneous" counts of fishing activity. This chapter proposes alternate estimators for use with both access and roving designs in conjunction with effort surveys for which sample times are not random. Ratio type estimators based on activity counts are developed under various scenarios and their performance examined through simulation. In addition, optimizing strategies for use with multiple activity counts are explored. Finally, data from an in-river gill net fishery on the Fraser River is used to illustrate these results.

3.1 Introduction and Motivation

When estimating catch in a fishery, one useful method is to employ two independent surveys: one to measure $CPUE$, the catch per unit effort; and another to measure E, the total effort. With unbiased estimators for $CPUE$ and E it is then possible to estimate total catch C by their product. This technique, however, requires that both the $CPUE$ survey and the effort survey be random.

Typically, the $CPUE$ survey is a ground survey in which fishers are contacted to produce catch and time fished information. This is then expressed as catch per unit time of fishing effort. The estimate of total effort is derived from the product of an "instantaneous" count of active fishing episodes made at a randomly chosen time (or average of such counts) and the number of time units in the fishing day. A true measure of total effort is essential for this technique and it is only by randomizing the time of the count over the full fishing day that the effort survey can ensure an unbiased estimate of total effort.

In practice, certain factors can make full randomization of the effort survey impractical or even impossible (e.g. darkness, aircraft availability). One approach to dealing with this problem has been to use angler trips as the measure for effort. With estimators for angler trips and catch per angler trip, catch can then be estimated. McNeish and Trial (1991) used interview data to construct activity curves that would give the proportion of the day's trips that were active at each hour. A count at any time could then be used to determine the total number of trips for that day. Parker (1956) discussed a similar approach. One difficulty with this approach is the loss of independence for the effort and catch rate estimators, and the accompanying complexity of the variance estimator for catch.

This chapter considers catch estimation when the overflight surveys are scheduled at pre–chosen times. The product of a ratio estimator with an effort–related measure is also developed *i.e.* a measure of catch scaled by a count of active fishing episodes made at a non–randomly selected time. "Totalness" in the effort is now based on the randomness in the ground survey rather than in the effort survey. To accomplish this, additional information must be recorded in the ground survey, namely, the start and end times (or estimates of) of each episode sampled, so that it can be determined whether or not the episode was also counted by the effort survey.

Catch estimators and their variances, using single and multiple overflight surveys in conjunction with both access and roving designs, are developed and examined using simulated data. Also, an optimizing procedure is given for use when multiple effort counts are to be combined as a weighted average. Effects on the estimators from different patterns in the effort profiles (as caused by episodes overlapping from one day into the next or time of day fishing preferences) are also investigated. Finally, results are applied to an in-river gill net fishery on the Fraser River, Canada where performance of the standard and proposed forms of the catch estimator are compared.

3.2 Notation

T	Time period for one repetition of the fishing effort pattern.
N	Number of fishing episodes in the population.
n	Number of fishing episodes sampled. For the access survey n is fixed and known while for the roving survey n is random.
$n_t^{(O)}$	Overflight survey count of active episodes at time t *i.e.* $\sum_{j=1}^{N} \delta_j(t)$.
$n_t^{(A)}$	Access survey count of active episodes at time t *i.e.* $\sum_{i=1}^{n} \delta_i(t)$.
$n_t^{(R)}$	Roving survey count of active episodes at time t *i.e.* $\sum_{i=1}^{n} \delta_i(t)$.
C_j^*	Catch from the j^{th} fishing episode when completed, where $j = 1, \ldots, N$.
C_j	Catch from the j^{th} fishing episode at time of interview, where $j = 1, \ldots, N$. Define $C_j = 0$ if the j^{th} episode is not selected in the roving sample.
L_j^*	Length of time fished from the j^{th} fishing episode when completed, where $j = 1, \ldots, N$.
L_j	Length of time fished during the j^{th} fishing episode at time of interview, where $j = 1, \ldots, N$. Define $L_j = 0$ if the j^{th} episode is not selected in the roving sample.
C^*	Total catch $= \sum_{j=1}^{N} C_j^*$.
λ_j	Catch rate for the j^{th} episode.
$\delta_j()$	Indicator variable for the j^{th} episode used to denote inclusion in sample s or activity at time t.

As a means of distinguishing, the subscript j is used with population units while i is used with sample units.

3.3 Survey Design and Assumptions

Consider a fishery on a given fishing resource with N fishers whose fishing activities are completely independent. For simplicity, consider a single fishing day of length T where the pattern of fishing repeats from day to day. Each fisher has a fishing episode of length L_j^* for which fishing is a stationary Poisson process with parameter λ_j (*i.e.* catch rate) constant over the time of its episode producing a catch of C_j^*. The objective is to find an estimate for the total catch $C^* = \sum_{j=1}^{N} C_j^*$. A stationary Poisson process was also used by Hoenig et al. (1997) and Pollock et al. (1997) to develop catch estimators using fully randomized effort surveys.

Assume an overflight survey is scheduled for a specific time, t, during T and provides an accurate count $n_t^{(O)}$ of the fishing episodes active at that time. Such a count is viewed as "instantaneous" (Pollock et al., 1994). Moreover, $n_t^{(O)}$ is a known constant. As a variation, there may be multiple overflight surveys during T at times t_i providing $n_{t_i}^{(O)}$ for $i = 1, \ldots, m$. Values of t, which are not random, would be based on practical concerns such as aircraft availability and darkness together with an objective of producing large values of $n_t^{(O)}$.

Ground survey methods can be categorized in two ways (Pollock et al., 1994). "Access" surveys are those that sample fishers after completion of their fishing episodes, typically as they pass some point of access to the fishing resource. Information for the completed episode is immediate. In contrast, "roving" surveys are those in which the fishers are sampled while their fishing episode is in progress by an interviewer roving through the fishing resource. Information for the completed episode is not available and must be model based relying on some assumed model for catch rate. These two sample designs result in different estimators for the catch, the more complex usually associated with roving designs (Pollock et al., 1997).

If the j^{th} episode is selected, its sample inclusion indicator $\delta_i(s)$ equals 1 and 0 otherwise.

If the ground survey has an access design, assume that a simple random sample of n episodes is selected. This gives a probability of sample selection for the j^{th} episode of $P(\delta_j(s) = 1) = n/N$. The data collected are C_i^*, the catch at completion and the start and end times of the episode. From this, it is known whether or not the episode was active at the time of the overflight count and the value of $\delta_i(t)$ equals 1 if the episode was active at time t or 0 if not active. Sample activity counts are then $n_t^{(A)} = \sum_{i=1}^{n} \delta_i(t)$.

If the ground survey has a roving design, assume that for each episode the time of intercept is randomly and independently selected according to a uniform distribution over T by adopting the sampling design used by Hoenig et al. (1997) and Pollock et al. (1997) *i.e.* a random starting point and random direction of travel by the interviewer for a complete pass through the fishing resource. It is also assumed that fishers will be able to provide reliable estimates of end times as well as accurate start time information. Hence the interview provides start and end times; C_i and L_i, the catch and time fished up to time of interview; L_i^*, the length of the completed episode; and $\delta_i(t)$, the state of activity at time t for each episode. A constant catch rate, λ_j, is assumed for each episode and sample activity counts are calculated as $n_t^{(R)} = \sum_{i=1}^{n} \delta_i(t)$.

3.4 Estimators for the Access Designs

Randomness in the access survey stems from the simple random sampling of all fishing episodes. In practice, there may be further variability associated with values which require recollection on the part of the fishers such as start times and end times if sampling is not immediate. However, in the development of estimators, these are assumed to be known and reported without error.

3.4.1 A Single Overflight Survey

A natural way to produce an estimate for total catch is to expand the value for catch recorded from the access survey by the ratio of activity counts from the overflight survey to the activity counts from the access survey. An estimator for catch with an overflight survey at time t is then

$$\hat{C} = \sum_{i=1}^{n} C_i^* \frac{n_t^{(O)}}{n_t^{(A)}}. \tag{3.1}$$

By defining $\hat{R} = \sum_{i=1}^{n} C_i^* \Big/ \sum_{i=1}^{n} \delta_i(t)$, $\hat{C}$ is seen to be a ratio estimator. Note that for $\hat{C}$ to exist there must be at least one episode in the access survey, and hence the population, active at time t.

3.4.1.1 The Expected Value of $\hat{C}$

Using the standard linear approximation for $\hat{R}$ (Cochran, 1977), the expectation of $\hat{R}$ can be found as,

$$\begin{aligned} E\left[\hat{R}\right] &= E\left[\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i(t)}\right] \\ &\approx \frac{E\left[\sum_{i=1}^{n} C_i^*\right]}{E\left[\sum_{i=1}^{n} \delta_i(t)\right]} \\ &= \frac{E\left[\sum_{j=1}^{N} C_j^* \delta_j(s)\right]}{E\left[\sum_{j=1}^{N} \delta_j(t,s)\right]}. \end{aligned}$$

Evaluating the numerator:

$$\begin{aligned} E\left[\sum_{j=1}^{N} C_j^* \delta_j(s)\right] &= \sum_{j=1}^{N} \left\{ E\left[C_j^* \delta_j(s) \middle| \delta_j(s) = 0 \right] P(\delta_j(s) = 0) \right. \\ &\quad \left. +\ E\left[C_j^* \delta_j(s) \middle| \delta_j(s) = 1 \right] P(\delta_j(s) = 1) \right\} \\ &= 0 + \sum_{j=1}^{N} C_j^* \frac{n}{N}. \end{aligned}$$

Evaluating the denominator:

$$\begin{aligned} E\left[\sum_{j=1}^{N} \delta_j(t,s)\right] &= \sum_{j=1}^{N} \left\{ E\left[\delta_j(t,s) \middle| \delta_j(s) = 0 \right] P(\delta_j(s) = 0) \right. \\ &\quad \left. +\ E\left[\delta_j(t,s) \middle| \delta_j(s) = 1 \right] P(\delta_j(s) = 1) \right\} \\ &= 0 + \sum_{j=1}^{N} \delta_j(t) \frac{n}{N}. \end{aligned}$$

Therefore

$$\begin{aligned} E\left[\hat{R}\right] &= \frac{n}{N} \sum_{j=1}^{N} C_j^* \bigg/ \frac{n}{N} \sum_{j=1}^{N} \delta_j(t) \\ &= \sum_{j=1}^{N} C_j^* \bigg/ n_t^{(O)} \end{aligned}$$

and

$$\begin{aligned} E\left[\hat{C}\right] &= \frac{\sum_{j=1}^{N} C_j^*}{n_t^{(O)}} \cdot n_t^{(O)} \\ &= \sum_{j=1}^{N} C_j^* \\ &= C^*. \end{aligned}$$

showing that $\hat{C} = \sum_{i=1}^{n} C_i^* \, n_t^{(O)} / n_t^{(A)}$ is an unbiased estimator for C^* for any time t.

3.4.1.2 The Variance of $\hat{C}$

Since it is assumed that the time of the overflight survey is predetermined, $n_t^{(O)}$ is not random and therefore

$$Var\left(\hat{C}\right) = \left[n_t^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i\left(t\right)}\right).$$

Note that both C_i^* and $\delta_i\left(t\right)$ are measured on the same sample unit enabling the variance of $\hat{C}$ to be constructed using the standard results for a ratio estimator (Cochran, 1977) giving

$$Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i\left(t\right)}\right) = \frac{1-f}{\left(\overline{n_t^{(O)}}\right)^2} \frac{1}{n} S_z^2$$

where f is the sampling fraction, $\overline{n_t^{(O)}} = \left(\sum_{j=1}^{N} \delta_j\left(t\right)\right) \Big/ N$ is the population proportion of episodes active at time t, and $S_z^2 = \frac{1}{N-1} \sum_{j=1}^{N} \left(z_j\right)^2$ with $z_j = C_j^* - \left(\sum_{j=1}^{N} C_j^* \Big/ \sum_{j=1}^{N} \delta_j\left(t\right)\right) \delta_j\left(t\right)$. The variance of $\hat{C}$ can then be simplified to

$$\begin{aligned} Var\left(\hat{C}\right) &= \frac{N^2}{n}\left(1-f\right) S_z^2 \\ &= \frac{n}{f}\left(\frac{1}{f}-1\right) S_z^2. \end{aligned} \tag{3.2}$$

To form a sample estimator, it is necessary to have an estimator for the sampling fraction f. A logical candidate for this estimator is $\hat{f} = n_t^{(A)} \Big/ n_t^{(O)}$. The sample estimator then becomes

$$\widehat{Var}\left(\hat{C}\right) = n \cdot \frac{1}{\hat{f}}\left(\frac{1}{\hat{f}}-1\right) s_z^2 \tag{3.3}$$

where

$$s_z^2 \quad = \quad \frac{1}{n-1}\sum_{i=1}^{n}(z_i)^2 \text{ using } z_i = C_i^* - \frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i(t)} \cdot \delta_i(t).$$

3.4.2 Multiple Overflight Surveys

In practice, a reasonable design might be to combine, for a single T, the results of one access survey with those of two or more overflight surveys. Since ratio estimators will be involved, combining multiple results leads to a choice between a mean of ratios and a ratio of means type of estimator. To illustrate the principles involved, only two overflight surveys are used, however, the results are easily extended to the more general case.

3.4.2.1 A Mean of Ratios Estimator

With multiple overflight surveys it is possible to make multiple estimates of total catch using the catch measurement from a single access survey. A natural choice for a single measure of total catch is to weight these estimators equally and use their mean. For a design involving two overflight surveys at times t_1 and t_2, the estimator for total catch is thus

$$\hat{C}_{mor} = \frac{1}{2}\left\{\sum_{i=1}^{n} C_i^* \cdot \frac{n_{t_1}^{(O)}}{n_{t_1}^{(A)}} + \sum_{i=1}^{n} C_i^* \cdot \frac{n_{t_2}^{(O)}}{n_{t_2}^{(A)}}\right\}. \tag{3.4}$$

3.4.2.1.1 The Expected Value of $\hat{C}_{mor}$

Previously it was shown that $\hat{C} = \sum_{i=1}^{n} C_i^* \cdot \frac{n_t^{(O)}}{n_t^{(A)}}$ is an approximately unbiased estimator of total catch using a single overflight survey at any time t. It then follows that for the average of two overflight surveys at times t_1 and t_2

$$
\begin{aligned}
E\left[\hat{C}_{mor}\right] &= \frac{1}{2}\left\{E\left[\sum_{i=1}^{n} C_i^* \cdot \frac{n_{t_1}^{(O)}}{n_{t_1}^{(A)}}\right] + E\left[\sum_{i=1}^{n} C_i^* \cdot \frac{n_{t_2}^{(O)}}{n_{t_2}^{(A)}}\right]\right\} \\
&\approx \frac{1}{2}\left\{C^* + C^*\right\} \\
&= C^*,
\end{aligned}
$$

showing that $\hat{C}_{mor}$ is an unbiased estimator for C^*.

3.4.2.1.2 The Variance of $\hat{C}_{mor}$

Since $\hat{C}_{mor}$ is a linear combination of dependent random variables, its variance is formed as

$$
\begin{aligned}
Var\left(\hat{C}_{mor}\right) &= \left(\frac{1}{2}\right)^2 \left\{\left[n_{t_1}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i\left(t_1\right)}\right) + \left[n_{t_2}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i\left(t_2\right)}\right)\right. \\
&\left. + 2\left[n_{t_1}^{(O)}\right]\left[n_{t_2}^{(O)}\right] Cov\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i\left(t_1\right)}, \frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i\left(t_2\right)}\right)\right\}. \qquad (3.5)
\end{aligned}
$$

Forming a sample estimator for $Var\left(\hat{C}_{mor}\right)$ involves sample estimators for the components of Equation 3.5. Estimates for $\left[n_{t_1}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i(t_1)}\right)$ and $\left[n_{t_2}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i(t_2)}\right)$ can be found as per Equation 3.3. An estimate for $Cov\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i(t_1)}, \frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i(t_2)}\right)$ can be constructed using first order approximations as

$$
\widehat{Cov}\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i\left(t_1\right)}, \frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i\left(t_2\right)}\right) = \frac{N^2}{n_{t_1}^{(O)} \cdot n_{t_2}^{(O)}} \frac{1-\hat{f}}{n} c_z^2
$$

where

$$c_z^2 \quad = \quad \tfrac{1}{n-1} \sum_{i=1}^{n} z_{t_1,i}\, z_{t_2,i} \;\; \text{using} \;\; z_{t,i} = C_i^* - \frac{\sum_{i=1}^{n} C_i^*}{n_t^{(A)}} \cdot \delta_i\,(t) \text{ for } t_1 \text{ and } t_2.$$

A sample estimator for the final term in Equation 3.5 is then

$$2\left[n_{t_1}^{(O)} \cdot n_{t_2}^{(O)}\right] \widehat{Cov}\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i\,(t_1)}, \frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i\,(t_2)}\right) \quad = \quad 2\left[\frac{n_{t_1}^{(O)} \cdot n_{t_2}^{(O)}}{n_{t_1}^{(O)} \cdot n_{t_2}^{(O)}}\right]\left(\frac{n}{\hat{f}}\right)^2 \frac{1}{n}\left(1-\hat{f}\right) s_z^2$$

$$= \quad 2\,n\,\left(\frac{1}{\hat{f}}\right)\left(\frac{1}{\hat{f}}-1\right) c_z^2$$

where the estimator for f may be formed as $\hat{f} = \frac{1}{2}\left\{\frac{n_{t_1}^{(A)}}{n_{t_1}^{(O)}} + \frac{n_{t_2}^{(A)}}{n_{t_2}^{(O)}}\right\}$ or as $\hat{f} = \frac{n_{t_1}^{(A)}+n_{t_2}^{(A)}}{n_{t_1}^{(O)}+n_{t_2}^{(O)}}$, the latter offering protection against equal weighting of possible atypical values involving small numbers of observations that might exist should overflight times be selected near the start or end of the fishing day.

3.4.2.1.3 The General Case and Optimal Weightings

Expanding to the general case, with overflight surveys at times t_i for $i = 1, \ldots, m$ and n episodes selected for the access survey, the unbiased estimators for total catch using each overflight survey (as per Equation 3.1) can be arranged in an $m \times 1$ vector $\mathbf{C}$. The $(i,j)^{th}$ element of $\mathbf{V}$, the estimated variance-covariance matrix of $\mathbf{C}$, is formed as

$$v_{ij} = n\,\left(\frac{1}{\hat{f}}\right)\left(\frac{1}{\hat{f}}-1\right)\frac{1}{n-1}\sum_{k=1}^{n} \hat{z}_{t_i,k}\hat{z}_{t_j,k}$$

where $\hat{z}_{t,k} = C_k^* - \frac{\sum_{k=1}^{n} C_k^*}{\sum_{k=1}^{n} \delta_k(t)} \cdot \delta_k\,(t)$ for t_i and t_j and $\hat{f} = \sum_{i=1}^{m} n_{t_i}^{(A)} \Big/ \sum_{i=1}^{m} n_{t_i}^{(O)}$. If $\mathbf{w}$ is an $m \times 1$ vector of weights for each element of $\mathbf{C}$ such that $\sum_{i=1}^{m} w_i = 1$, then

$$\hat{C}_{mor} = \mathbf{w}'\mathbf{C}$$

for which $E\left[\hat{C}_{mor}\right] = \mathbf{w}'E\left[\mathbf{C}\right] = C^*$ and

$$\widehat{Var}\left(\hat{C}_{mor}\right) = \mathbf{w}'\mathbf{V}\mathbf{w}.$$

Expressed this way, it is possible to find weightings that minimize the variance of $\hat{C}_{mor}$ using the generalized Cauchy inequality (Olkin, 1958). These optimal weightings are approximated as

$$\hat{\mathbf{w}}_{\mathbf{opt}} = \frac{\mathbf{e}'\mathbf{V}^{-\mathbf{1}}}{\mathbf{e}'\mathbf{V}^{-\mathbf{1}}\mathbf{e}}$$

where $\mathbf{e}$ is an $m \times 1$ vector of 1's.

3.4.2.2 A Ratio of Means Estimator

Multiple overflight surveys also allow for a pooling of the information used to expand the measurement of catch from the access survey. This results in a ratio of means type of ratio estimator of the form

$$\hat{C}_{rom} = \sum_{i=1}^{n} C_i^* \cdot \frac{n_{t_1}^{(O)} + n_{t_2}^{(O)}}{n_{t_1}^{(A)} + n_{t_2}^{(A)}}. \tag{3.6}$$

3.4.2.2.1 The Expected Value of $\hat{C}_{rom}$

To show that $\hat{C}_{rom}$ is an unbiased estimator for total catch C^*, note that $n_{t_1}^{(O)} + n_{t_2}^{(O)}$ is not random. Then by using first order approximations

$$\begin{aligned} E\left[\sum_{i=1}^{n} C_i^* \cdot \frac{n_{t_1}^{(O)} + n_{t_2}^{(O)}}{n_{t_1}^{(A)} + n_{t_2}^{(A)}}\right] &\approx \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right] \cdot \frac{E\left[\sum_{i=1}^{n} C_i^*\right]}{E\left[\sum_{i=1}^{n} \delta_i(t_1) + \sum_{i=1}^{n} \delta_i(t_2)\right]} \\ &= \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right] \cdot \frac{\sum_{j=1}^{N} C_j^* \cdot \frac{n}{N}}{\sum_{j=1}^{N} \delta_j(t_1) \cdot \frac{n}{N} + \sum_{j=1}^{N} \delta_j(t_2) \cdot \frac{n}{N}} \end{aligned}$$

$$
\begin{aligned}
&= \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right] \cdot \frac{\sum_{j=1}^{N} C_j^*}{n_{t_1}^{(O)} + n_{t_2}^{(O)}} \\
&= \sum_{j=1}^{N} C_j^* \\
&= C^*.
\end{aligned}
$$

3.4.2.2.2 The Variance of $\hat{C}_{rom}$

The variance of $\hat{C}_{rom}$ is found by using the standard variance formula for a ratio estimator. Since $n_{t_1}^{(O)} + n_{t_2}^{(O)}$ is constant

$$
Var\left(\hat{C}_{rom}\right) = \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i(t_1) + \sum_{i=1}^{n} \delta_i(t_2)}\right) \tag{3.7}
$$

for which

$$
Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n} \delta_i(t_1) + \sum_{i=1}^{n} \delta_i(t_2)}\right) = \frac{1-f}{\left(\overline{n_{t_1}^{(O)} + n_{t_2}^{(O)}}\right)^2} \frac{1}{n} S_z^2
$$

where

$$
\begin{aligned}
\overline{n_{t_1}^{(O)} + n_{t_2}^{(O)}} &= \tfrac{1}{N} \sum_{j=1}^{N} \left[\delta_j(t_1) + \delta_j(t_2)\right] \text{ , and} \\
S_z^2 &= \tfrac{1}{N-1} \sum_{j=1}^{N} (z_j)^2 \text{ for } z_j = C_j^* - \frac{\sum_{j=1}^{N} C_j^*}{n_{t_1}^{(O)} + n_{t_2}^{(O)}} \cdot \left[\delta_j\left(t_1\right) + \delta_j\left(t_2\right)\right] .
\end{aligned}
$$

Simplifying,

$$
\begin{aligned}
Var\left(\hat{C}_{rom}\right) &= \frac{N^2}{n} (1-f) S_z^2 \\
&= \frac{n}{f} \left(\frac{1}{f} - 1\right) S_z^2.
\end{aligned}
$$

A sample estimator for $Var\left(\hat{C}_{rom}\right)$ is formed using the simplified form and substituting $s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i)^2$ for S_z^2 where $z_i = C_i^* - \frac{\sum_{i=1}^{n} C_i^*}{n_{t_1}^{(A)}+n_{t_2}^{(A)}} \cdot [\delta_i(t_1) + \delta_i(t_2)]$. Also, for f substitute $\hat{f} = \frac{1}{2}\left\{\frac{n_{t_1}^{(A)}}{n_{t_1}^{(O)}} + \frac{n_{t_2}^{(A)}}{n_{t_2}^{(O)}}\right\}$ or $\hat{f} = \frac{n_{t_1}^{(A)}+n_{t_2}^{(A)}}{n_{t_1}^{(O)}+n_{t_2}^{(O)}}$.

As presented, it is easy to see that both $Var\left(\hat{C}_{mor}\right)$ and $Var\left(\hat{C}_{rom}\right)$ decrease with increased sample sizes. It is also important to note that the variances decrease with an increase in the number of overflight surveys through changes in S_z^2.

3.4.2.3 The Relationship Between $\hat{C}_{mor}$ and $\hat{C}_{rom}$

For multiple overflight surveys, both $\hat{C}_{mor}$ and $\hat{C}_{rom}$ were shown to be unbiased, however, when extreme values occur a ratio of means estimator is, in general, more stable than a mean of ratios estimator. Simulation results will show that with optimal weightings $Var\left(\hat{C}_{mor}\right)$ is approximately equal to $Var\left(\hat{C}_{rom}\right)$ and in some instances, marginally smaller. When two overflight surveys are being considered optimal weightings for the mean of ratios estimator can be found as

$$w_1 = \frac{V_2 - Cov}{V_1 + V_2 - 2Cov} \quad \text{and} \quad w_2 = \frac{V_1 - Cov}{V_1 + V_2 - 2Cov}$$

where V_i is the variance of the catch estimator associated with the i^{th} overflight and Cov is the covariance between $\hat{C}_1$ and $\hat{C}_2$.

To find weightings for which $Var\left(\hat{C}_{mor}\right)$ equals $Var\left(\hat{C}_{rom}\right)$, choose the following weight conditional on the outcome of the access survey

$$w = \frac{\sum_{i=1}^{n}\delta_i(t_1)}{\sum_{i=1}^{n}\delta_i(t_1) + \sum_{i=1}^{n}\delta_i(t_2)} = \frac{n_{t_1}^{(A)}}{n_{t_1}^{(A)} + n_{t_2}^{(A)}}.$$

Equation 3.7 can now be written as

$$Var\left(\hat{C}_{rom}\middle| s\right) = \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]^2 Var\left(\frac{1}{2}\left\{\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_1\right)} \cdot w + \frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_2\right)} \cdot (1-w)\right\}\right)$$

for which

$$\begin{aligned} &Var\left(\frac{1}{2}\left\{\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_1\right)} \cdot w + \frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_2\right)} \cdot (1-w)\right\}\right) \\ &\quad = \left(\frac{w}{2}\right)^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_1\right)}\right) + \left(\frac{1-w}{2}\right)^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_2\right)}\right) \\ &\quad\quad +2\left(\frac{w}{2}\right)\left(\frac{1-w}{2}\right) Cov\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_1\right)}, \frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_2\right)}\right). \end{aligned}$$

Next suppose that $w = n_{t_1}^{(A)} \Big/ \left[n_{t_1}^{(A)} + n_{t_2}^{(A)}\right] = n_{t_1}^{(O)} \Big/ \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]$. Then

$$\begin{aligned} &Var\left(\hat{C}_{rom}\right) \\ &\quad = \left(\frac{1}{2}\right)^2 \left\{\left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]^2 \left[\frac{n_{t_1}^{(O)}}{n_{t_1}^{(O)} + n_{t_2}^{(O)}}\right]^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_1\right)}\right)\right. \\ &\quad\quad + \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]^2 \left[\frac{n_{t_2}^{(O)}}{n_{t_1}^{(O)} + n_{t_2}^{(O)}}\right]^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_2\right)}\right) \\ &\quad\quad \left. + 2\left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]^2 \cdot \frac{n_{t_1}^{(O)} \cdot n_{t_2}^{(O)}}{\left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]^2} \cdot Cov\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_1\right)}, \frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_2\right)}\right)\right\} \end{aligned}$$

$$\approx \left(\frac{1}{2}\right)^2 \left\{ \left[n_{t_1}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_1\right)}\right) + \left[n_{t_2}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_2\right)}\right) \right.$$
$$\left. + 2\left[n_{t_1}^{(O)}\right]\left[n_{t_2}^{(O)}\right] Cov\left(\frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_1\right)}, \frac{\sum_{i=1}^{n} C_i^*}{\sum_{i=1}^{n}\delta_i\left(t_2\right)}\right)\right\}.$$

That is, if the overflight counts are proportional to the population counts of active fishing episodes, *i.e.* $w = n_{t_1}^{(A)} \Big/ \left[n_{t_1}^{(A)} + n_{t_2}^{(A)}\right] = n_{t_1}^{(O)} \Big/ \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]$, then using weights which are ratios of the associated overflight count to the total of all overflight counts will result in $Var\left(\hat{C}_{mor}\right)$ equal to $Var\left(\hat{C}_{rom}\right)$.

3.5 Estimators for the Roving Designs

For the development of the estimators for use with the roving survey, it is assumed that the sampling design is that of a random start, time, and direction for the pass through the resource. That is, for the j^{th} episode in the population, there is a potential interview time randomly chosen between 0 and T. As noted by Hoenig et al. (1997) and Pollock et al. (1997) the probability of selection for any episode is $P(\delta_j(s) = 1) = L_j^*/T$ and the $\delta_j(s)$ are Bernoulli random variables with $E[\delta_j(s)] = L_j^*/T$ and $Var(\delta_j(s)) = (L_j^*/T)(1 - L_j^*/T)$. Then, since L_j, the time of intercept is Uniform over L_j^*, given selection, $E[L_j|L_j^*, \delta_j(s) = 1] = L_j^*/2$ and $Var(L_j|L_j^*, \delta_j(s) = 1) = L_j^{*2}/12$. Also, by assuming a constant catch rate for each episode it is possible to estimate C_i^* as $C_i(L_i^*/L_i)$. By assuming a stationary homogeneous Poisson process, for the j^{th} episode, given that it is selected

$$P(C_j|\, L_j) = \frac{(\lambda_j L_j)^{C_j} e^{-\lambda_j L_j}}{C_j\,!}$$

and

$$P\big(C_j^*|\, L_j^*\big) = \frac{\left(\lambda_j L_j^*\right)^{C_j^*} e^{-\lambda_j L_j^*}}{C_j^*\,!}.$$

Also, given the catch at some interim time,

$$P\big(C_j^*|\, C_j, L_j, L_j^*\big) = \frac{\left[\lambda_j \left(L_j^* - L_j\right)\right]^{C_j^*-C_j} e^{-\lambda_j \left(L_j^*-L_j\right)}}{\left(C_j^* - C_j\right)!}.$$

It then follows that

$$P\big(C_j|\, L_j, C_j^*, L_j^*\big) = \left(\begin{array}{c} C_j^* \\ C_j \end{array}\right) \left(\frac{L_j}{L_j^*}\right)^{C_j} \left(1 - \frac{L_j}{L_j^*}\right)^{C_j^*-C_j}.$$

Note: While $2L_j$, the unbiased estimator for L_j^* is available, it is assumed that L_j^* is also known (Section 3.3) and is used in subsequent sample quantities.

3.5.1 A Single Overflight Survey

As with designs in which the overflight survey is random, the form of estimators appropriate for use with an access survey are, in general, not of the same form as those that are appropriate for use with a roving design. In Section 3.4.1 it was shown that $\hat{R} = \sum_{i=1}^{n} C_i^* \Big/ \sum_{i=1}^{n} \delta_i(t)$, or equivalently $\sum_{j=1}^{N} C_j^* \delta_j(s) \Big/ \sum_{j=1}^{N} \delta_j(t,s)$, was appropriate for use with an access survey when the overflight survey is not random. It is instructive to demonstrate that this form of the estimator is not appropriate for use with a roving survey where catch and length of episode are observed only up to the time of interview. Under a constant catch rate assumption with knowledge of L_j^*, the roving equivalent for this form is one which estimates C_j^* with $C_j L_j^*/L_j$. Hence for any sample s and intercept time t,

$$
\begin{aligned}
E\left[\sum_{j=1}^{N} C_j \frac{L_j^*}{L_j}\delta_j(s)\right] &= \sum_{j=1}^{N}\left\{E\left[C_j \frac{L_j^*}{L_j}\delta_j(s)\middle|\,\delta_j(s)=0\right]P(\delta_j(s)=0)\right. \\
&\qquad \left. + E\left[C_j \frac{L_j^*}{L_j}\delta_j(s)\middle|\,\delta_j(s)=1\right]P(\delta_j(s)=1)\right\} \\
&= 0+\sum_{j=1}^{N} E\left[C_j\frac{L_j^*}{L_j}\right]\frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} E_{L_j}\left[\frac{L_j^*}{L_j}E_{C_j}\left[C_j|\,L_j, C_j^*, L_j^*\right]\right]\frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} E_{L_j}\left[\frac{L_j^*}{L_j}C_j^*\frac{L_j}{L_j^*}\right]\frac{L_j^*}{T} \\
&= \sum_{j=1}^{N}\frac{C_j^* L_j^*}{T}
\end{aligned}
$$

and

$$
\begin{aligned}
E\left[\sum_{j=1}^{N}\delta_j(s,t)\right] &= \sum_{j=1}^{N}\left\{E\left[\delta_j(t,s)|\,\delta_j(s)=0\right]P(\delta_j(s)=0)\right. \\
&\qquad \left. + E\left[\delta_j(t,s)|\,\delta_j(s)=1\right]P(\delta_j(s)=1)\right\} \\
&= 0+\sum_{j=1}^{N}\delta_j(t)\frac{L_j^*}{T}.
\end{aligned}
$$

That is, the expected value of this ratio approximately equals $\left(\sum_{j=1}^{N} C_j^* L_j^*\right)\Big/\left(\sum_{j=1}^{N}\delta_j(t)L_j^*\right)$ which is not equal to the required $C^*/\,n_t^{(O)}$.

Alternatively, a proposed estimator is $\hat{R}=\sum_{i=1}^{n} C_i\frac{L_i^*}{L_i}/L_i^*\Big/\sum_{i=1}^{n}\delta_i(t)/L_i^*$. This leads to the catch estimator

$$
\hat{C}=\sum_{i=1}^{n}\frac{C_i}{L_i}\cdot\frac{n_t^{(O)}}{\sum_{i=1}^{n}\frac{\delta_i(t)}{L_i^*}} \tag{3.8}
$$

which exists only if at least one episode in the roving survey is active at time t. When working with this estimator it is important to note that here, the sample size n, as well as C_i and L_i are random quantities. It is also important to note that determining the value of $\delta_i(t)$, the activity indicator for the i^{th} episode at time t, requires L_i^* or its estimate if t is greater than the time of interview.

3.5.1.1 The Expected Value of $\hat{C}$

Again assuming that the expectation of the ratio can be approximated by the ratio of the expectations,

$$E\left[\hat{R}\right] = E\left[\frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t)}{L_i^*}}\right] \approx \frac{E\left[\sum_{i=1}^{n} \frac{C_i}{L_i}\right]}{E\left[\sum_{i=1}^{n} \frac{\delta_i(t)}{L_i^*}\right]}.$$

Examining the numerator and denominator separately,

$$\begin{aligned}
E\left[\sum_{i=1}^{n} \frac{C_i}{L_i}\right] &= E\left[\sum_{j=1}^{N} \frac{C_i}{L_i} \delta_j(s)\right] \\
&= E\left[\sum_{j=1}^{N} \left\{ \left. \frac{C_j}{L_j} \right| \delta_j(s) = 1 \right\} \cdot P\left(\delta_j(s) = 1\right)\right] \\
&= \sum_{j=1}^{N} E_{L_j}\left[\frac{1}{L_j} E_{C_j}\left[C_j | \, L_j, C_j^*, L_j^* \right]\right] \frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} E_{L_j}\left[\frac{1}{L_j} C_j^* \frac{L_j}{L_j^*}\right] \frac{L_j^*}{T} \\
&= \sum_{j=1}^{N} \frac{C_j^*}{T} \\
&= \frac{C^*}{T}
\end{aligned}$$

and

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}\frac{\delta_i(t)}{L_i^*}\right] &= E\left[\sum_{j=1}^{N}\frac{\delta_i(t)}{L_i^*}\delta_j(t,s)\right] \\
&= E\left[\sum_{j=1}^{N}\left\{\left.\frac{\delta_j(t,s)}{L_j^*}\right|\delta_j(t,s)=1\right\}\cdot P\left(\delta_j(t,s)=1\right)\right] \\
&= \sum_{j=1}^{N} E\left[\frac{\delta_j(t)}{L_j^*}\cdot\frac{L_j^*}{T}\right] \\
&= \sum_{j=1}^{N} E\left[\frac{\delta_j(t)}{T}\right] \\
&= \frac{\sum_{j=1}^{N}\delta_j(t)}{T} \\
&= \frac{n_t^{(O)}}{T}.
\end{aligned}
$$

This then gives

$$
E\left[\hat{R}\right] \approx \frac{C^*/T}{n_t^{(O)}/T} = \frac{C^*}{n_t^{(O)}}.
$$

Therefore $E\left[\hat{C}\right] = n_t^{(O)}E\left[\hat{R}\right] \approx n_t^{(O)}\cdot C^*/n_t^{(O)} = C^*$, showing that $\hat{C}$ is an unbiased estimator for the total catch given any overflight time t.

3.5.1.2 The Variance of $\hat{C}$

Under the roving design, the variance of $\hat{C}$ is complicated by having the sample inclusion probabilities vary as the duration of the individual episodes. The design is a variant of a multistage design in which episodes are primary units and interview times are subunits, both determined by the same random time of intercept. As such, s, the set of episodes selected and its size, n, are random quantities. Therefore, conditioning on s with $\left[n_t^{(O)}\right]$

constant

$$Var\left(\hat{C}\right) = \left[n_t^{(O)}\right]^2 \left\{E\left[Var\left(\hat{R}\middle| s\right)\right] + Var\left(E\left[\hat{R}\middle| s\right]\right)\right\}.$$

The within episode variance can be found by first noting that episodes are assumed independent and given s, then $\delta_i(t)$ is constant leading to

$$Var\left(\hat{R}\middle| s\right) = \frac{\sum_{i=1}^{n} Var\left(\frac{C_i}{L_i}\middle| s\right)}{\left(\sum_{i=1}^{n} \frac{\delta_i(t)}{L_i^*}\right)^2}$$

for which

$$\begin{aligned}
Var\left(\frac{C_i}{L_i}\middle| s\right) &= Var_{L_i}\left(\frac{E_{C_i}\left[C_i \middle| L_i\right]}{L_i}\right) + E_{L_i}\left[\frac{1}{L_i^2} Var_{C_i}(C_i | L_i)\right] \\
&= Var_{L_i}\left(\frac{C_i^*\left(\frac{L_i}{L_i^*}\right)}{L_i}\right) + E_{L_i}\left[\frac{1}{L_i^2} C_i^*\left(\frac{L_i}{L_i^*}\right)\left(1 - \frac{L_i}{L_i^*}\right)\right] \\
&= Var_{L_i}\left(\frac{C_i^*}{L_i^*}\right) + E_{L_i}\left[\frac{1}{L_i^2} C_i^*\left(\frac{L_i}{L_i^*} - \frac{{L_i}^2}{L_i^{*2}}\right)\right] \\
&= 0 + E_{L_i}\left[\frac{C_i^*}{L_i^*}\left(\frac{1}{L_i} - \frac{1}{L_i^*}\right)\right] \qquad \text{See Note} \\
&\approx \frac{C_i^*}{L_i^*}\left(\frac{1}{L_i^*/2} - \frac{1}{L_i^*}\right) \\
&= \frac{C_i^*}{L_i^{*2}}.
\end{aligned}$$

Note: Since L_i is uniform over $(0\,, L_i^*]$, values can approach 0 resulting in extreme values for $E\left[1/L_i\right] = \frac{1}{L_i^* - \epsilon}\left\{log(L_i^*) - log(\epsilon)\right\}$ for ϵ arbitrarily close to 0. Hence $Var\left(\hat{R} \middle| s\right)$ is unstable. One strategy to overcome this difficulty is to require a minimum time to interview (*i.e.* on L_i) for sample inclusion (Hoenig et al., 1997; Pollock et al., 1997). Here it is assumed that L_i is suitably bounded from 0 and a first order Taylor approximation can be used.

Then, approximating the expectation of a ratio with a ratio of expectations and passing the expectation operator through the square then gives

$$\begin{aligned}
E_s\left[Var\left(\hat{R}\middle|\, s\right)\right] &\approx \frac{E\left[\sum_{i=1}^{n}\frac{C_i^*}{L_i^{*2}}\right]}{E\left[\left(\sum_{i=1}^{n}\frac{\delta_i(t)}{L_i^*}\right)^2\right]} \\
&< \frac{E\left[\sum_{i=1}^{n}\frac{C_i^*}{L_i^{*2}}\right]}{\left(E\left[\sum_{i=1}^{n}\frac{\delta_i(t)}{L_i^*}\right]\right)^2} \\
&= \frac{\sum_{j=1}^{N}\frac{C_j^*}{L_j^{*2}}E\left[\delta_j\left(s\right)\right]}{\left(\sum_{j=1}^{N}\frac{\delta_j(t)}{L_j^*}E\left[\delta_j\left(s\right)\right]\right)^2} \\
&= \frac{\sum_{j=1}^{N}\frac{C_j^*}{L_j^{*2}}\frac{L_j^*}{T}}{\left(\sum_{j=1}^{N}\frac{\delta_j(t)}{L_j^*}\frac{L_j^*}{T}\right)^2} \\
&= \frac{T\sum_{j=1}^{N}\frac{C_j^*}{L_j^*}}{\left[n_t^{(O)}\right]^2}.
\end{aligned}$$

To find the between episode variance, note that given s

$$\begin{aligned}
E\left[\sum_{i=1}^{n}\frac{C_i}{L_i}\right] &= \sum_{i=1}^{n}E_{L_i}\left[\frac{1}{L_i}E_{C_i}\left[C_i\middle|\, L_i\right]\right] \\
&= \sum_{i=1}^{n}E_{L_i}\left[\frac{1}{L_i}C_i^*\frac{L_i}{L_i^*}\right] \\
&= \sum_{i=1}^{n}\frac{C_i^*}{L_i^*}
\end{aligned}$$

while $\delta_i\left(t\right)$ is constant. Then,

$$Var_s\left(E\left[\hat{R}\middle|\, s\right]\right) = Var_s\left(\frac{\sum_{i=1}^{n}\frac{C_i^*}{L_i^*}}{\sum_{i=1}^{n}\frac{\delta_i(t)}{L_i^*}}\right) = Var_s\left(\frac{\sum_{i=1}^{n}\frac{C_i^*}{L_i^*/T}}{\sum_{i=1}^{n}\frac{\delta_i(t)}{L_i^*/T}}\right)$$

which is the variance of the ratio of two Horvitz-Thompson estimators, $\hat{\tau}_y$ and $\hat{\tau}_x$ with selection probabilities $\pi_i = L_i^*/T$. Viewed this way,

$$E\left[\hat{\tau}_y\right] = \sum_{j=1}^{N} \frac{C_j^*}{L_j^*/T} P\left(\delta_j\left(s\right)=1\right) = \sum_{j=1}^{N} \frac{C_j^*}{L_j^*/T}\frac{L_j^*}{T} = \sum_{j=1}^{N} C_j^* = C^*$$

and

$$E\left[\hat{\tau}_x\right] = \sum_{j=1}^{N} \frac{\delta_j\left(t\right)}{L_j^*/T} P\left(\delta_j\left(s\right)=1\right) = \sum_{j=1}^{N} \frac{\delta_j\left(t\right)}{L_j^*/T}\frac{L_j^*}{T} = \sum_{j=1}^{N} \delta_j\left(t\right) = n_t^{(O)}$$

to give $\hat{R} = \hat{\tau}_y/\hat{\tau}_x$ and $R = \tau_y/\tau_x = C^*/n_t^{(O)}$ and

$$\hat{R} - R \approx \frac{\hat{\tau}_y - R\hat{\tau}_x}{\tau_x} = \frac{1}{n_t^{(O)}} \sum_{i=1}^{n} \frac{C_i^* - R\,\delta_i\left(t\right)}{\pi_i}$$

which is again a Horvitz-Thompson estimator (Thompson, 1992). Denoting $z_i = C_i^* - R\,\delta_i\left(t\right)$, the usual Horvitz-Thompson variance estimator is

$$Var\left(\hat{R}\right) = Var\left(\hat{R} - R\right) = \frac{1}{\left[n_t^{(O)}\right]^2} \left\{ \sum_{i=1}^{N} \left(\frac{1-\pi_i}{\pi_i}\right) {z_i}^2 + \sum_{i=1}^{N} \sum_{j \neq i} \left(\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}\right) z_i z_j \right\}.$$

However, since fishing episodes are assumed to be independent, the joint inclusion probability of any two episodes is the product of their individual inclusion probabilities $i.e. \pi_{ij} = \pi_i\pi_j$ and the covariance portion of the variance formula vanishes. Therefore

$$Var_s\left(E\left[\hat{R}\middle| s\right]\right) = \frac{1}{\left[n_t^{(O)}\right]^2} \sum_{j=1}^{N} \left(\frac{1-\pi_j}{\pi_j}\right) {z_j}^2.$$

Finally, combining within and between components

$$Var\left(\hat{C}\right) \approx T \sum_{j=1}^{N} \frac{C_j^*}{L_j^*} + \sum_{j=1}^{N} \left(\frac{1-\pi_j}{\pi_j}\right) \left(C_j^* - R\,\delta_j\left(t\right)\right)^2. \tag{3.9}$$

To develop a sample estimator for $Var\left(\hat{C}\right)$, estimators are found for each component. For the within variance write $\sum_{j=1}^{N} C_j^*/L_j^* = N\left(\sum_{j=1}^{N} C_j^*/L_j^*\right)\Big/N$ for which $\left(\sum_{j=1}^{N} C_j^*/L_j^*\right)\Big/N$

can be estimated with $\left(\sum_{i=1}^{n} C_i \frac{L_i^*}{L_i} / L_i^*\right) \Big/ n = \left(\sum_{i=1}^{n} C_i / L_i\right) \Big/ n$. To find an estimator for N, note that each episode is selected with probability $\frac{L_i^*}{T}$. It can then be argued that for each episode selected, there are $1 \Big/ \frac{L_i^*}{T}$ other episodes of equal length in the population. An estimate for N would then be $\hat{N} = \sum_{i=1}^{n} \frac{T}{L_i^*}$.

For the between episode variance, note that for any s,

$$E\left[C_i \frac{L_i^*}{L_i}\bigg| s\right] = E_{L_i}\left[\frac{L_i^*}{L_i} E_{C_i}\left[C_i | L_i\right]\right] = E_{L_i}\left[\frac{L_i^*}{L_i} C_i^* \left(\frac{L_j}{L_j^*}\right)\right] = C_i^*.$$

Then, with $\hat{\tau}_y = T \sum_{i=1}^{n} C_i \frac{L_i^*}{L_i} / L_i^* = T \sum_{i=1}^{n} C_i / L_i$ and $\hat{\tau}_x = T \sum_{i=1}^{n} \delta_i\left(t\right) / L_i^*$,

$$E\left[\hat{\tau}_y\right] = E_s\left[\sum_{i=1}^{n} \frac{E\left[C_i \frac{L_i^*}{L_i}\right]}{L_i^*/T}\Bigg| s\right] = \sum_{j=1}^{N} \frac{C_j^*}{L_j^*/T} P\left(\delta_j\left(s\right) = 1\right) = \sum_{j=1}^{N} \frac{C_j^*}{L_j^*/T} \frac{L_j^*}{T} = \sum_{j=1}^{N} C_j^* = C^*$$

and

$$E\left[\hat{\tau}_x\right] = \sum_{j=1}^{N} \frac{\delta_j\left(t\right)}{L_j^*/T} P\left(\delta_j\left(s\right) = 1\right) = \sum_{j=1}^{N} \frac{\delta_j\left(t\right)}{L_j^*/T} \frac{L_j^*}{T} = \sum_{j=1}^{N} \delta_j\left(t\right) = n_t^{(O)}$$

which preserves $\hat{R} = \hat{\tau}_y / \hat{\tau}_x$ and $R = \tau_y / \tau_x$ now giving

$$\hat{R} - R \approx \frac{\hat{\tau}_y - R\hat{\tau}_x}{\tau_x} = \frac{1}{n_t^{(O)}} \sum_{i=1}^{n} \frac{C_i \frac{L_i^*}{L_i} - R\,\delta_i\left(t\right)}{\pi_i} = \frac{1}{n_t^{(O)}} \sum_{i=1}^{n} \frac{z_i}{\pi_i}$$

for which the usual Horvitz-Thompson variance formula applies and where again the covariance term is dropped due to the assumed independence of the episodes. Substituting $\hat{R}$ for R, the sample estimator for this variance is then

$$\widehat{Var}\left(\hat{R}\right) = \frac{1}{\left[n_t^{(O)}\right]^2} \sum_{i=1}^{n} \left(\frac{1 - \pi_i}{\pi_i^2}\right) \hat{z}_i^2.$$

Combining components, the sample estimator for the variance of $\hat{C}$ becomes

$$\widehat{Var}\left(\hat{C}\right) = \frac{\hat{N}}{n} \left(T \sum_{i=1}^{n} \frac{C_i}{L_i}\right) + \sum_{i=1}^{n} \left(\frac{1 - \pi_i}{\pi_i^2}\right) \hat{z}_i^2 \qquad (3.10)$$

where

$$\pi_i = L_i^*/T \ ,$$

$$\hat{N} = \sum_{i=1}^{n} T/L_i^* \ , \text{ and}$$

$$\hat{z}_i = C_i \frac{L_i^*}{L_i} - \hat{R}\delta_i(t) \text{ with } \hat{R} = \left(\sum_{i=1}^{n} \frac{C_i}{L_i}\right) \Big/ \left(\sum_{i=1}^{n} \frac{\delta_i(t)}{L_i^*}\right) .$$

3.5.2 Multiple Overflight Surveys

As with access designs, the need to combine the results of two or more non random overflight surveys with a single, random roving survey is a scenario that might arise in practice. Again there is a choice between a mean of ratios and a ratio of means type of estimator for total catch.

3.5.2.1 A Mean of Ratios Estimator

To form the mean of ratios type of estimator, the total catch estimators using the single roving survey in conjunction with each of the overflight surveys at times t_1 and t_2 are combined as a simple average. This results in the following estimator,

$$\hat{C}_{mor} = \frac{1}{2} \left\{ \frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*}} \cdot n_{t_1}^{(O)} + \frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}} \cdot n_{t_2}^{(O)} \right\} . \tag{3.11}$$

3.5.2.1.1 The Expected Value of $\hat{C}_{mor}$

Since the estimators for total catch at times t_1 and t_2 are approximately unbiased, the expected value of $\hat{C}_{mor}$, which is a linear combination of these estimators, is also approximately unbiased. Further, since the unconditional expectation equal the conditional expectation for each estimator, $E[\hat{C}_{mor}]$ equals $E[\hat{C}_{mor}|n]$. That is

$$\begin{aligned} E\left[\hat{C}_{mor}\right] &= \frac{1}{2}\left\{E\left[\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}}\cdot n_{t_1}^{(O)}\right]+E\left[\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\cdot n_{t_2}^{(O)}\right]\right\} \\ &\approx \frac{1}{2}\left\{C^*+C^*\right\} \\ &= C^*. \end{aligned}$$

3.5.2.1.2 The Variance of $\hat{C}_{mor}$

Since the total catch estimators at times t_1 and t_2 are not independent, the variance of $\hat{C}_{mor}=\frac{1}{2}\left(\hat{C}_1+\hat{C}_2\right)$ conditional on any roving survey of size n is given by

$$\begin{aligned} Var\left(\hat{C}_{mor}\right) &= \left(\frac{1}{2}\right)^2\left\{\left[n_{t_1}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}}\right)+\left[n_{t_2}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\right)\right. \\ &\left. \quad + 2\left[n_{t_1}^{(O)}\cdot n_{t_2}^{(O)}\right] Cov\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}},\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\right)\right\}. \end{aligned} \tag{3.12}$$

A sample estimator for the unconditioned variance of $\hat{C}_{mor}$ can be formed by using unconditioned estimators for each of the components in Equation 3.12. That is, use Equation 3.10 to find the unconditioned estimators for $Var\left(\hat{C}_1\right)=\left[n_{t_1}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}}\right)$ and $Var\left(\hat{C}_2\right)=\left[n_{t_2}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\right)$. An estimator for $Cov\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}},\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\right)$ can be formed by making a Horvitz-Thompson type adaptation to the first order covariance approximations for use with multivariate ratio estimators under equal inclusion probabilities. Also, note that for a given roving sample, the variance for any ratio $\hat{R}_t$ associated with overflight t has a within episode component that is constant, being the product of $n_t^{(O)}$ and a function of catch and time fished values common to other $\hat{R}_t$. As such, given the roving sample, the covariance between any two $\hat{R}_t$ involves only the between episode variability. This then gives as an estimator

$$\widehat{Cov}\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}}, \frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\right) = \sum_{i=1}^{n}\left(\frac{1-\pi_i}{\pi_i^2}\right)\frac{\hat{z}_{t_1,i}}{n_{t_1}^{(O)}}\frac{\hat{z}_{t_2,i}}{n_{t_2}^{(O)}}$$

and, for the final term in the estimator for the variance of $\hat{C}_{mor}$

$$\begin{aligned} 2\left[n_{t_1}^{(O)}\cdot n_{t_2}^{(O)}\right]\widehat{Cov}\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}}, \frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\right) &= 2\left[\frac{n_{t_1}^{(O)}\cdot n_{t_2}^{(O)}}{n_{t_1}^{(O)}\cdot n_{t_2}^{(O)}}\right]\sum_{i=1}^{n}\left(\frac{1-\pi_i}{\pi_i^2}\right)\hat{z}_{t_1,i}\hat{z}_{t_2,i} \\ &= 2\sum_{i=1}^{n}\left(\frac{1-\pi_i}{\pi_i^2}\right)\hat{z}_{t_1,i}\hat{z}_{t_2,i} \end{aligned}$$

where

$$\begin{aligned} \pi_i &= L_i^*/T \text{ , and} \\ \hat{z}_{t,i} &= C_i\frac{L_i^*}{L_i} - \frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t)}{L_i^*}}\cdot\delta_i(t) \text{ for } t_1 \text{ and } t_2\,. \end{aligned}$$

3.5.2.1.3 The General Case and Optimal Weightings

For the general case with overflight surveys selected at times t_i for $i = 1, \ldots, m$ and n episodes selected for the single roving survey, the unbiased estimators for total catch corresponding to each overflight survey (as per Equation 3.10) can be arranged in the $m \times 1$ vector $\mathbf{C}$. Elements of $\mathbf{V}$, the estimated variance-covariance matrix of $\mathbf{C}$, have first order approximations

$$v_{i,j} = \sum_{k=1}^{n} \left(\frac{1-\pi_k}{\pi_k^2}\right) \hat{z}_{t_i,k}\hat{z}_{t_j,k} \quad \text{where} \quad \hat{z}_{t,k} = C_k \frac{L_k^*}{L_k} - \frac{\sum\limits_{k=1}^{n} \frac{C_k}{L_k}}{\sum\limits_{k=1}^{n} \frac{\delta_k(t)}{L_k^*}} \cdot \delta_k(t) \text{ for } t_i \text{ and } t_j.$$

If $\mathbf{w}$ is an $m \times 1$ vector of weightings for each element of $\mathbf{C}$ such that $\sum\limits_{i=1}^{m} w_i = 1$, then $\hat{C}_{mor} = \mathbf{w}'\mathbf{C}$ and $E\left[\hat{C}_{mor}\right] = \mathbf{w}'E[\mathbf{C}] = C^*$. Also, $\widehat{Var}\left(\hat{C}_{mor}\right) = \mathbf{w}'\mathbf{V}\mathbf{w}$. Optimal weightings to minimize $Var\left(\hat{C}_{mor}\right)$ can be constructed using the generalized Cauchy inequality (Olkin, 1958). Approximations to these weightings can be found as

$$\hat{\mathbf{w}}_{\mathbf{opt}} = \frac{\mathbf{e}'\mathbf{V}^{-1}}{\mathbf{e}'\mathbf{V}^{-1}\mathbf{e}}$$

where $\mathbf{e}$ is an $m \times 1$ vector of 1's.

3.5.2.2 A Ratio of Means Estimator

As with the access survey, the ratio of means estimator pools information on activity counts to construct a single expansion factor for use with the value of catch obtained from the roving survey. Such an estimator has the form

$$\hat{C}_{rom} = \sum_{i=1}^{n} \frac{C_i}{L_i} \cdot \frac{n_{t_1}^{(O)} + n_{t_2}^{(O)}}{\sum\limits_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*} + \sum\limits_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}}. \tag{3.13}$$

3.5.2.2.1 The Expected Value of $\hat{C}_{rom}$

Using first order approximations and noting that $n_{t_1}^{(O)} + n_{t_2}^{(O)}$ is a constant, then, conditional on any roving survey of size n

$$E\left[\sum_{i=1}^{n} \frac{C_i}{L_i} \cdot \frac{n_{t_1}^{(O)} + n_{t_2}^{(O)}}{\sum\limits_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*} + \sum\limits_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}}\right] \approx \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right] \cdot \frac{E\left[\sum\limits_{i=1}^{n} \frac{C_i}{L_i}\right]}{E\left[\sum\limits_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*} + \sum\limits_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}\right]}$$

$$
\begin{aligned}
&= \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right] \cdot \frac{\sum_{j=1}^{N} E_{L_j}\left[\frac{1}{L_j} E_{C_j}\left[C_j \middle| L_j, C_j^*, L_j^*\right]\right] \frac{L_j^*}{T}}{\sum_{j=1}^{N} E\left[\frac{\delta_j(t_1)}{L_j^*}\right] \frac{L_j^*}{T} + \sum_{j=1}^{N} E\left[\frac{\delta_j(t_2)}{L_j^*}\right] \frac{L_j^*}{T}} \\
&= \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right] \cdot \frac{\sum_{j=1}^{N} E_{L_j}\left[\frac{1}{L_j} C_j^* \frac{L_j}{L_j^*}\right] \frac{L_j^*}{T}}{\sum_{j=1}^{N} E\left[\frac{\delta_j(t_1)}{L_j^*}\right] \frac{L_j^*}{T} + \sum_{j=1}^{N} E\left[\frac{\delta_j(t_2)}{L_j^*}\right] \frac{L_j^*}{T}} \\
&= \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right] \cdot \frac{\sum_{j=1}^{N} \frac{C_j^*}{L_j^*} \cdot \frac{L_j^*}{T}}{\sum_{j=1}^{N} \frac{\delta_j(t_1)}{L_j^*} \cdot \frac{L_j^*}{T} + \sum_{j=1}^{N} \frac{\delta_j(t_2)}{L_j^*} \cdot \frac{L_j^*}{T}} \\
&= \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right] \cdot \frac{\sum_{j=1}^{N} \frac{{C_j}^*}{T}}{\sum_{j=1}^{N} \frac{\delta_j(t_1)}{T} + \sum_{j=1}^{N} \frac{\delta_j(t_2)}{T}} \\
&= \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right] \cdot \frac{C^*}{n_{t_1}^{(O)} + n_{t_2}^{(O)}} \\
&= C^*
\end{aligned}
$$

which is independent of n showing that $\hat{C}_{rom}$ is an unbiased estimator for C^* and that $E\left[\hat{C}_{rom}\middle| n\right]$ and $E\left[\hat{C}_{rom}\right]$ are equal.

3.5.2.2.2 The Variance of $\hat{C}_{rom}$

With $\left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]$ constant,

$$
Var\left(\hat{C}_{rom}\right) = \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*} + \sum_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}}\right) . \tag{3.14}
$$

Also, as with a single overflight, the roving portion of the design is multistage and

$$Var\left(\hat{C}_{rom}\right) = \left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]^2 \left\{E_n\left[Var\left(\hat{R}\middle| s\right)\right] + Var_n\left(E\left[\hat{R}\middle| s\right]\right)\right\}.$$

To find a sample estimator for the within episode portion of this variance, recall that

$$Var\left(\frac{C_i}{L_i}\middle| s\right) \approx \frac{C_i^*}{L_i^{*2}}$$

and

$$E_s\left[\sum_{i=1}^{n} \frac{C_i^*}{L_i^{*2}}\right] = \sum_{j=1}^{N} \frac{C_j^*}{L_j^*/T}.$$

Then, as was done in the single overflight, passing the expected value operator through the square gives

$$\begin{aligned} E_s\left[\left(\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*} + \sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}\right)^2\right] &> \left(E_s\left[\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}\right] + E_s\left[\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}\right]\right)^2 \\ &= \left(\sum_{j=1}^{N}\frac{\delta_j(t_1)}{L_j^*}E\left[\delta_j\left(s\right)\right] + \sum_{j=1}^{N}\frac{\delta_j(t_2)}{L_j^*}E\left[\delta_j\left(s\right)\right]\right)^2 \\ &= \left(\sum_{j=1}^{N}\frac{\delta_j(t_1)}{L_j^*}\frac{L_i^*}{T} + \sum_{j=1}^{N}\frac{\delta_j(t_2)}{L_j^*}\frac{L_i^*}{T}\right)^2 \\ &= \left[n_{t_1}^{(O)}/T + n_{t_2}^{(O)}/T\right]^2. \end{aligned}$$

This leads to

$$\begin{aligned} E_s\left[Var\left(\hat{R}\middle| s\right)\right] &= \frac{\sum_{j=1}^{N}\frac{C_j^*}{L_j^*/T}}{\left[n_{t_1}^{(O)}/T + n_{t_2}^{(O)}/T\right]^2} \\ &= \frac{TN\left(\sum_{j=1}^{N}\frac{C_j^*}{L_j^*} \Big/ N\right)}{\left[n_{t_1}^{(O)} + n_{t_2}^{(O)}\right]^2} \end{aligned}$$

and a sample estimator for the within episode variance can be formed by substituting the estimator $\hat{N}\left(\sum_{i=1}^{n} C_i/L_i\right)\Big/n$ for $N\left(\sum_{j=1}^{N} C_j^*/L_j^*\right)\Big/N$ where again $\hat{N}=\sum_{i=1}^{n} T/L_i^*$.

To find a sample estimator for the between episode variance, again cast $\hat{R}$ as a Horvitz-Thompson estimator. That is, with

$$\hat{R}=\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}+\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}=\frac{\sum_{i=1}^{n}\frac{C_i\frac{L_i^*}{L_i}}{L_i^*/T}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*/T}+\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*/T}}=\frac{\sum_{i=1}^{n}\frac{C_i\frac{L_i^*}{L_i}}{\pi_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{\pi_i}+\sum_{i=1}^{n}\frac{\delta_i(t_2)}{\pi_i}}$$

where

$$E\left[\sum_{i=1}^{n}\frac{C_i\frac{L_i^*}{L_i}}{\pi_i}\right]=C^* \quad \text{and} \quad E\left[\sum_{i=1}^{n}\frac{\delta_i(t_1)}{\pi_i}+\sum_{i=1}^{n}\frac{\delta_i(t_2)}{\pi_i}\right]=n_{t_1}^{(O)}+n_{t_2}^{(O)}\,,$$

it follows that $R=C^*\Big/\left[n_{t_1}^{(O)}+n_{t_2}^{(O)}\right]$ and

$$\hat{R}-R\approx\frac{1}{n_{t_1}^{(O)}+n_{t_2}^{(O)}}\sum_{i=1}^{n}\frac{C_i\frac{L_i^*}{L_i}-R\left[\delta_i\left(t_1\right)+\delta_i\left(t_2\right)\right]}{\pi_i}=\frac{1}{n_{t_1}^{(O)}+n_{t_2}^{(O)}}\sum_{i=1}^{n}\frac{z_i}{\pi_i}$$

for which

$$Var\left(\hat{R}\right)=\frac{1}{\left[n_{t_1}^{(O)}+n_{t_2}^{(O)}\right]^2}Var\left(\sum_{i=1}^{n}\frac{z_i}{\pi_i}\right)$$

where covariance terms in the usual sample estimator can be dropped. The sample estimator for the variance of $\hat{C}_{rom}$ then becomes

$$\begin{aligned}\widehat{Var}\left(\hat{C}_{rom}\right) &= \left[n_{t_1}^{(O)}+n_{t_2}^{(O)}\right]^2\left\{\frac{\frac{\hat{N}}{n}\left(T\sum_{i=1}^{n}\frac{C_i}{L_i}\right)}{\left[n_{t_1}^{(O)}+n_{t_2}^{(O)}\right]^2}+\frac{1}{\left[n_{t_1}^{(O)}+n_{t_2}^{(O)}\right]^2}\sum_{i=1}^{n}\left(\frac{1-\pi_i}{\pi_i}\right)\frac{\hat{z}_i^2}{\pi_i}\right\}\\ &= \frac{\hat{N}}{n}\left(T\sum_{i=1}^{n}\frac{C_i}{L_i}\right)+\sum_{i=1}^{n}\left(\frac{1-\pi_i}{\pi_i}\right)\frac{\hat{z}_i^2}{\pi_i}\end{aligned}$$

where

$$\begin{aligned}\pi_i &= L_i^*/T \ , \\ \hat{N} &= \sum_{i=1}^{n} T/L_i^* \ \text{, and} \\ \hat{z}_i &= C_i \frac{L_i^*}{L_i} - \hat{R}\left[\delta_i\left(t_1\right) + \delta_i\left(t_2\right)\right] \text{ with } \hat{R} = \left(\sum_{i=1}^{n} \frac{C_i}{L_i}\right) \Bigg/ \left(\sum_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*} + \sum_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}\right) .\end{aligned}$$

In general, samples that are larger will perform better, however, n is random in the roving surveys and is not under the control of the investigator. Note that n does not explicitly appear in $Var\left(\hat{C}_{mor}\right)$ nor in $Var\left(\hat{C}_{rom}\right)$. Rather the collection of observations from a sample of any size are expanded to a full population set according to their probabilities of selection. As with access designs, increasing the number of overflight surveys does result in an increased use of auxiliary data and an accompanying decrease in variance and, as a variance decreasing device, is available to the investigator.

3.5.2.3 The Relationship Between $\hat{C}_{mor}$ and $\hat{C}_{rom}$

As with access surveys, $\hat{C}_{mor}$ and $\hat{C}_{rom}$ were shown to be unbiased and as per Table 3.3, $Var\left(\hat{C}_{rom}\right)$ is, in general, less than $Var\left(\hat{C}_{mor}\right)$ with equally weighted $\hat{C}_i$. To compare $Var\left(\hat{C}_{rom}\right)$ and $Var\left(\hat{C}_{rom}\right)$ again construct a weight conditional on the outcome of the sample. Given the results of the roving survey, choose

$$w = \frac{\sum_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*}}{\sum_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*} + \sum_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}} .$$

Now, conditional on any s,

$$Var\left(\frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*} + \sum_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}}\right) =$$

$$
\begin{aligned}
& Var\left(\frac{1}{2}\left\{\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}}\cdot w+\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\cdot(1-w)\right\}\right)\\
= & \left(\frac{w}{2}\right)^2 Var\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}}\right)\\
& +\left(\frac{1-w}{2}\right)^2 Var\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\right)\\
& +2\left(\frac{w}{2}\right)\left(\frac{1-w}{2}\right) Cov\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}},\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\right).
\end{aligned}
$$

It then follows that

$$
\begin{aligned}
Var\left(\hat{C}_{rom}\right) = & \left(\frac{1}{2}\right)^2\left\{\left[\frac{n_{t_1}^{(O)}+n_{t_2}^{(O)}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}+\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}\right]^2 Var\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}}\right)\right.\\
& +\left[\frac{n_{t_1}^{(O)}+n_{t_2}^{(O)}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}+\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}\right]^2 Var\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\right)\\
& +2\left[\frac{n_{t_1}^{(O)}+n_{t_2}^{(O)}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}+\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}\right]\times\\
& \left[\frac{n_{t_1}^{(O)}+n_{t_2}^{(O)}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}+\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}\right]\times\\
& \left. Cov\left(\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_1)}{L_i^*}},\frac{\sum_{i=1}^{n}\frac{C_i}{L_i}}{\sum_{i=1}^{n}\frac{\delta_i(t_2)}{L_i^*}}\right)\right\}.
\end{aligned}
$$

Note that even though w is a constant, the values used in its construction are realizations from a random sample for which L_i^*/T is the probability of selecting the i^{th} episode. Therefore, in a Horvitz-Thompson sense, $\sum_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*/T}$ is an approximation for $\sum_{j=1}^{N} \delta_j(t_1) = n_{t_1}^{(O)}$. Similarly $\sum_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*/T}$ for $n_{t_2}^{(O)}$. Thus, introducing T in both numerator and denominator, $Var\Big(\hat{C}_{rom}\Big)$ can then be approximated as

$$
\begin{aligned}
Var\Big(\hat{C}_{rom}\Big) \quad &\approx \quad \left(\frac{1}{2}\right)^2 \Bigg\{ \left[\frac{n_{t_1}^{(O)} + n_{t_2}^{(O)}}{\tilde{n}_{t_1}^{(O)} + \tilde{n}_{t_2}^{(O)}} \cdot \tilde{n}_{t_1}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*}}\right) \\
&\quad + \left[\frac{n_{t_1}^{(O)} + n_{t_2}^{(O)}}{\tilde{n}_{t_1}^{(O)} + \tilde{n}_{t_2}^{(O)}} \cdot \tilde{n}_{t_2}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}}\right) \\
&\quad + 2\left[\frac{n_{t_1}^{(O)} + n_{t_2}^{(O)}}{\tilde{n}_{t_1}^{(O)} + \tilde{n}_{t_2}^{(O)}} \cdot \tilde{n}_{t_1}^{(O)} \cdot \frac{n_{t_1}^{(O)} + n_{t_2}^{(O)}}{\tilde{n}_{t_1}^{(O)} + \tilde{n}_{t_2}^{(O)}} \cdot \tilde{n}_{t_2}^{(O)}\right] \times \\
&\quad Cov\left(\frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*}}, \frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}}\right)\Bigg\} \\
&\approx \quad \left(\frac{1}{2}\right)^2 \Bigg\{ \left[n_{t_1}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*}}\right) \\
&\quad + \left[n_{t_2}^{(O)}\right]^2 Var\left(\frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}}\right) \\
&\quad + 2\left[n_{t_1}^{(O)} \cdot n_{t_2}^{(O)}\right] Cov\left(\frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_1)}{L_i^*}}, \frac{\sum_{i=1}^{n} \frac{C_i}{L_i}}{\sum_{i=1}^{n} \frac{\delta_i(t_2)}{L_i^*}}\right)\Bigg\}.
\end{aligned}
$$

Thus, to the accuracy of the Horvitz-Thompson estimation, $Var\left(\hat{C}_{mor}\right)$ using w equals $Var\left(\hat{C}_{rom}\right)$ under equal weightings.

3.6 Simulation Results

Performance of the estimators was investigated using a simulated population consisting of 225 fishing episodes placed over a 24–hour fishing day (Table 3.1). As shown by Figure 3.1, these episodes ranged in length from 0.5 to 24.0 hours with the majority of lengths being between 1.0 and 5.0 hours. Catch histories were assigned to each episode using an exponential distribution with parameter $\lambda = 1.5$ fish/hour to generate the random times between

Table 3.1: Defining characteristics of the simulated population.

Start of the fishing day	00:00
End of the fishing day	24:00
Total number of fishing episodes (N)	225
Minimum length of any fishing episode in minutes	30
Initial catch rate as catch/hr	1.5
Total catch	2,007
Total effort in minutes (E)	81,150

successive catches. Figure 3.2 shows a plot of the resulting catch versus length of time fished for the completed episodes. Two scenarios were considered when placing the episodes over the fishing day: a uniform distribution and a more realistic distribution that uses a strong preference for fishing at 7:30 a.m. Within each placement method, scenarios allowing for 0 and 24 hours of overlap from the previous or into the following day were considered. When an episode was positioned such that an overlap would occur, then the overlapping portion

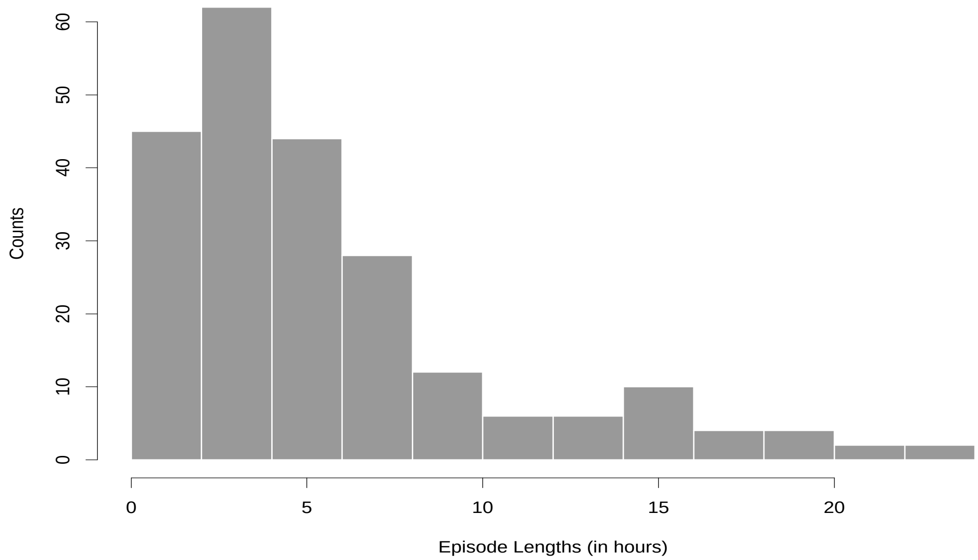

Figure 3.1: Fishing episodes by length for the simulated population (total count = 225).

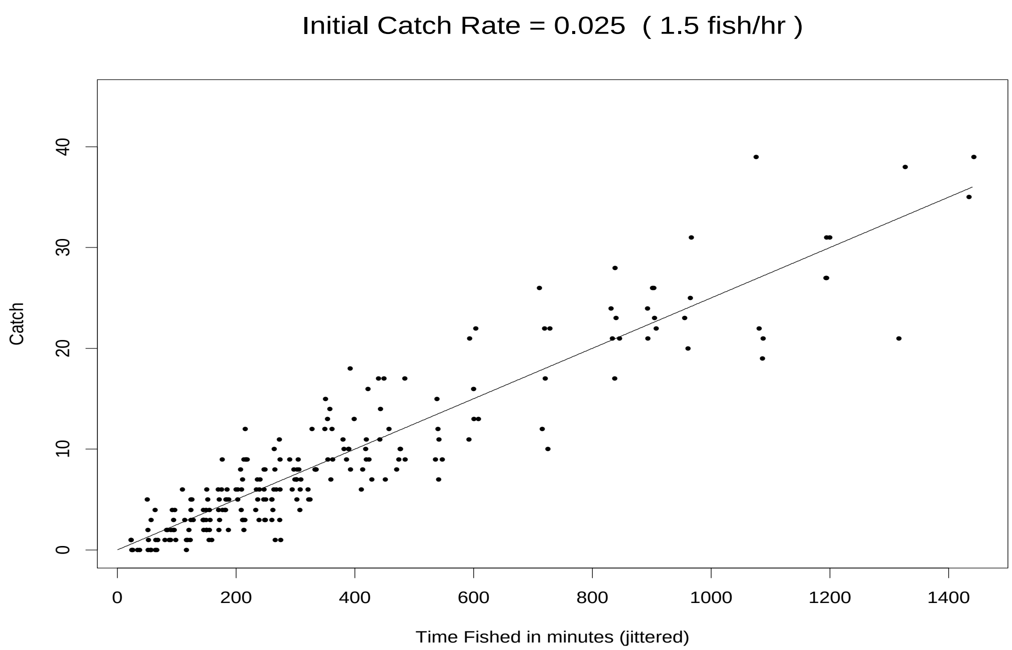

Figure 3.2: Catch vs. Time Fished for completed episodes of the simulated population. Also shown is a plot of the expected catch using the least squares fit.

was placed at the opposite end of the fishing day where it was then treated as an episode overlapping from (or into) the adjacent day. This was justified by the assumption of a repeating fishing pattern from day to day. When a "split" portion of an episode was selected for a roving sample, data using the full episode were recorded ensuring that all scenarios being compared had the the same total effort and total catch. Effort plots for the various episode placement scenarios are given in Figure 3.3. Since these plots are counts of active

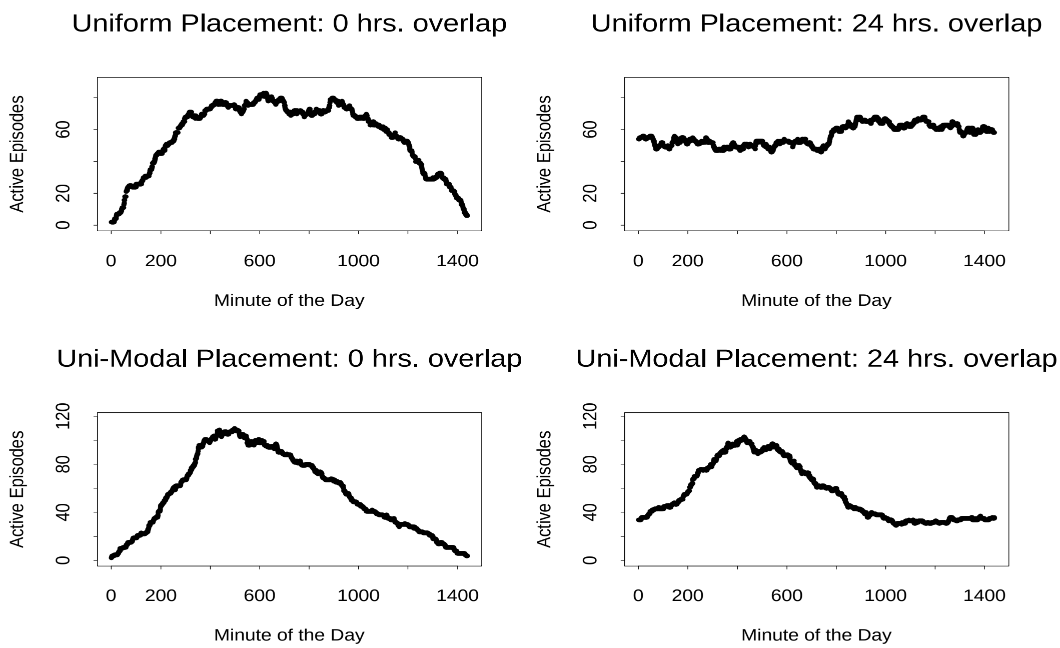

Figure 3.3: Count of active fishing episodes by minute of the fishing day. Overflights occur at minute 720 for the single overflight scenario or at minute 480 and minute 1080 for the multiple overflight scenario.

fishing episodes by time of day they are also values of $n_t^{(O)}$, the overflight survey results for a given time t, and impact on the values assumed by $n_t^{(A)}$ or $n_t^{(R)}$. Single overflight scenarios use $t = 12{:}00$ noon while for multiple overflight scenarios, $t_1 = 8{:}00$ a.m. and $t_2 = 6{:}00$ p.m.

Results of the access simulations are given in Table 3.2. For each scenario, random

Table 3.2: Performance comparison of the access survey catch estimators and their variance estimators using 500 replicated samples (simulated population).

		Overlap: 0 hours		Overlap: 24 hours	
Fishing Preference Pattern:		Uniform	Uni-Modal	Uniform	Uni-Modal
Actual Catch		2,007	2,007	2,007	2,007
Single Overflight at 12:00 noon					
Mean($\hat{C}_s$)	n=50	2,039	2,037	2,099	2,116
	n=125	2,008	2,017	2,036	2,014
Var($\hat{C}_s$)	n=50	103,793	70,743	241,241	195,906
	n=125	20,655	14,101	50,639	29,197
Mean $\left(\widehat{\text{Var}}_s(\hat{C})\right)$	n=50	102,716	72,962	289,515	193,306
	n=125	19,640	15,193	51,525	30,458
Multiple Overflights at 8:00 a.m. and 6:00 p.m - mor					
Mean($\hat{C}_s$)	n=50	2,054	2,066	2,077	2,030
	n=125	2,017	2,019	2,022	2,023
Var($\hat{C}_s$)	n=50	44,745	55,911	74,629	51,726
	n=125	8,041	11,273	13,568	14,511
Mean $\left(\widehat{\text{Var}}_s(\hat{C})\right)$	n=50	36,994	54,559	63,324	65,345
	n=125	7,767	11,445	12,390	14,893
Multiple Overflights at 8:00 a.m. and 6:00 p.m - rom					
Mean($\hat{C}_s$)	n=50	2,012	2,018	2,027	1,998
	n=125	2,010	2,008	2,012	2,012
Var($\hat{C}_s$)	n=50	36,489	36,746	60,417	40,427
	n=125	7,635	8,421	12,749	10,199
Mean $\left(\widehat{\text{Var}}_s(\hat{C})\right)$	n=50	34,366	36,621	59,456	46,067
	n=125	7,476	8,202	12,107	10,677
Optimal Weightings - mor					
Mean($\hat{C}_s$)	n=50	2,012	2,020	2,026	1,993
	n=125	2,010	2,009	2,012	2,011
Var($\hat{C}_s$)	n=50	36,754	36,714	60,253	39,068
	n=125	7,623	8,332	12,779	9,972
Mean $\left(\widehat{\text{Var}}_s(\hat{C})\right)$	n=50	33,806	35,790	58,747	44,479
	n=125	7,427	8,089	12,074	10,435

samples of size 50 and 125 were used. Bias, as a ratio to standard error, is less than 0.2 and therefore can be considered no worse than "mild" (Cochran, 1977). In general, the results for the 0–hour overlap scenarios are better than those for the 24–hour overlap scenarios. This is explained by Figure 3.3 which shows that, for the 0–hour overlap scenarios, there would be greater counts at times of the overflight surveys which would result in greater use of the auxiliary data and in turn, estimates that are less biased with lower variances. The lower variances in the multiple overflight scenarios, where a larger portion of the sample is included in the overflight counts, underlines the value of the ratio estimator. Also, better results are obtained from the ratio of means estimators demonstrating their greater stability, however, with optimized weightings, results from the mean of ratios estimators are comparable. In all cases, as would be expected, results improve with a larger sample. The larger than expected change in variance for a 2.5 times increase in sample size can be explained by the finite population correction factor.

To emulate the roving design, one minute was uniformly selected from the fishing day for each episode. This minute determined both sample selection for the episode and time of interview given selection. Note that, for each episode selected, it is assumed that L_i^* is known and for estimation purposes, is used in place of $2L_i$. In general, patterns in simulation results for all roving scenarios were similar to those of the corresponding access simulations. See Table 3.3. Note that since the total effort $E = 81,150$ episode-minutes and length of a fishing day $T = 1,440$ minutes, the average sample size for a roving survey is $E[n] = E/T \approx 55$ episodes making the results of the roving survey comparable, in a sense, to those of the access survey of size $n = 50$. Comparing results, the relatively more simple estimators of the access designs, with their fewer sources of variability, performed

Table 3.3: Performance comparison of the roving survey catch estimators and their variance estimators using 500 replicated samples (simulated population).

		Overlap: 0 hours		Overlap: 24 hours	
Fishing Preference Pattern:		Uniform	Uni-Modal	Uniform	Uni-Modal
Actual Catch		2,007	2,007	2,007	2,007
Single Overflight at 12:00 noon					
Mean($\hat{C}_s$)	E[n]=55	1,975	1,992	2,020	2,013
Var($\hat{C}_s$)	E[n]=55	110,068	107,159	251,750	150,076
Mean $\left(\widehat{\text{Var}}_s(\hat{C})\right)$	E[n]=55	93,382	89,032	160,834	115,478
Multiple Overflights at 8:00 a.m. and 6:00 p.m - mor					
Mean($\hat{C}_s$)	E[n]=55	2,004	1,997	2,014	2,031
Var($\hat{C}_s$)	E[n]=55	67,567	64,221	84,817	86,738
Mean $\left(\widehat{\text{Var}}_s(\hat{C})\right)$	E[n]=55	53,103	50,857	63,375	62,674
Multiple Overflights at 8:00 a.m. and 6:00 p.m - rom					
Mean($\hat{C}_s$)	E[n]=55	1,960	1,972	1,960	1,989
Var($\hat{C}_s$)	E[n]=55	58,837	64,555	78,266	73,850
Mean $\left(\widehat{\text{Var}}_s(\hat{C})\right)$	E[n]=55	56,681	56,664	69,388	68,308
Optimal Weightings - mor					
Mean($\hat{C}_s$)	E[n]=55	1,998	2,000	2,013	2,027
Var($\hat{C}_s$)	E[n]=55	62,902	63,956	85,577	79,026
Mean $\left(\widehat{\text{Var}}_s(\hat{C})\right)$	E[n]=55	51,001	51,088	62,595	61,438

better. It must be remembered, however, that with inclusion based on L_i^*, the roving designs will tend to sample episodes of longer duration making them more likely to be counted in the overflight surveys. This could result in greater use of the auxiliary data and hence, improved estimation. Note that attempts to increase the sample size by making a second pass through the resource can result in double sampling of some episodes making use of the proposed estimators inappropriate. It should also be noted that a 15–minute minimum time to interview was required of all sampled episodes. While this appeared to have little effect on the estimates of C, variance estimates became more stable but experienced increasingly negative bias with greater values of the minimum (results not shown). Also, use of L_i^* in place of $2L_i$ has a stabilizing effect which is more effectively used in the construction of $\widehat{Var}(\hat{C})$ than in $\hat{C}$ which contributes to the persistent underestimation of $Var(\hat{C})$.

3.7 The Fraser River Study

In 1995 the Canadian Department of Fisheries and Oceans (DFO) initiated an ongoing catch estimation program on the Fraser River, British Columbia, Canada (Palermo et al. 1997). This was a major undertaking encompassing a large area with changing fishing technologies and a growing number of fishing sites. One objective of the project is an estimate of sockeye catch from the gill net component, for which a complemented survey approach has been implemented. The effort estimates are made using helicopter overflights to count nets actively fishing, while the catch rate estimates are made from interview data on catch and time fished information, obtained from interviews using access and roving surveys at selected sites and on boat patrols over navigable stretches of the river.

To illustrate the use and compare the performance of the proposed estimators, 1996 data from a single Monday to Friday period over a continuous stretch of water was selected. Because activity counts at times of overflight were sparse, it was assumed that catch rates and fishing patterns were consistent from day to day and the data was folded into a single day. The overflight counts, which occurred on Wednesday and Thursday at 8:35 and 12:24, respectively, were then multiplied by a factor of five to act as two overflight counts on the constructed day. Despite improvements made over the initial year of the project, a number of problems had remained, many of which related to logistics and the difficult terrain. Night operations were limited and complete randomization was not always possible, particularly in the overflight surveys. Unattended nets and unwilling participation by some fishers remained a problem. No corrections were made for these and other such factors, instead, a single edit of the data file was used. Episodes longer than 3 days (*i.e.* a soak time of 72 hours in peak run time) were removed, as were episodes with catch in excess of 100 fish (*i.e.* a net load well in excess of 200 kg).

Estimates for catch over the 5–day period were made using the access survey approach and again using the roving survey approach using only the data collected from the respective design. Table 3.4 summarizes these results. Apart from their equally weighted mean of ratios estimates, both approaches estimate catch as being between 19,000 and 22,000 fish and both have little difference in variation between their respective ratio of means and optimized mean of ratios methods. The lower variances in the roving survey approach is explained by its greater use of the overflight auxiliary data, particularly on the 12:24 count. Compared with standard methods (*i.e.* based on an estimate of CPUE and total effort), the proposed methods are seen to be more consistent. The large discrepancy in size of the estimates for proposed versus standard methods is not surprising since restrictions in randomization of

Table 3.4: Catch estimates and their estimated variances for a Monday through Friday period over a continuous stretch of water on the Fraser River gill net fishery (1996).

			Access	Roving
Proposed Methods				
Overflight counts at		8:35	235	235
		12:24	175	175
Episodes sampled			49	60
Episodes active at		8:35	17	31
		12:24	5	35
mor:	$\hat{C}$		28,635	18,017
	$\widehat{Std}(\hat{C})$		5,369	1,592
	$\hat{C}_{Opt}$		19,976	21,842
	$\widehat{Std}(\hat{C}_{Opt})$		4,117	1,448
rom:	$\hat{C}$		21,860	19,432
	$\widehat{Std}(\hat{C})$		4,185	1,471
Standard Methods ($\hat{C} = \hat{E} \times \widehat{CPUE}$)				
$\hat{E}^1$ in fishing minutes			295,200	295,200
$\widehat{CPUE}$ as catch per minute			0.0525	0.0927
$\hat{C}$			77,424	136,806

[1] average of counts at 8:35 and 12:24 multiplied by 1440 minutes.

the overflights to prime times produce an overestimate of total effort (calculated as average count times 1440 minutes) which then inflates the catch estimates for the standard method. Note that actual DFO estimates, made using the total effort approach, did account for problems with implementation and would be smaller than those shown here. It should also be noted that, while a stationary Poisson process was used to develop estimators for the roving survey, a decreasing model for catch rate might have been more appropriate to describe gill net fishing.

3.8 Conclusions and Discussion

The above methods provide a defensible approach to estimating catch when the effort survey is not randomized. These methods appear to work well under a variety of scenarios for both access and roving designs. Fundamental to success is a properly designed and implemented ground survey and adequate resources must be allocated to this end. Remaining resources can be directed to increasing the number of overflight surveys. Overflight surveys are most effective when they yield high total counts and contain a high proportion of the episodes included in the ground survey. Consequently a single overflight should be scheduled at the peak fishing time. Choosing optimum times for additional overflights is less straightforward since now the lengths of the fishing episodes are also a consideration. In general, the more uniform the effort profile (see Figure 3.3), the more evenly spaced should be the scheduled times. When multiple overflights are employed, estimators should be based on either mean of ratios with optimized weightings, or ratio of means which is, in a sense, already optimally weighted.

In general, access designs are preferable to roving designs. Difficulties with design and implementation of the sampling scheme; random sample size with no means of adjusting to achieve a desired level of confidence; additional variability associated with estimating L_i^*; and the greater complexity of the estimators and difficulties relating to division by L_i are potentially serious problems inherent to roving designs. Protocol must include a minimum time to interview. Hoenig et al. (1997) used 30 minutes for measuring catch rate in a sports fishery. Here catch is being measured which may be more sensitive to a "cropped" sample, and 15 minutes may be more appropriate. To explore the usefulness of a known L_i^*, simulations were also made with $2L_i$ substituted for L_i^* (results not provided). In general, variances were two to three times larger.

In addition to overcoming the problem of division by L_i, it might also be possible to reduce variability in the roving survey variance estimators by using λL_i^* in place of $C_i L_i^*/L_i$ to estimate the catch of the i^{th} episode where λ is an overall catch rate. Variance estimators are then conditional on λ. Unconditioned formulae would be formed by substituting an estimate for λ and adding a component of extra variability due estimating λ. This is the subject of further research.

This chapter has shown that for a simple ground survey, estimation is possible even if the overflight survey is not random, and can be improved by increasing the number of overflights. In applications, the sampling design of the ground survey is apt to be complex. This presents a special problem because, for estimators to exist, each sampling session must include at least one episode that is in the overflight count. The next chapter examines estimation when the access survey is complex.

3.9 Acknowledgments

Data from the Fraser River gill net Fishery provided by the Department of Fisheries and Oceans, 3225 Stephenson Point Road, Nanaimo, B.C., V9T 1K3. Supported by Research Grants from the Natural Science and Engineering Research Council (NSERC) of Canada.

Chapter 4

Applications in Complex Surveys

Focusing on access designs, this chapter examines the problem when restrictions in randomization occur in the effort survey, and the sampling design of the access survey is complex. In particular, a catch estimator and its variance estimator are developed when the access survey has a three–stage sampling design with unequal probabilities of selection at each stage. Simulated data are then used to examine the performance of these estimators for differing amounts of variability at the various stages of sampling. Data from the Georgia Strait Creel Survey (Department of Fisheries and Oceans, Canada) are used to illustrate an application.

4.1 Introduction

Chapter 3 showed that, when a simple sampling scheme is used with the access survey, complemented angler surveys can still be used to estimate total catch in creel surveys, even if randomization in the effort survey is compromised. This chapter shows how, when the access survey is embedded in a complex sampling design, acceptable results can again be achieved, but with suitable modifications to the estimators.

In simple sampling designs, typically the sum of catch from an access sample is simply inflated to a population estimate, using the ratio of the observed auxiliary data in the sample to the population value of the auxiliary data observed by the overflight survey *i.e.* a simple ratio estimator. In complex sampling designs, however, it will be seen that the structure of the design must be built into the estimators. In effect, a population estimate of catch is made from the access survey alone, using the sample catch and the selection probabilities of the design. A similar estimate is made for the auxiliary information which is then used to "truth" the expansion procedure and adjust the catch estimate. Note that with unequal selection probabilities, the single expansion factor approach is not appropriate, especially when working with variances. It will also be seen that the population from which sampling units are selected may include units that catch no fish but carry the auxiliary information. Such units, e.g. non-fishers or fishers targeting a species not of interest, are merely assigned a zero catch. This feature can be exploited to improve the estimates and, in many cases, can lend practical convenience.

A complemented angler survey, with restricted randomization in the overflight component and a three–stage sampling design in the access component, is used to illustrate development of an estimator for catch and its variance. Performance of the estimators is examined using simulated data. For generality, unequal selection probabilities were assumed at each stage of sampling. Scenarios were constructed in which all the variability was placed in each level of the structure alone, as well as one in which the variability was placed across all levels. Simulations were then conducted, using combinations of high (75%) and low (25%) sampling rates, to study the effectiveness of different sampling strategies. The variance estimates were also compared with variance estimates computed using a jackknife.

The Georgia Strait Creel Survey is a continuing program, initiated in 1980 by the Canadian Department of Fisheries and Oceans to evaluate the Strait of Georgia sports fishery off the west coast of Canada. Originally the study was conceived as a complemented survey with a stratified three–stage access design and targeted only certain species of salmon. From the outset, the overflight component was not random. Over the course of the study, broadening of scope and practical considerations have since resulted in some randomization restrictions in the access component, and other deviations from the original design. The basic sampling design, however, remains multistage and provides an opportunity to apply the new estimators.

4.2 Notation

Subscripting for the stratified multistage design uses the following convention.

		Population Number	Sample Size	index	Inclusion Probability
Area or Stratum		NA	nA	a	1.0
Launch Site	given Area	NL_a	nL_a	l	τ_{al}
Block Time	given Area given Launch Site	NB_{al}	nB_{al}	b	π_{alb}
Interview	given Area given Launch Site given Block Time	NI_{alb}	nI_{alb}	i	ω_{albi}

For, the a^{th} area, l^{th} launch site and b^{th} time block,

C^*_{albi}	is the total catch for the i^{th} episode and
$m^{(O)}_{albi}$	is the number of times the i^{th} episode was counted by all overflight surveys. (Exists for all subscript values.)
$m^{(A)}_{albi}$	is the count of overflight surveys for which the i^{th} episode was active. (Exists for sampled subscript values.)

When the meaning is clear from the context, the a, l, b and i subscripting may be dropped.

$n^{(O)}_t$	Overflight survey count of active episodes at time t.
$n^{(A)}_t$	Access survey count of active episodes at time t.
$N^{(O)}$	Total of $n^{(O)}_t$ over all t.
$N^{(A)}$	Total of $n^{(A)}_t$ over all t.
$\delta_j()$	The indicator variable for the j^{th} episode to denote activity at time t.

4.3 Survey Design and Assumptions

Suppose that a sports fishery is stratified into NA geographic areas for which $\hat{C}$, an estimate of total catch is required, as well as an estimates of $\hat{C}_a$ for each stratum a. For simplicity, assume that the estimates are to be made for a single day, and that fishing is restricted to daylight hours which is divided into NB adjacent time blocks e.g. morning, afternoon, and evening. Also assume that the fishing activity in one area is independent of that in any other area. In addition, assume independence of the fishing activity of individuals within areas. In particular, fishing episodes do not cross area boundaries. Access by fishers to each area is through any of NL_a launch sites within each area.

There are two components for the survey: aerial counts of active fishing episodes and a ground access survey conducted at the launch sites. The access survey is a properly randomized multistage design. For each of the $a = 1, \ldots, NA$ *Areas* (or strata), $l = 1, \ldots, nL_a$ of NL_a *Launch* sites (*i.e.* primary units) are selected without replacement with probability τ_{al}. Within these, $b = 1, \ldots, nB_{al}$ of NB_{al} time *Blocks* are selected without replacement with probability π_{alb}. Finally, within each of these, $i = 1, \ldots, nI_{alb}$ of NI_{alb} episodes (*i.e.* angling parties) are selected without replacement with probability ω_{albi}. For generality, assume that all sampling is done with unequal probabilities of selection, but in many cases equal probability designs are used.

The data collected on each selected episode include C^*_{albi}, the number of fish caught, and the start and end times of the fishing episode. From the start and end times $\delta_{albi}(t_p)$, an indicator for fishing activity during times t_p of the $p = 1, \ldots, q$ aerial counts, can be determined. This "count" data from the access survey can be summarized as $m^{(A)}_{albi} = \sum_{p=1}^{q} \delta_{albi}(t_p)$ for each episode. Alternatively, over episodes, counts from each of the $p = 1, \ldots, q$ overflights at times t_p for a given *Area*, *Launch*, and *Block* combination are denoted as $n^{(A)}_{albt_p} = \sum_{i=1}^{nI_{alb}} \delta_{albi}(t_p)$. It is assumed that the start and end times used to determine δ are accurate.

The second component is an aerial survey. It is also assumed that "instantaneous" counts made by these surveys (Pollock et. al., 1994) are accurate. That is, $n^{(O)}_{t_p}$ are available and accurate at the area level from each of the overflight surveys. Note that the timings of the aerial counts are not random, possibly determined by aircraft availability or other such practical reasons. As such, they cannot be used to measure the fishing effort.

4.4 The Estimation of C

For a complemented access survey designed as a simple random sample of n fishing episodes each with catch C_i^* and for which q overflight counts have occurred at times $t_1, \ldots t_q$, it was shown in Chapter 3 that an unbiased estimator for total catch is

$$\begin{aligned}
\hat{C} &= \sum_{i=1}^{n} C_i^* \cdot \frac{n_{t_1}^{(O)} + \cdots + n_{t_q}^{(O)}}{n_{t_1}^{(A)} + \cdots + n_{t_q}^{(A)}} \\
&= \frac{\sum\limits_{i=1}^{n} C_i^*}{\sum\limits_{i=1}^{n} \sum\limits_{p=1}^{q} \delta_i(t_p)} \cdot \sum_{p=1}^{q} n_{t_p}^{(O)} \\
&= \hat{R} \cdot N^{(O)}
\end{aligned}$$

where $n_{t_p}^{(O)}$ and $n_{t_p}^{(A)}$ denote counts of episodes active at times t_p by the aerial and access surveys respectively. Note that $N^{(O)}$ is a known constant. It can also be shown that $\sum\limits_{i=1}^{n} C_i^* \Big/ \sum\limits_{p=1}^{q} \sum\limits_{i=1}^{n} \delta_i(t_p) = \sum\limits_{i=1}^{n} C_i^* \Big/ \sum\limits_{i=1}^{n} m_i^{(A)}$ estimates $C/N^{(O)}$. Thus, if $\hat{R}$ is formed as the ratio of unbiased estimators for C and $N^{(O)}$, then $\hat{C}$ can be generalized to any sampling design if unbiased estimators for C and $N^{(O)}$ can be developed for that design. With stratified designs, two methods for ratio estimation are commonly used (Cochran, 1977). "Separate" ratio estimates are formed by constructing separate ratio estimates in each stratum and then combining across strata in the appropriate manner. This method can be used only if the auxiliary data are available by stratum. "Combined" ratio estimates are formed by combining the elements of numerator and denominator across the strata before forming the ratio which is then used with the auxiliary data after they too have been combined over the strata. Here it is assumed that $N_a^{(O)}$ is known for each stratum a and, because estimates by stratum are generally required, only the method of separate estimators is considered. (The situation involving a single overflight count for all strata requires only minor additional modifications and is omitted.) From values for $N_a^{(O)}$ and independent estimates of R_a for each stratum, the estimate of total catch is formed using separate ratio estimates as,

$$\hat{C} = \sum_a \hat{R}_a N_a^{(O)}.$$

For the proposed multistage sampling plan, an unbiased estimator for catch for each of the $a = 1, \ldots, NA$ areas or strata is formed using a Horvitz-Thompson estimator at each stage.

$$\hat{C}_a = \sum_{l=1}^{nL_a} \frac{1}{\tau_{al}} \left\{ \sum_{b=1}^{nB_{al}} \frac{1}{\pi_{alb}} \left\{ \sum_{i=1}^{nI_{alb}} \frac{1}{\omega_{albi}} C^*_{albi} \right\} \right\}$$

and similarly for counts of activity at times of overflight

$$\begin{aligned} \hat{N}_a^{(O)} &= \sum_{l=1}^{nL_a} \frac{1}{\tau_{al}} \left\{ \sum_{b=1}^{nB_{al}} \frac{1}{\pi_{alb}} \left\{ \sum_{i=1}^{nI_{alb}} \frac{1}{\omega_{albi}} \left\{ \sum_{p=1}^{q} \delta_{albi}(t_p) \right\} \right\} \right\} \\ &= \sum_{l=1}^{nL_a} \frac{1}{\tau_{al}} \left\{ \sum_{b=1}^{nB_{al}} \frac{1}{\pi_{alb}} \left\{ \sum_{i=1}^{nI_{alb}} \frac{1}{\omega_{albi}} m_{albi}^{(A)} \right\} \right\}. \end{aligned}$$

An estimator for R for the a^{th} stratum is then

$$\hat{R}_a = \frac{\hat{C}_a}{\hat{N}_a^{(O)}} = \frac{\sum_{l=1}^{nL_a} \frac{1}{\tau_{al}} \left\{ \sum_{b=1}^{nB_{al}} \frac{1}{\pi_{alb}} \left\{ \sum_{i=1}^{nI_{alb}} \frac{1}{\omega_{albi}} C^*_{albi} \right\} \right\}}{\sum_{l=1}^{nL_a} \frac{1}{\tau_{al}} \left\{ \sum_{b=1}^{nB_{al}} \frac{1}{\pi_{alb}} \left\{ \sum_{i=1}^{nI_{alb}} \frac{1}{\omega_{albi}} m_{albi}^{(A)} \right\} \right\}}. \tag{4.1}$$

4.4.1 The Unbiasedness of $\hat{R}$

To show unbiasedness, note that for any stratum a, an unbiased estimate for catch is a nested sequence of Horvitz-Thompson estimators for which

$$
\begin{aligned}
E\left[\hat{C}_a\right] &= E\left[E\left[E\left[\hat{C}_a \middle| \begin{array}{c} Launch \\ Block \end{array}\right] \middle| Launch\right]\right] \\
&= E\left[E\left[\sum_{l=1}^{nL} \frac{1}{\tau_l}\left\{\sum_{b=1}^{nB_l} \frac{1}{\pi_{lb}}\left\{\sum_{i=1}^{NI_{lb}} C_{lbi}^*\right\}\right\} \middle| Launch\right]\right] \\
&= E\left[\sum_{l=1}^{nL} \frac{1}{\tau_l}\left\{\sum_{b=1}^{NB_l}\sum_{i=1}^{NI_{lb}} C_{lbi}^*\right\}\right] \\
&= \sum_{l=1}^{NL}\sum_{b=1}^{NB_l}\sum_{i=1}^{NI_{lb}} C_{lbi}^* \\
&= C_a.
\end{aligned}
$$

Similarly

$$
\begin{aligned}
E\left[\hat{N}_a^{(O)}\right] &= \sum_{l=1}^{NL}\sum_{b=1}^{NB_l}\sum_{i=1}^{NI_{lb}} m_{lbi}^{(A)} \\
&= N_a^{(O)}.
\end{aligned}
$$

Therefore, to a first order Taylor approximation,

$$
E\left[\hat{R}_a\right] = E\left[\hat{C}_a\right] \Big/ E\left[\hat{N}_a^{(O)}\right] = C_a / N_a^{(O)} = R_a
$$

and the estimators for R at the stratum level are approximately unbiased as is $\hat{C}$ since $E\left[\hat{C}\right] = \sum_a E\left[\hat{R}_a\right] N_a^{(O)}$. Note that strata subscripts will be suppressed and used only where needed to avoid confusion.

4.4.2 The Variance of $\hat{R}$

The variance of $\hat{R}$ can be determined by standard methods used in multistage sampling by decomposing it into a sum of components relating to each stage. An estimator for $Var\left(\hat{R}\right)$ can then be made by applying estimation techniques to each component. With the proposed multistage sampling plan, for any stratum

$$\begin{aligned} Var\left(\hat{R}\right) &= Var\begin{pmatrix}\text{AMONG}\\ Launches\end{pmatrix} + Var\begin{pmatrix}\text{WITHIN}\\ Launches\end{pmatrix} \\ &= Var\begin{pmatrix}\text{AMONG}\\ Launches\end{pmatrix} + Var\begin{pmatrix}\text{AMONG}\\ Blocks\end{pmatrix} + Var\begin{pmatrix}\text{WITHIN}\\ Blocks\end{pmatrix} \end{aligned} \tag{4.2}$$

where, as the final stage of decomposition, $Var\begin{pmatrix}\text{WITHIN}\\ Blocks\end{pmatrix} = Var\begin{pmatrix}\text{AMONG}\\ Episodes\end{pmatrix}$. Next, each components is expressed as nested conditional expectations and variances as follows:

- $Var\begin{pmatrix}\text{AMONG}\\ Launches\end{pmatrix} = Var\left(E\left[\ \hat{R}\ \middle|\ Launch\right]\right)$ for which, using a ratio of expectations to approximate the expectation of the ratio,

$$\begin{aligned} E\left[\ \hat{R}\ \middle|\ Launch\right] &= E\left[E\left[\ \hat{R}\ \middle|\ \begin{matrix}Launch\\ Block\end{matrix}\right]\ \middle|\ Launch\right] \\ &= \frac{\sum_{l=1}^{nL}\frac{1}{\tau_l}\sum_{b=1}^{NB_l}\sum_{i=1}^{NI_{lb}} C^*_{lbi}}{\sum_{l=1}^{nL}\frac{1}{\tau_l}\sum_{b=1}^{NB_l}\sum_{i=1}^{NI_{lb}} m^{(A)}_{lbi}} \end{aligned}$$

  is a ratio of averaged stratum estimates of total catch and overflight counts made for each level of Launch. A Taylor expansion about ($Y = \sum_{l=1}^{NL}\sum_{b=1}^{NB_l}\sum_{i=1}^{NI_{lb}} C^*_{lbi}, X = \sum_{l=1}^{NL}\sum_{b=1}^{NB_l}\sum_{i=1}^{NI_{lb}} m^{(O)}_{lbi}$) gives

$$\begin{aligned} Var\left(E\left[\ \hat{R}\ \middle|\ Launch\right]\right) = Var\Bigg(&\frac{Y}{X} - \frac{Y}{X^2}\left\{\sum_{l=1}^{nL}\frac{1}{\tau_l}\sum_{b=1}^{NB_l}\sum_{i=1}^{NI_{lb}} m^{(A)}_{lbi} - X\right\} \\ &+\frac{1}{X}\left\{\sum_{l=1}^{nL}\frac{1}{\tau_l}\sum_{b=1}^{NB_l}\sum_{i=1}^{NI_{lb}} C^*_{lbi} - Y\right\}\Bigg) \end{aligned}$$

$$= Var\left(R + \frac{1}{X}\left\{ \sum_{l=1}^{nL} \frac{1}{\tau_l} \sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} C^*_{lbi} - R \sum_{l=1}^{nL} \frac{1}{\tau_l} \sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} m^{(A)}_{lbi} \right\} \right)$$

$$= \frac{1}{[X]^2} Var\left(\sum_{l=1}^{nL} \frac{1}{\tau_l} \left\{ \sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} C^*_{lbi} - R \sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} m^{(A)}_{lbi} \right\} \right).$$

Note that $\sum_{l=1}^{nL} \frac{1}{\tau_l} \left\{ \sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} C^*_{lbi} - R \sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} m^{(A)}_{lbi} \right\}$ is a Horvitz-Thompson estimator. A biased (tending to overestimate) approximation for the sample estimator for its variance can be formed by converting the estimator to a sample average of totals and then applying the usual estimate for an average (Brewer and Hanif, 1983). Extending this sample technique back to the full set of population values,

$$Var\left(E\left[\hat{R} \,\middle|\, Launch \right] \right) \approx$$

$$\frac{1}{\left[\sum_{l=1}^{NL} \sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} m^{(O)}_{lbi} \right]^2} \left(\frac{NL - nL}{NL} \right) \left(\frac{1}{nL} \right) \frac{1}{NL-1} \sum_{l=1}^{NL} \left(Z_l - \bar{Z}_l \right)^2 \tag{4.3}$$

where $\qquad Z_l = \frac{nL}{\tau_l} \left(\sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} C^*_{lbi} - R \sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} m^{(A)}_{lbi} \right)$

for which the sample estimator is then

$$\widehat{Var}\begin{pmatrix} \text{AMONG} \\ Launches \end{pmatrix} =$$

$$\frac{1}{\left[\hat{N}^{(O)} \right]^2} \left(\frac{NL - nL}{NL} \right) \left(\frac{1}{nL} \right) \frac{1}{nL-1} \sum_{l=1}^{nL} \left(z_l - \bar{z}_l \right)^2 \tag{4.4}$$

where

$$\hat{N}^{(O)} = \sum_{l=1}^{nL} \frac{1}{\tau_l} \sum_{b=1}^{nB_l} \frac{1}{\pi_{lb}} \sum_{i=1}^{nI_{lb}} \frac{1}{\omega_{lbi}} m_{lbi}^{(A)}$$

and

$$z_l = \frac{nL}{\tau_l} \left(\sum_{b=1}^{nB_l} \frac{1}{\pi_{lb}} \sum_{i=1}^{nI_{lb}} \frac{1}{\omega_{lbi}} C^*_{lbi} - \hat{R} \sum_{b=1}^{nB_l} \frac{1}{\pi_{lb}} \sum_{i=1}^{nI_{lb}} \frac{1}{\omega_{lbi}} m_{lbi}^{(A)} \right).$$

- $Var\begin{pmatrix}\text{AMONG}\\ Blocks\end{pmatrix} = E\left[Var\left(E\left[\hat{R} \,\middle|\, \begin{matrix} Launch \\ Block \end{matrix}\right] \,\middle|\, Launch\right)\right]$ for which, again approximating an expectation of a ratio with the ratio of its expectations,

$$E\left[\hat{R} \,\middle|\, \begin{matrix} Launch \\ Block \end{matrix}\right] = \frac{\sum_{l=1}^{nL} \frac{1}{\tau_l} \left\{ \sum_{b=1}^{nB_l} \frac{1}{\pi_{lb}} \left\{ \sum_{i=1}^{NI_{lb}} C^*_{lbi} \right\} \right\}}{\sum_{l=1}^{nL} \frac{1}{\tau_l} \left\{ \sum_{b=1}^{nB_l} \frac{1}{\pi_{lb}} \left\{ \sum_{i=1}^{NI_{lb}} m_{lbi}^{(A)} \right\} \right\}}$$

which is a ratio of nested sequences of weighted averaged estimates of total catch and overflight counts, culminating in stratum estimates. A Taylor expansion about $(Y = \sum_{l=1}^{NL} \sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} C^*_{lbi}, X = \sum_{l=1}^{NL} \sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} m_{lbi}^{(O)})$ then gives

$$\begin{aligned} E\left[\hat{R} \,\middle|\, \begin{matrix} Launch \\ Block \end{matrix}\right] &= R + \frac{1}{X} \left[\sum_{l=1}^{nL} \frac{1}{\tau_l} \left\{ \sum_{b=1}^{nB_l} \frac{1}{\pi_{lb}} \left\{ \sum_{i=1}^{NI_{lb}} C^*_{lbi} \right\} \right\} \right. \\ &\quad \left. - R \sum_{l=1}^{nL} \frac{1}{\tau_l} \left\{ \sum_{b=1}^{nB_l} \frac{1}{\pi_{lb}} \left\{ \sum_{i=1}^{NI_{lb}} m_{lbi}^{(A)} \right\} \right\} \right] \\ &= R + \frac{1}{X} \left[\sum_{l=1}^{nL} \frac{1}{\tau_l} \left\{ \sum_{b=1}^{nB_l} \frac{1}{\pi_{lb}} \left(\sum_{i=1}^{NI_{lb}} C^*_{lbi} - R \sum_{i=1}^{NI_{lb}} m_{lbi}^{(A)} \right) \right\} \right] \end{aligned}$$

for which, again extending the sample estimator of Brewer and Hanif, 1983 to the full set of population values

$$Var\left(E\left[\hat{R} \,\middle|\, \begin{array}{c} Launch \\ Block \end{array} \right] \,\middle|\, Launch \right) \approx \frac{1}{[X]^2}$$

$$\times \left\{ \frac{1}{nL} \sum_{l=1}^{nL} \frac{nL}{(\tau_l)^2} \left(\frac{NB_l - nB_l}{NB_l} \right) \left(\frac{1}{nB_l} \right) \frac{1}{NB_l - 1} \sum_{b=1}^{NB_l} \left(Z_{lb} - \bar{Z_{lb}} \right)^2 \right\}$$

where $\quad Z_{lb} = \frac{nB_l}{\pi_{lb}} \left(\sum_{i=1}^{NI_{lb}} C^*_{lbi} - R \sum_{i=1}^{NI_{lb}} m^{(A)}_{lbi} \right).$

Finally, to the accuracy of the approximations,

$$E\left[Var\left(E\left[\hat{R} \,\middle|\, \begin{array}{c} Launch \\ Block \end{array} \right] \,\middle|\, Launch \right) \right] = \frac{1}{\left[\sum_{l=1}^{NL} \sum_{b=1}^{NB_l} \sum_{i=1}^{NI_{lb}} m^{(O)}_{lbi} \right]^2}$$

$$\times \frac{1}{NL} \sum_{l=1}^{NL} \frac{nL}{(\tau_l)^2} \left(\frac{NB_l - nB_l}{NB_l} \right) \left(\frac{1}{nB_l} \right) \frac{1}{NB_l - 1} \sum_{b=1}^{NB_l} \left(Z_{lb} - \bar{Z_{lb}} \right)^2 \qquad (4.5)$$

for which the sample estimator is

$$\widehat{Var}\left(\begin{array}{c} \text{AMONG} \\ Blocks \end{array} \right) = \frac{1}{\left[\hat{N}^{(O)} \right]^2}$$

$$\times \left\{ \sum_{l=1}^{nL} \frac{1}{(\tau_l)^2} \left(\frac{NB_l - nB_l}{NB_l} \right) \left(\frac{1}{nB_l} \right) \frac{1}{nB_l - 1} \sum_{b=1}^{nB_l} (z_{lb} - \bar{z_{lb}})^2 \right\} \qquad (4.6)$$

where

$$\hat{N}^{(O)} = \sum_{l=1}^{nL} \frac{1}{\tau_l} \sum_{b=1}^{nB_l} \frac{1}{\pi_{lb}} \sum_{i=1}^{nI_{lb}} \frac{1}{\omega_{lbi}} m^{(A)}_{lbi}$$

and

$$z_{lb} = \frac{nB_l}{\pi_{lb}} \left(\sum_{i=1}^{nI_{lb}} \frac{1}{\omega_{lbi}} C^*_{lbi} - \hat{R} \sum_{i=1}^{nI_{lb}} \frac{1}{\omega_{lbi}} m^{(A)}_{lbi} \right).$$

- $Var\left(\begin{matrix}\text{WITHIN}\\ Blocks\end{matrix}\right) = E\left[E\left[Var\left(\hat{R}\;\middle|\;\begin{matrix}Launch\\ Block\end{matrix}\right)\;\middle|\;Launch\right]\right]$

for which, again using a Taylor expansion about the values $Y = \sum_{l=1}^{NL}\sum_{b=1}^{NB_l}\sum_{i=1}^{NI_{lb}} C^*_{lbi}$ and $X = \sum_{l=1}^{NL}\sum_{b=1}^{NB_l}\sum_{i=1}^{NI_{lb}} m^{(O)}_{lbi}$,

$$
\begin{aligned}
Var\left(\hat{R}\;\middle|\;\begin{matrix}Launch\\ Block\end{matrix}\right) &= Var\left(\frac{\sum_{l=1}^{nL}\frac{1}{\tau_l}\left\{\sum_{b=1}^{nB_l}\frac{1}{\pi_{lb}}\left\{\sum_{i=1}^{nI_{lb}}\frac{1}{\omega_{lbi}}C^*_{lbi}\right\}\right\}}{\sum_{l=1}^{nL}\frac{1}{\tau_l}\left\{\sum_{b=1}^{nB_l}\frac{1}{\pi_{lb}}\left\{\sum_{i=1}^{nI_{lb}}\frac{1}{\omega_{lbi}}m^{(A)}_{lbi}\right\}\right\}}\right)\\
&= Var\Bigg(R + \frac{1}{X}\Bigg[\sum_{l=1}^{nL}\frac{1}{\tau_l}\left\{\sum_{b=1}^{nB_l}\frac{1}{\pi_{lb}}\left\{\sum_{i=1}^{nI_{lb}}\frac{1}{\omega_{lbi}}C^*_{lbi}\right\}\right\}\\
&\qquad -R\sum_{l=1}^{nL}\frac{1}{\tau_l}\left\{\sum_{b=1}^{nB_l}\frac{1}{\pi_{lb}}\left\{\sum_{i=1}^{nI_{lb}}\frac{1}{\omega_{lbi}}m^{(A)}_{lbi}\right\}\right\}\Bigg]\Bigg)\\
&= Var\left(R + \frac{1}{X}\left[\sum_{l=1}^{nL}\frac{1}{\tau_l}\left\{\sum_{b=1}^{nB_l}\frac{1}{\pi_{lb}}\left\{\sum_{i=1}^{nI_{lb}}\frac{1}{\omega_{lbi}}\left(C^*_{lbi} - Rm^{(A)}_{lbi}\right)\right\}\right\}\right]\right).
\end{aligned}
$$

As done previously when expressed in this form, the variance of the Horvitz-Thompson estimator is next approximated with the usual sample estimator (Brewer and Hanif, 1983) and extended to the full set of population values to give

$$
\begin{aligned}
Var\left(\hat{R}\;\middle|\;\begin{matrix}Launch\\ Block\end{matrix}\right) &\approx \frac{1}{[X]^2}\\
&\quad\times\left\{\sum_{l=1}^{nL}\frac{1}{(\tau_l)^2}\left\{\sum_{b=1}^{nB_l}\frac{1}{(\pi_{lb})^2}\left(\frac{NI_{lb} - nI_{lb}}{NI_{lb}}\right)\left(\frac{1}{nI_{lb}}\right)S^2_Z\right\}\right\}
\end{aligned}
$$

where

$$S^2_Z = \frac{1}{NI_{lb}-1}\sum_{i=1}^{NI_{lb}}\left(Z_{lbi} - \bar{Z_{lbi}}\right)^2$$

with

$$Z_{lbi} = \frac{nI_{lb}}{\omega_{lbi}}\left(C^*_{lbi} - R\,m^{(A)}_{lbi}\right).$$

Expressed this way it can be seen that

$$E\left[Var\left(\hat{R}\;\middle|\;\begin{matrix}Launch\\Block\end{matrix}\right)\;\middle|\;Launch\right]=\frac{1}{[X]^2}$$
$$\times\left\{\sum_{l=1}^{nL}\frac{1}{(\tau_l)^2}\left\{\frac{1}{NB_l}\sum_{b=1}^{NB_l}\frac{nB_l}{(\pi_{lb})^2}\left(\frac{NI_{lb}-nI_{lb}}{NI_{lb}}\right)\left(\frac{1}{nI_{lb}}\right)S_Z^2\right\}\right\}$$

and that

$$E\left[E\left[Var\left(\hat{R}\;\middle|\;\begin{matrix}Launch\\Block\end{matrix}\right)\;\middle|\;Launch\right]\right]=\frac{1}{[X]^2}$$
$$\times\left\{\frac{1}{NL}\sum_{l=1}^{NL}\frac{nL}{(\tau_l)^2}\left\{\frac{1}{NB_l}\sum_{b=1}^{NB_l}\frac{nB_l}{(\pi_{lb})^2}\left(\frac{NI_{lb}-nI_{lb}}{NI_{lb}}\right)\left(\frac{1}{nI_{lb}}\right)S_Z^2\right\}\right\} \qquad (4.7)$$

for which a sample estimator is

$$\widehat{Var}\begin{pmatrix}\text{WITHIN}\\Blocks\end{pmatrix}=\frac{1}{\left[\hat{N}^{(O)}\right]^2}$$
$$\times\left\{\sum_{l=1}^{nL}\frac{1}{(\tau_l)^2}\left\{\sum_{b=1}^{nB_l}\frac{1}{(\pi_{lb})^2}\left(\frac{NI_{lb}-nI_{lb}}{NI_{lb}}\right)\left(\frac{1}{nI_{lb}}\right)s_z^2\right\}\right\} \qquad (4.8)$$

where

$$\hat{N}^{(O)} = \sum_{l=1}^{nL}\frac{1}{\tau_l}\sum_{b=1}^{nB_l}\frac{1}{\pi_{lb}}\sum_{i=1}^{nI_{lb}}\frac{1}{\omega_{lbi}}m_{lbi}^{(A)}$$

and

$$s_z^2 = \frac{1}{nI_{lb}-1}\sum_{i=1}^{nI_{lb}}\left(z_{lbi}-\bar{z_{lbi}}\right)^2$$

with

$$z_{lbi} = \frac{nI_{lb}}{\omega_{lbi}}\left(C_{lbi}^{*}-\hat{R}\,m_{lbi}^{(A)}\right)\cdot$$

4.4.3 Jackknife Estimates

Finding the required formula for variance estimators can become exceedingly tedious as the sampling design becomes more complex. Jackknife estimates offer an alternative that work with the relatively more simple formulation of $\hat{R}$ which can be expressed as

$$\hat{R}_a = \frac{\sum_{l=1}^{nL} \sum_{b=1}^{nB} \sum_{i=1}^{nI} C^*_{albi} W_{albi}}{\sum_{l=1}^{nL} \sum_{b=1}^{nB} \sum_{i=1}^{nI} m^*_{albi} W_{albi}}$$

where weights W_{albi} are formed using the sample inclusion probabilities. If sampling at all stages is done using equal probabilities, then $\hat{R}$ is further simplified with cancelation of the W_{albi}.

To form the jackknife estimates for $\hat{R}$ in a multistage design, estimates using the reduced-by-one sample are made only at the primary level of sampling. That is, for the three–stage design being considered, for any stratum estimates $\hat{R}^*_{-l}$ for $l = 1, \ldots, nL$ are constructed using Equation 4.1 and the sampled data each with *Launch* l and its levels of *Blocks* and episodes removed. The jackknife estimate for each stratum is then formed as

$$\widehat{Var}_J\left(\hat{R}\right) = \frac{nL-1}{nL} \sum_{l=1}^{nL} \left(\hat{R}^*_{-l} - \bar{\hat{R}}^*\right)^2. \tag{4.9}$$

In general, jackknife estimates work well for continuous functions of means and, therefore, it can be expected that jackknifing will work well for $\hat{R}$. Another "rule of thumb" is that good results can be expected if a Taylor approximation also works well. Because the formula variance estimator uses a Taylor approximation, similar results might be expected for both methods. Note that small values of $m^{(A)}_{albi}$ can cause problems in both $\hat{R}^*_{-l}$ and the formula variance estimator. Referring to simple sampling designs and citing work by J.N.K. Rao, Sukhatme et. al. (1984) notes that a general conclusion is that the jackknife estimator tends to overestimate, while the usual formula estimator for a ratio estimator tends to underestimate. It was also noted that further work was necessary before any final conclusions could be made.

4.4.4 Design Variations

The above approach to finding an estimator for R and its variance remains the same whether the sampling plan changes by varying the number of stages or the method of selecting sampling units at each stage. Three methods of unit selection at the final stage of sampling that are likely to arise in practice, however, are worthy of comment.

Select Units With Equal Probability. Inclusion probabilities ω_{lbi} now become nI_{lb}/NI_{lb} for all episodes within a given *Launch* and *Block* combination. As a result, all $1/\omega_{lbi}$ factor outside their summations as NI_{lb}/nI_{lb}.

Select All Units. All nI_{lb} terms now become NI_{lb} so that quantities such as $\sum_{i=1}^{nI_{lb}} \frac{1}{\omega_{lbi}} C^*_{lbi}$ are replaced with $\sum_{i=1}^{NI_{lb}} C^*_{lbi}$. The finite population correction factors are also dropped from the variance formulae at the episode or interview level.

Bernoulli Sampling. At some interview sites, the anticipated number of fishers might exceed the manpower available to accommodate them. Bernoulli sampling, in which sample inclusion is determined according to a pre-chosen probability as episodes are encountered, offers a means of reducing the number of interviews to be taken, while preserving unbiasedness since interviews are randomly selected from across the full *Block* time. With this type of sampling, the inclusion probabilities ω_{lbi} now equal ω_{lb} for all episodes, while the sample size nI_{lb} becomes random. With a target sample size of n and using $\omega_{lb} = n/NI_{lb}$, the unbiased ratio estimator is the same as that for selection with equal probability (Särndal, 1992). The variance, however, must include an extra factor to account for the randomness of nI_{lb} so that given *Launch* and *Block*, the form in the estimator for the variance of $\hat{R}$ is now

$$\left(\frac{NI_{lb}-n}{NI_{lb}}\right)\left(\frac{1}{nI_{lb}}\right)\left(\frac{nI_{lb}-1}{nI_{lb}}\right)\frac{1}{nI_{lb}-1}\sum_{i=1}^{nI_{lb}}\left(z_{lbi}-\bar{z_{lbi}}\right)^2$$

where

$$z_{lbi} \quad = \quad \tfrac{nI_{lb}\cdot NI_{lb}}{n}\left(C^{*}_{lbi}-\hat{R}\,m^{(A)}_{lbi}\right).$$

4.5 The Variance of Catch

It has been assumed that aerial counts are made accurately. Thus, with $\hat{C}=\sum_a \hat{R}_a N_a^{(O)}$ and independence amongst strata,

$$Var\left(\hat{C}\right)=\sum_a\left(N_a^{(O)}\right)^2 Var\left(\hat{R}_a\right) \tag{4.10}$$

where, for any stratum a, $Var\left(\hat{R}_a\right)$ is found as the sum of Equations 4.3, 4.5 and 4.7. An estimate is found using the sum of Equations 4.4, 4.6 and 4.8.

If there is variability in $N^{(O)}$, then, provided that there is an unbiased estimate of $N^{(O)}$ and an estimate of its variance, the variance of the catch estimator is found as

$$\widehat{Var}\left(\hat{C}\right)=\widehat{Var}\left(\hat{N}^{(O)}\right)\left(\hat{R}\right)^2+\widehat{Var}\left(\hat{R}\right)\left(\hat{N}^{(O)}\right)^2-\widehat{Var}\left(\hat{N}^{(O)}\right)\widehat{Var}\left(\hat{R}\right)$$

(Goodman, 1960).

4.6 Simulation Results

In general, an estimator performs best when the sampling effort is concentrated where the variability is greatest. To assess the performance of $\hat{R}$ and its variance estimator, simulations were made using four variability scenarios, each with a different catch rate for *Launch* site, interview *Block* and *Episode*. Combinations of high ($H = 75\%$) and low ($L = 25\%$) sampling rates applied to each scenario could then be used to compare effectiveness of various sampling strategies. For this purpose, a single fishing day lasting from 5:00 until 21:00 was used. Overflight counts were scheduled at 8:00, 13:00, and 16:00.

The test population was constructed by first generating a total of 2,670 episodes, ranging in length from 1.5 to 10 hours for 20 launch sites. The distribution of these episode lengths is given in Figure 4.1. These were then randomly placed over the fishing day according to a

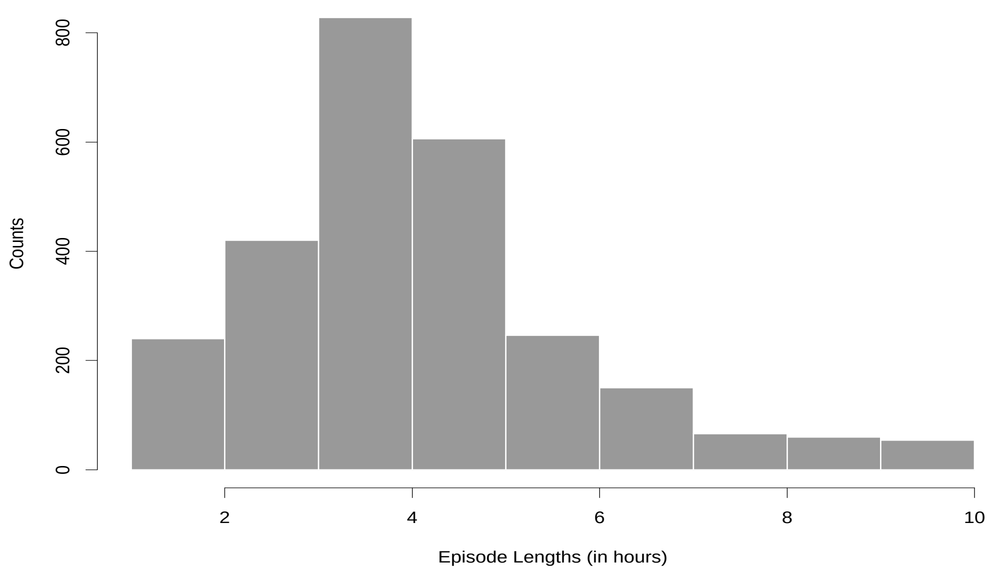

Figure 4.1: Simulated Data: Distribution of episode lengths for each scenario (total count=2,670).

fishing preference curve with a major peak at 7:00 and a lesser peak at 19:30 (Figure 4.2). This resulted in the effort profile shown in Figure 4.3 and determined into which interview

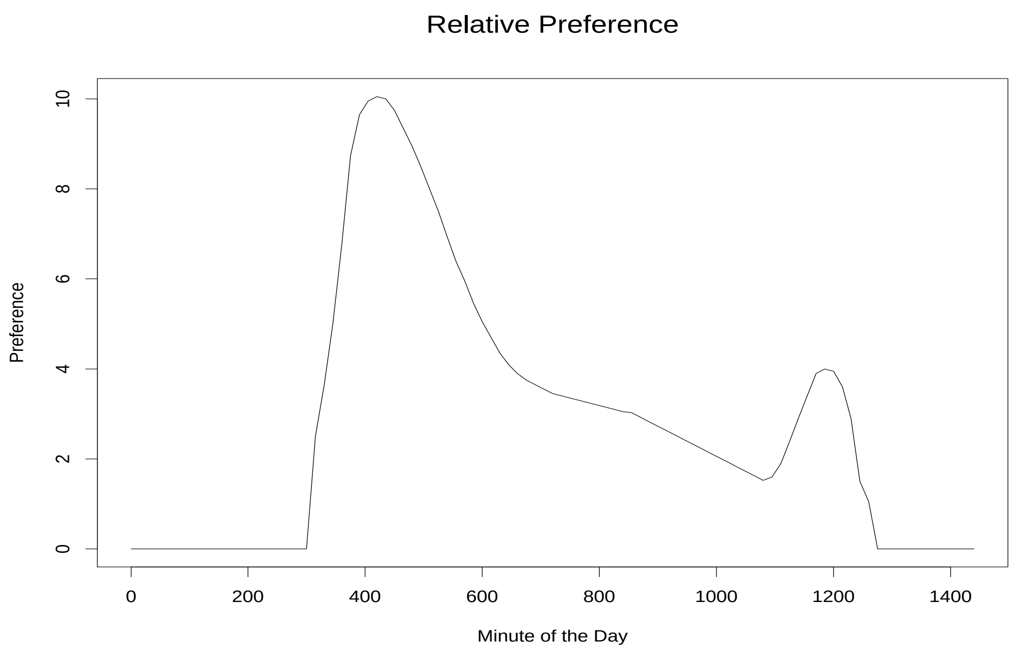

Figure 4.2: Simulated Data: Fishing Time Preference Curve.

block an episode would fall (*i.e.* episode end time) and, together with its start time, whether or not, if sampled, it would be included in a particular overflight count. Block times were defined as Block I from 5:00 to 8:59; Block II from 9:00 to 12:59; Block III from 13:00 to 16:59; and Block IV from 17:00 to 21:00. Table 4.1 summarizes the generation and placement of episodes. This "shell" was then replicated to form the basis for the four scenarios for a total of 10,680 episodes.

Catch was assumed to be a homogeneous Poisson process for each episode but with different catch rates, depending on the scenario. For Scenario I all the variability was put in the episodes. Each episode had a catch rate randomly chosen between 1.5 and 2.5 fish

Table 4.1: Simulated data: Episode counts for each scenario.

	Total	By Length (hrs)			By Block			
Launch	Episodes	0 to 3	3 to 6	6 to 16	1	2	3	4
1	180	24	126	30	13	64	54	49
2	165	21	111	33	8	55	42	60
3	150	27	99	24	7	61	46	36
4	135	27	87	21	9	55	38	33
5	135	27	87	21	8	60	31	36
6	120	21	75	24	10	58	26	26
7	120	21	75	24	8	54	32	26
8	120	21	75	24	7	43	40	30
9	105	21	63	21	15	45	24	21
10	105	21	63	21	9	42	27	27
11	180	24	126	30	12	78	53	37
12	165	21	111	33	10	70	54	31
13	150	27	99	24	7	54	43	46
14	135	27	87	21	12	60	33	30
15	135	27	87	21	13	58	33	31
16	120	21	75	24	9	47	38	26
17	120	21	75	24	11	52	31	26
18	120	21	75	24	9	43	24	44
19	105	21	63	21	10	37	35	23
20	105	21	63	21	6	47	25	27
	2,670	462	1,722	486	193	1,083	729	665

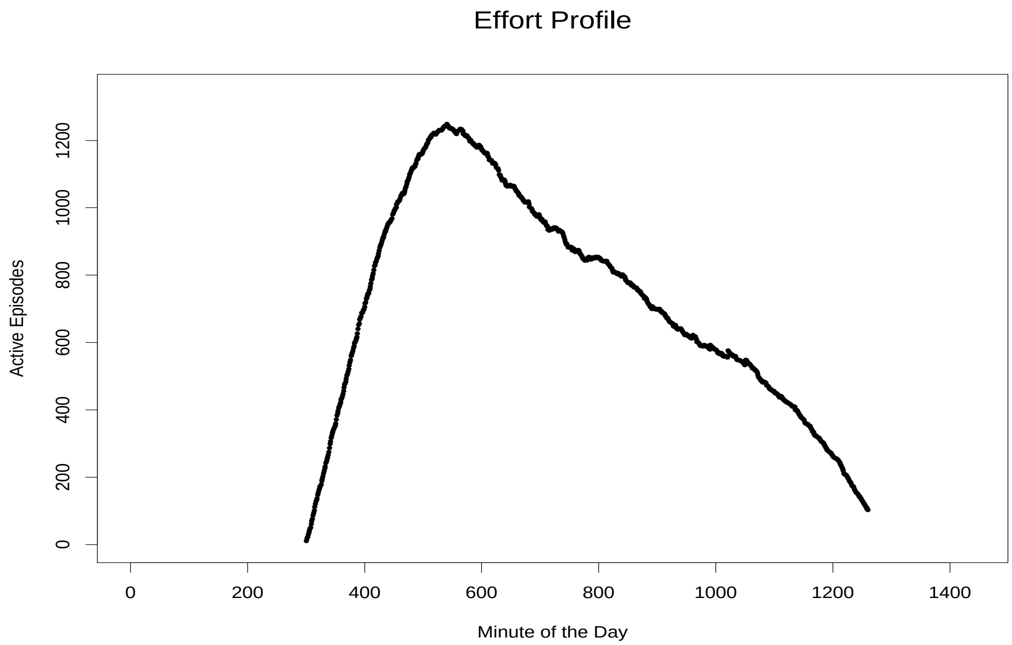

Figure 4.3: Simulated Data: Effort profile for each scenario.

per hour. For Scenario II the variability was restricted to among Blocks *i.e.* a time of day catch rate. While an episode was active during Block I times, it was given a catch rate of 2.0 fish per hour; while active during Block II times, it was given a catch rate of 1.5 fish per hour; while active during Block III times, it was given a catch rate of 1.0 fish per hour; and while active during Block IV times, it was given a catch rate of 0.5 fish per hour. For Scenario III the variability was restricted to among Launches *i.e.* a location catch rate. If an episode was in either Launch 1, 2, 11, or 12, it was given a catch rate of 2.0; if in Launch 2, 4, 5, 13, 14, or 15, it was given a catch rate of 1.5; if in Launch 6, 7, 8, 16, 17, or 18, it was given a catch rate of 1.0; and if in Launch 9, 10, 19, or 20, it was given a catch rate of 0.5. For Scenario IV the variability was present at both the Block and the Launch levels. Catch rates were again 2.0, 1.5, 1.0, or 0.5 for Block I to IV, respectively, but now with an an augmentation depending on Launch. These were an additional rate of 1.5 for Launches 1, 2, 11, or 12; an additional rate of 1.0 for Launches 3, 4, 5, 13, 14, or 15; an additional rate of 0.5 for Launches 6, 7, 8, 16, 17, or 18; and 0.0 additional rate for Launches 9, 10, 19, or 20. A total of 70,591 catches were generated. A Tabulation by Scenario and Block of catch is given in Table 4.2. Block I episodes could be active only during Block I times, however,

Table 4.2: Simulated data: Population catch by scenario.

		Scenario I	Scenario II	Scenario III	Scenario IV
Block 1	Episodes	193	193	193	193
	R	4.5403	6.9435	4.6774	9.0968
	Catch	563	861	580	1,128
Block 2	Episodes	1,083	1,083	1,083	1,083
	R	6.4889	8.8746	6.5650	12.4302
	Catch	5,535	7,570	5,600	10,603
Block 3	Episodes	729	729	729	729
	R	4.5756	5.1197	4.8067	7.9286
	Catch	4,356	4,874	4,576	7,548
Block 4	Episodes	665	665	665	665
	R	6.4448	4.6205	6.6096	8.4479
	Catch	4,144	2,971	4,250	5,432
Total	Episodes	2,670	2,670	2,670	2,670
	R	5.6757	6.3281	5.8344	9.6077
	Catch	14,598	16,276	15,006	24,711

Block II episodes could be active during both Block I and Block II times. Similarly, Block III and Block IV episodes could be active during preceding block times. Also catch rates vary by time of day (*i.e.* across Blocks) for Scenarios II and IV. These factors are reflected in the differing values of catch. For any scenario, the overflight counts are 124, 853, 952, and 643 for Blocks I, II, III, and IV respectively. Since R is a ratio estimator, it can be expected that in general, estimates produced from scenarios making greater use of the auxiliary data, and for which there is less variability in catch between episodes, should have greater success. Plots of episode catch versus number of overflight surveys, in which the episode was counted, indicate linear patterns with positive slopes (Figure 4.4). While not through the origin, it is

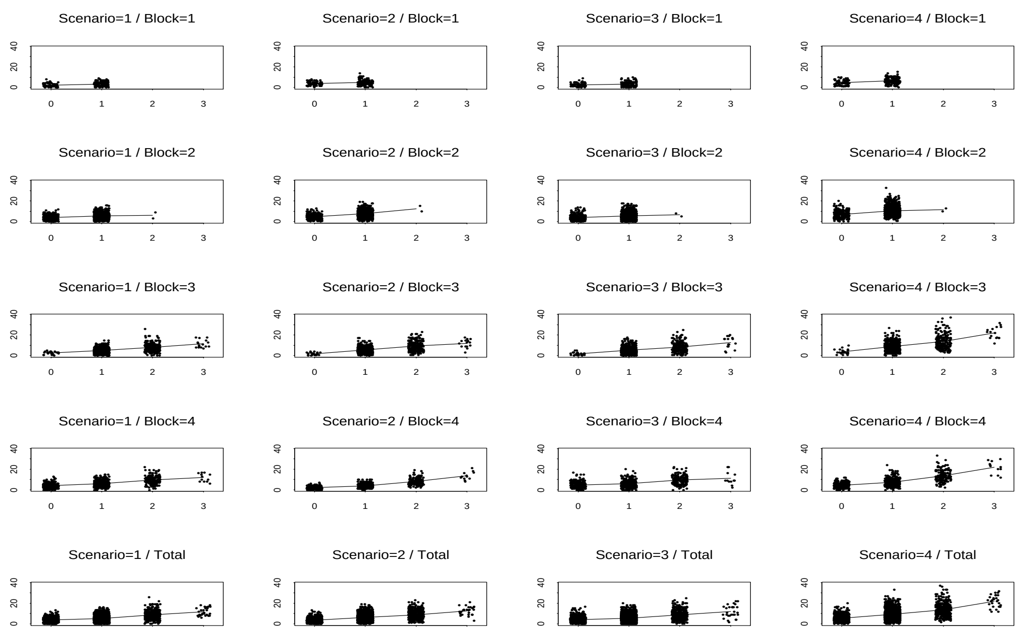

Figure 4.4: Simulated Data: Fish caught by number of times counted in the overflight surveys. (Lines connect means of catch in each group.)

still felt that overflight counts should be useful auxiliary information for a ratio estimator.

For initial selection, Launches 1 and 11 had a 0.0675 probability of selection; Launches 2 and 12, a 0.0625 probability; Launches 3 and 13, a 0.0550 probability; Launches 4, 5, 14, and 15, a 0.0500 probability; and Launches 9, 10, 19, and 20, a 0.0400 probability. Block

I was selected with probability 0.10; Block II with probability 0.40; and Blocks III and IV with probability 0.25. For any Launch-Block combination, episodes were selected with equal probability. For each combination of high and low sampling rate on *Launch*, *Block*, and *Episode*, 500 replicated simulations were conducted. Sample selection was done without replacement using sample inclusion probabilities based on these initial selection probabilities, however, the sample inclusion probabilities τ_l, π_{lb}, and ω_{lbi} used in the estimation formulae were approximated by these selection probabilities multiplied by sample size. This approximation was done to test the robustness of the estimators, in particular, in a scenario similar to that of the Georgia Strait Creel Study. See Appendix C for more exact results based on proper estimates of τ_l, π_{lb} and ω_{lbi}.

The performance of $\hat{R}$ is given in Table 4.3. As expected, precision is dependent on sampling rates and placement of variation within the multistage structure. Inherent to ratio estimators, $\hat{R}$ shows negative bias (Cochran, 1977), that, with few exceptions does not exceed 5 %. A trend of increasing bias (more noticeable at lower levels of sampling) can be seen moving from Scenario I to Scenario III. This reflects the structure of Equation 4.1 where an inaccurate result at the "inner" (*i.e.* *Episode*) level has a greater expansion than one at the "outer" (*i.e.* *Launch*) level. Scenario IV has amplified variability at all levels and generally shows the greatest amount of bias.

The performance of the variance estimators are given in Table 4.4. Predictably, the $Var_s(\hat{R})$ variance estimates, calculated using $\hat{R}$ from the 500 simulated samples, varied according to scenario (*i.e.* variability at the different sampling stages) and sampling intensity. The formula variance estimator, $\widehat{Var}(\hat{R})$, based on the conservative Brewer-Hanif approximation, tends to overestimate the variance. In general, the jackknife estimator, $\widehat{Var}_J(\hat{R})$,

Table 4.3: Simulated data: Performance of the $\hat{R}$ estimator for differing sampling rates on Launch Site, Interview Blocks and fishing Episodes (L=25%; H=75%). Catch rates vary by Episode in Scenario I, vary by Block in Scenario II, vary by Launch in Scenario III and vary by Block and Launch in Scenario IV. (500 replicated simulations.)

Sampling Percentages nL	nB	nI		Scenario I	Scenario II	Scenario III	Scenario IV
			R (population)	5.676	6.328	5.834	9.608
L	L	L	$Mean_s(\hat{R})$	5.672	6.216	5.711	9.452
			Bias (%)	-0.1	-1.8	-2.1	-1.6
			c.v. (%)	8.0	8.9	17.5	10.8
L	L	H	$Mean_s(\hat{R})$	5.646	6.120	5.715	9.455
			Bias (%)	-0.5	-3.3	-2.0	-1.6
			c.v. (%)	5.8	7.1	16.6	9.8
L	H	L	$Mean_s(\hat{R})$	5.608	6.040	5.688	9.242
			Bias (%)	-1.2	-4.6	-2.5	-3.8
			c.v. (%)	5.8	5.9	15.8	9.8
L	H	H	$Mean_s(\hat{R})$	5.575	6.050	5.690	9.222
			Bias (%)	-1.8	-4.4	-2.5	-4.0
			c.v. (%)	3.6	3.7	15.1	8.8
H	L	L	$Mean_s(\hat{R})$	5.614	6.212	5.598	9.281
			Bias (%)	-1.1	-1.8	-4.0	-3.4
			c.v. (%)	4.6	5.0	7.0	4.8
H	L	H	$Mean_s(\hat{R})$	5.621	6.222	5.583	9.279
			Bias (%)	-1.0	-1.7	-4.3	-3.4
			c.v. (%)	3.3	4.1	6.4	4.0
H	H	L	$Mean_s(\hat{R})$	5.566	6.081	5.564	9.120
			Bias (%)	-1.9	-3.9	-4.6	-5.1
			c.v. (%)	3.3	3.4	6.3	4.0
H	H	H	$Mean_s(\hat{R})$	5.564	6.077	5.550	9.118
			Bias (%)	-2.0	-4.0	-4.9	-5.1
			c.v. (%)	1.9	2.3	5.6	3.2

Table 4.4: Simulated data: Comparison of variance and variance estimates of $\hat{R}$ for differing sampling rates on Launch Site, Interview Blocks and fishing Episodes (L=25%; H=75%). $V_s(\hat{R})$ denotes the variance of the simulated $\hat{R}$ values, $\widehat{V}(\hat{R})$ the mean of formula estimates and $\widehat{V}_j(\hat{R})$ the mean of jackknife estimates (without a finite population correction factor). Catch rates vary by Episode in Scenario I, vary by Block in Scenario II, vary by Launch in Scenario III and vary by Block and Launch in Scenario IV. (500 replicated simulations.)

Sampling Percentages nL	nB	nI		Scenario I	Scenario II	Scenario III	Scenario IV
L	L	L	$V_s(\hat{R})$	0.204	0.307	0.998	1.051
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	2.18	2.18	1.16	1.63
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.15	1.02	1.23	1.30
L	L	H	$V_s(\hat{R})$	0.108	0.190	0.896	0.865
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.79	2.24	1.05	1.42
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.06	1.07	1.26	1.54
L	H	L	$V_s(\hat{R})$	0.105	0.127	0.812	0.819
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	2.10	2.19	1.21	1.31
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.10	1.10	1.42	1.28
L	H	H	$V_s(\hat{R})$	0.040	0.050	0.737	0.666
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.90	2.64	1.15	1.17
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.25	1.30	1.47	1.33
H	L	L	$V_s(\hat{R})$	0.066	0.098	0.155	0.199
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.59	1.74	1.23	1.67
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.06	1.04	2.70	2.14
H	L	H	$V_s(\hat{R})$	0.034	0.066	0.126	0.140
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.32	1.65	1.06	1.55
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.09	0.98	3.05	2.60
H	H	L	$V_s(\hat{R})$	0.033	0.044	0.123	0.134
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.64	1.59	1.10	1.34
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.18	1.05	3.02	2.51
H	H	H	$V_s(\hat{R})$	0.011	0.019	0.096	0.087
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.64	1.79	1.06	1.30
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.36	1.16	3.63	3.34

performed better than $\widehat{Var}(\hat{R})$ when variability was confined to the fishing episodes (Scenario I) or to time of day (Scenario II) where its better performance was more noticeable at the lower sampling rates. When variability was confined to area (Scenario III), the formula estimator performed better, particularly with higher sampling rates amongst landing sites. With variability at all levels (Scenario IV), formula estimates generally performed better than jackknife estimates. A more uniform performance throughout was given by the jackknife estimator. One reason for this is its avoidance of estimating the individual variance components, given in Equation 4.2, which require values for quantities such as s_z^2 which, in turn, use quantities such as $z_{lbi} = \frac{1}{\omega_{lbi}} \left(C^*_{lbi} - \hat{R}\, m^{(A)}_{lbi} \right)$. With small sample sizes, relatively large variations in the $m^{(A)}_{lbi}$ are possible resulting in extremes in z_{lbi} and the associated variance components. The larger variances occurring with the high sampling rates of Launches in Scenarios III and IV can be explained by the lack of a finite population correction factor.

Table 4.3 also shows an increase in bias accompanying an increase in sampling rate. This can be explained by the method of calculating the inclusion probabilities. As constructed, they tend to $N \times$ {initial probability of selection} as sample size increases which fails to adjust for the probability of other units already selected. Thus, when sample size reaches N, individual units will not necessarily have inclusion probabilities (*i.e.* weights) equal to 1.0. Then quantities such as Total Catch $\neq \sum_j^N C_j$. Variance estimates are also affected by these incorrect inclusion probabilities. See Appendix C.

4.7 The Georgia Strait Creel Survey

Commercial, Native, and Recreational fishers from both Canada and the United States depend on fish stocks of the Georgia Strait, a region in excess of 5,900 km^2 of water surface and 2,400 km of shoreline located between Vancouver Island and the west coast of Canada. In 1980 the Department of Fisheries and Oceans (DFO) piloted its Strait of Georgia Creel Survey program for the sports fishery sector. The region under study is divided into nine areas from which data are collected using two complemented surveys: an access survey interviewing boaters on completion of their outings and an aerial program providing counts of boats actively fishing. The initial objective was to provide "accurate and timely sport catch statistics primarily for chinook and coho". Over the years the study has remained relatively unchanged in design but has been expanded to include all species including ground and shell fish, and other monitoring programs (Shardlow and Collicutt, 1989). The 1998, July and August data for this ongoing study were used to illustrate the proposed methodology.

Stratification for the access survey is done over month, area, day type (*i.e.* weekend/holiday versus weekday) and guided versus unguided anglers. Using a multistage design on selected days, interviews are conducted on completion of boat-trips within selected shifts (*i.e.* "interview work blocks") at selected launching sites within each area.

> "In each region, various landing sites were chosen as location for surveyors to conduct interviews. Site selection was based on 4 criteria: representativeness, traffic volume, site accessibility and adequate observation points. Discussions with local fishers, marina operators and Fisheries Officers and data from previous surveys were used to choose sites that were representative of local sport fishing activity (*i.e.* sites which were used by a wide cross-section of anglers). Sites with expected traffic volumes of more than 15 boats per day in the summer were considered as possible sampling locations. Expected traffic volumes for sites were compiled from previous surveys or from discussions with marina operators

> or local Fisheries Officers.
>
> Site accessibility refers to whether an interviewer can easily reach a site by car or ferry during the defined shift hours. Only sites with good accessibility were selected. As a result, landing sites on most of the islands in the Strait of Georgia were excluded from the survey. This was not expected to be a major factor, however, since most of the fishing that occurs in these areas is from boats launching from an accessible site. The final criterion, adequate observation points, was essential for interviewers to obtain an accurate count of all boats returning to a landing site. At some large marinas, where the number of access points made it impossible to see all boats returning, the facility was defined as two separate sites.
>
> Allocation of sampling effort among months followed the same general pattern as fishing effort, that is, more effort was allocated during the summer when fishing effort is at its highest. Allocation of sampling effort among regions ... also followed fishing effort patterns. Within each month, each chosen site was allocated between 6 and 10 shifts. These shifts were divided equally among weekend and mid-week days and early and late daily time periods. (Hardie et. al., 1999)."

For every interview, the date, area, landing site, and hour of trip completion (and hence shift) are recorded. If it is determined that the purpose of the boat-trip was fishing, then the interview continued, recording the length of time fished (and hence start time). In addition, if the trip was successful, information on effort, catch, release, and other data of interest by species at the sub-area level is also recorded.

Overflight counts were conducted using one fixed wing aircraft. Routing and time of departure were selected to maximize the count of active fishing episodes over the entire study region.

> "Aerial surveys, conducted from airplanes travelling along pre-defined routes, allowed observers to count vessels actively sport fishing throughout the Strait of Georgia. Planes flew at an altitude of 150-210 m to facilitate a broad range of vision and still allow easy identification of vessel characteristics. ...
>
> The flight path and time of departure were designed to cover major concentrations of sport fishing activity at peak periods. To maximise precision, flying times during which fishing effort was rapidly changing were avoided. The number of overflights each month was governed by budget constraints, targets of desired precision and by the expected number of interviews from a given number of sampling shifts (English et. al. 1986). The days for overflights during a month were randomly selected for each day type. (Hardie et. al., 1999)."

These two components are then used to form a ratio estimator for catch.

> "These data are used to calculate catch per boat trip for a catch region (CPE) and the proportion of the days total fishing effort that occurred in a given hour of the day (P).
>
> The proportion of daily boat trips fishing at various times during the day, the activity pattern, is formed using the relationship:
>
> $$P_i = \frac{\sum_j^{NS} \sum_k^{NB} FT_{ijk} \cdot W_{jk}}{\sum_j^{NS} \sum_k^{NB} (\sum_i FT_{ijk}) \cdot W_{jk}}$$

where

P_i	=	proportion of daily fishing trips that included fishing time block i ["Fishing Time Block - There are 16 one hour fishing time blocks in each day."],
FT_{ijk}	=	number of fishing trips that included fishing time block i that landed at site j during interview work block k ["Interview Work Block - There are four four-hour interview work blocks in each day."],
W_{jk}	=	a weighting factor which adjusts interview data for sampling effort both within and between work blocks. (*i.e.*, if 10% of the morning work blocks in a month were worked and 50% of the boats landing were interviewed the weight factor would be 20.),
NS	=	number of landing sites sampled,
NB	=	number of interview work blocks (normally 4 per day).

Aerial surveys are conducted so observers can count all the sports boats (Y) actively fishing [defined as stationary or with only moderate bow wave] in specified sub-regions of Georgia Strait at the time of the survey. The above data can be combined to estimate catch and effort by statistical area and month:

$$\text{EFFORT} = Y \cdot \frac{1}{P}$$
$$\text{CATCH} = Y \cdot \frac{1}{P} \cdot \text{CPE}$$

... The means and standard deviations of catch per effort ... were calculated

... Variances of the means were calculated using the following standard variance formula:

$$S_{\bar{x}}^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n(n-1)}$$

where x is the unweighted catch by statistical area, month, method [e.g. bait versus lure] and species for each interview and n is the number of interviews in each category (English et. al., 1986)."

Notwithstanding a compromise of randomization in the access survey, the structure of the Strait of Georgia Creel Survey resembles that of Section 4.3. However, to apply the proposed methods, some adaptations were necessary. In particular, the proposed method requires that, with each interview session, a sample can be selected such that at least one episode is counted in an overflight survey. With interview sessions occurring on days with no overflight, this was not the case in the DFO survey. However, since overflights were flown at regularly scheduled times, it was possible to extend the days of operation of the overflight schedule to include all days with interview sessions using the means of actual entry times into the subareas by day type. Counts for these additional overflights were imputed based on the means of actual counts using same area and day type combinations. Note that variability is now introduced into $N^{(O)}$ by this estimation, however, in the analysis, no account was made for this extra variability.

With this extended overflight schedule it was then possible to view a month as a "single fishing day" with multiple overflights, and the design as multistage with boat interview within day-shift (defined as block) within landing site within area. Taken this way, stratification over day type was implicit.

Non-response was relatively low (1.5% in July). A random number and proportions by area, day type, landing site and shift were used to impute the fishing status of boaters that refused to be interviewed. Similar categorizing was used to determine δ status and number of fish caught on these boat trips.

Increasingly, site and shift selection has been based on an objective of maximizing interview counts rather than random selection. Accordingly, inclusion probabilities were approximated by observed frequency of occurrence. Thus, for a given block (*i.e.* day-shift) in July, the inclusion probability was taken to be the proportion of times that shift was actually selected within its area divided by 31. Boat interviews were taken to have equal probabilities and it was assumed that all potential interviews within a selected block were conducted.

Table 4.5 gives the July 1998 Chinook salmon catch estimates and their standard errors for each of the nine areas and the regional total using the proposed methodology, and compares these with the regional estimate given in the DFO study (Hardie et. al., 1999). This was also done using equal probabilities for site and block selection. Table 4.6 provides the same sets of comparisons for August 1998 Chinook catch, while Tables 4.7 and 4.8 provide July and August comparisons for Rockfish catch.

It was expected that regional DFO estimates and those made using the proposed methods would have been similar. Lack of randomness in the access survey makes it difficult to establish a claim as to which estimate might be more accurate, however, apparent discrepancies in documentation point to some potential theoretical difficulties. For example note that for time block t in a fishing day with T time blocks, each one hour long, and assuming

Table 4.5: Georgia Strait Creel Survey: Comparison of July 1998 Chinook catch estimates using various methods of analysis with estimates made by DFO. The first comparison applies the proposed multistage technique using unequal site and block selection probabilities based on observed frequencies while the second uses equal selection probabilities. The third comparison assumes a simple random sample of interviews.

		Unequal Probabilities		Equal Probabilities		Random Sample	
	Area	Est	s.e.	Est	s.e.	Est	s.e.
	13	1,256	609	1,314	671	1,228	209
	14	170	121	159	111	179	49
	15*	0	0	0	0	0	0
	16	232	254	224	250	216	104
	17	125	89	125	89	127	34
	18	23	38	23	38	25	29
	19	1,010	331	1,009	332	991	174
	28	166	123	163	119	122	47
	29	4	7	11	12	35	21
	Total Region	2,986	764	3,029	812	2,922	303
DFO	Total Region	4,143	1,150	4,143	1,150	4,143	1,150

* No catch was recorded by interviewers in Area 15.

Table 4.6: Georgia Strait Creel Survey: Comparison of August 1998 Chinook catch estimates using various methods of analysis with estimates made by DFO. The first comparison applies the proposed multistage technique using unequal site and block selection probabilities based on observed frequencies while the second uses equal selection probabilities. The third comparison assumes a simple random sample of interviews.

		Unequal Probabilities		Equal Probabilities		Random Sample	
	Area	Est	s.e.	Est	s.e.	Est	s.e.
	13	1,023	347	1,039	340	1,072	167
	14	695	469	798	587	853	138
	15	65	94	62	93	37	35
	16	168	201	203	230	267	154
	17	205	128	207	123	270	64
	18	190	98	190	98	218	61
	19	1,547	352	1,571	370	1,588	201
	28	33	56	33	56	38	38
	29	50	39	45	36	41	16
	Total Region	3,977	738	4,148	829	4,382	349
DFO	Total Region	7,163	1,721	7,163	1,721	7,163	1,721

Table 4.7: Georgia Strait Creel Survey: Comparison of July 1998 Rockfish catch estimates using various methods of analysis with estimates made by DFO. The first comparison applies the proposed multistage technique using unequal site and block selection probabilities based on observed frequencies while the second uses equal selection probabilities. The third comparison assumes a simple random sample of interviews.

		Unequal Probabilities		Equal Probabilities		Random Sample	
	Area	Est	s.e.	Est	s.e.	Est	s.e.
	13	851	826	827	795	737	276
	14	286	242	252	197	284	86
	15*	61	2,689	65	3,911	78	67
	16	642	511	621	490	592	272
	17	291	211	291	211	296	86
	18	815	798	815	798	844	462
	19	2,392	1,323	2,422	1,347	2,489	650
	28	429	580	396	572	471	280
	29	9	14	25	26	18	5
	Total Region	5,774	3,316	5,713	4,362	5,808	940
DFO	Total Region	19,163	5,044	19,163	5,044	19,163	5,044

* No overflight counts were made in Area 15. To avoid division by 0, a default of 1 was used.

Table 4.8: Georgia Strait Creel Survey: Comparison of August 1998 Rockfish catch estimates using various methods of analysis with estimates made by DFO. The first comparison applies the proposed multistage technique using unequal site and block selection probabilities based on observed frequencies while the second uses equal selection probabilities. The third comparison assumes a simple random sample of interviews.

		Unequal Probabilities		Equal Probabilities		Random Sample	
	Area	Est	s.e.	Est	s.e.	Est	s.e.
	13	955	854	960	810	982	316
	14	122	93	143	106	156	62
	15*	0	0	0	0	0	0
	16	893	728	854	670	713	345
	17	324	206	322	198	259	69
	18	123	96	123	96	117	45
	19	1,886	1,005	1,959	1,039	2,095	467
	28	623	642	623	642	682	443
	29*	0	0	0	0	0	0
	Total Region	4,926	1,656	4,984	1,631	5,003	803
DFO	Total Region	21,571	5,377	21,571	5,377	21,571	5,377

* No catch was recorded by interviewers in Areas 15 and 29.

that all episodes are sampled (*i.e.* weighting factors are equal to 1)

$$P_t = \frac{n_t}{\sum_{t=1}^{T} n_t} = \frac{n_t}{T \cdot E\left[n_t\right]} = \frac{n_t}{\sum_{j=1}^{N} L_j^*}$$

where L_j^* is the length of time fished by the j^{th} episode. Then

$$\text{EFFORT} = Y_t \cdot \frac{1}{P_t} = \frac{Y_t}{n_t} \sum_{j=1}^{N} L_j^*$$

where Y_t / n_t is a sample-to-population expansion factor and EFFORT is a measure with units "fishing hours". However, CPE is measured in catch per boat, and consequently CATCH $= Y_t \cdot \frac{1}{P_t} \cdot CPE$ would tend to overestimate the catch by a factor approximately equal to the average fishing duration in hours.

Note that if "fishing hours" were used to measure effort, then effort profiles must be constructed for each species for which estimates are to be made. This could prove difficult and even beyond the capability of the data for specified accuracy requirements.

Alternatively, disregard the formula and take P_i to be "the proportion of daily fishing trips that include time block i", then, if Y is the count of boats actively fishing at time block i, $Y \cdot (1/P)$ will estimate the number of boats actively fishing that day and $Y \cdot (1/P) \cdot CPE$ is an estimate of catch for the day. Concern now centres around the weights W_{jk} which scales the number of boats that fished to the number of boats that landed. Also, for this method of catch estimation, overflight counts must be made for each day. This was not the case but, assuming imputations, no methodology was given.

Further investigation has shown that DFO imputations were performed with correct understanding of the underlying theory. However, details of the programming were not available. Work is currently underway to discover the cause of the discrepancy between the two methods.

Table 4.9: Georgia Strait Creel Survey: Components of variance analysis 1998 for selected months and species.

		Variance Component (Percentage)			
		Stratum	Site	Block	Interview
Chinook	Jul	0.1	2.4	8.7	88.8
	Aug	0.1	3.2	3.8	92.9
Rockfish	Jul	0.3	0.1	7.2	92.4
	Aug	1.7	0.6	9.9	87.8

A variance component decomposition (Table 4.9) showed that most of the variation occurred at the episode level with little contribution from stratum, site, or block. The third set of estimates on Tables 4.5, 4.6, 4.7, and 4.8 were derived by ignoring the site and block stage of the design and treating the interviews, that had been selected within each stratum by day type, as a simple random sample. The methods of Chapter 3 can then be used. Accordingly,

$$\hat{C} = \sum_{i=1}^{n} C_i^* \cdot \frac{N^{(O)}}{N^{(A)}}$$

with

$$\widehat{Var}\left(\hat{C}\right) = \frac{n}{\hat{f}} \left(\frac{1}{\hat{f}} - 1\right) s_z^2$$

where

$$s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i)^2 \quad \text{using} \quad z_i = C_i^* - \frac{\sum_{i=1}^{n} C_i^*}{N^{(A)}} \cdot m_i^{(A)} \quad \text{and}$$

$$\hat{f} = \frac{N^{(A)}}{N^{(O)}}$$

was used to construct the estimates.

For this approach to be appropriate, it is necessary to assume that there is similarity in catch rates and fishing patterns from day to day in mornings versus afternoons. Then, with an approximately even split on morning versus afternoon shifts, it might be argued that the set of interviews actually sampled could have been realized using a simple random sample and the estimator $\hat{C} = N \sum_{i=1}^{n} C_i^* \Big/ n$, the usual estimate for population total in a simple random sample (Cochran, 1977), should be used. But, since $\hat{f} = N^{(A)}/N^{(O)} \approx n/N$, the estimator actually used becomes $\hat{C} = \sum_{i=1}^{n} C_i^* \cdot N^{(O)}/N^{(A)} \approx N \sum_{i=1}^{n} C_i^*/n$. Therefore, since the catch rate and fishing pattern assumptions seem reasonable, it is not surprising that estimates of catch by area are similar to those made using the full multistage structure.

Variance estimates may not be exact because of the lack of full randomization, however, results indicate that there is little difference if the data is analyzed as a multistage or as a simple random sample. This is not surprising given the components of variance (Table 4.9). Also, simulations (Table 4.4, Scenario I) have shown that when the variability is strictly among Episodes, the variance estimator for the proposed method can be expected to work reasonably well. (Jackknife estimators can also be expected to work well.) It is interesting to note that the coefficient of variation for the DFO estimate, using a variance estimator that ignores the ratio aspects of their catch estimator, is similar to that for the proposed methods (but not the analysis done as a simple random sample).

The components of variance clearly show that the greatest source of variance is between interviews and correspondingly, a best survey design should allocate most of the resources and effort to obtaining a large random sample of interviews. Concern about variability between blocks should be a distant second, however a strategy is less clear. Recall that block is actually day-shift and a large portion of the block variability may lie in only one of these components, *i.e.* among days or amongst shifts. Intuitively, a broad sample over both would seem appropriate. It is also appealing, though not indicated by the components analysis, that sampling be done over sites to some extent. The need to measure over area might well be only for administrative or reporting purposes.

The requirement of a minimum value for counts in the access survey is highlighted in Table 4.7. For Area 15, extreme variance estimates are generated by the leverage in small values of $m^{(A)}$. To avoid division by zero within the estimator, a minimum default value of one was used. Results should be monitored for such occurrences where one remedy is to collapse strata.

Sample inclusion probabilities for launch and block were calculated as the product of relative frequency of actual selection and sample size. With these frequencies nearly uniform, effects of this choice of inclusion probability on catch estimates were minor. In addition, because all boat trips were sampled, inclusion probabilities were not an issue with interviews and overall error was minimal, as seen if estimates in the unequal and equal probability columns in Tables 4.5 to 4.8 are compared. Also, since most of the variability was among interviews, there was little effect on variance estimates.

In applying the proposed estimator to the Georgia Strait Creel Survey data a number of issues, specific to the application and general in nature, were encountered.

1. The access survey must be completely randomized. With lack of randomization in the overflight survey, this is essential. Any compromise is a potential source of bias and destroys the defensibility of the estimates. This is true no matter how the data are analyzed.

2. The population of boats counted by the overflight surveys should be the same population that uses the landing site, and all landing sites should have positive probability of being selected. It is of little consequence if boats in this population actually fish. If not, a zero catch is assigned and there is no modification to the estimators.

3. It is of vital concern that it be known whether or not an interviewed boat was counted by an overflight survey. To this end, any combination of two of start time, end time or time fished should be recorded. This includes non-fishing interviews if they are part of the population of overflight counts. When start times for flights through subareas are separated by enough time to invalidate an "instantaneous" count using a single area time, care should be taken to accommodate this on the interview.

4. Overflight counts are useful only if some of the boats will be interviewed. Without a "match" to an interview session, such counts only serve to arbitrarily inflate $N^{(O)}$, and correspondingly $\hat{C}$. Data obtained from overflights in areas on days with no interviewing should, therefore, be discarded. Also, if there is not at least one interview with $\delta = 1$ (*i.e.* indicator for being counted), the estimator for R involves division by 0 and breaks down. To achieve this, overflights may have to be scheduled daily. Recall that in the analysis, block = day–shift was introduced, as was an extended overflight schedule. This arrangement preserved the multistage design by keeping blocks within

landing sites and, by now having 31 overflights in the single month-long fishing day, at least one interview from any block had $\delta = 1$. Note that if "extending" the schedule is done in practice, variability in $N^{(O)}$ is introduced which should be dealt with (not done in the above analysis). Catch is now the product of two unbiased, independent estimators (Goodman, 1960).

5. When landing sites service significant numbers of boats fishing in waters outside their area, the strict nesting required of a multistage design is violated. These sites might be considered an augmentation to the list of sites for the fished areas. Their interviews could then be included with the others of the area if an appropriate probability of selection were constructed. If so, some means of sorting out overflight counts would also be required. If these issues could not be resolved or a workable interviewing protocol devised, then combining areas might have to be considered.

4.8 Conclusions and Discussion

When catch estimation using complemented angler surveys with restricted randomization is embedded in a complex survey design, the estimator developed in Chapter 3, $\hat{C} = \hat{R} \cdot N^{(O)}$, must be adjusted accordingly. In simple sampling designs, $\hat{R}$ is the ratio of the raw sample catch to the raw overflight count in the sample. In this way the sample catch is "ratioed" up to the population level. In complex designs, $\hat{R}$ becomes the ratio of an estimated population catch to an estimated population overflight count. The catch estimate made using the access survey is adjusted according to how well the expansion "factor" used to expand the sample catch performs in expanding the sample overflight counts to population overflight counts. That is, the auxiliary information is used to refine the expansion of the catch to population estimate. The catch estimate must be able to stand alone. This is the reason, if results are to be defensible, that the access survey must be fully randomized.

One of the advantages of the proposed methodology is that the overflight surveys need not restrict their counts to boats which produce the catch that is to be estimated. The only requirement is that the population being counted must also be the population that is subject to possible interview. Boats with activity other than fishing or even fishing but targeting a species that is not of interest are merely assigned a catch of zero with no modification to the methodology. This eliminates the difficulties associated with selective counting and complex structuring of specific effort profiles. Such profiles add additional variability and because they combine results from both the overflight and access surveys, can present difficulties with theoretical aspects such as finding expected values of estimators (essential in determining an appropriate estimator, see for example Pollock et. al., 1997) and constructing variance formulae. Notice that while true effort curves are essential in the more standard techniques, exact measures of effort on the target species and issues of independence of effort between species are not a concern in the proposed methodology.

A requirement of the proposed methodology is that at least one of the sampled boat trips be active during an overflight count. Because small sample counts have high leverage, larger counts are desirable and overflights should be scheduled, if feasible, to result in high counts in any possible access sample. This goal of high counts complements tolerance in the overflight counting of boats other than those fishing the targeted species.

In general, the formula variance of the proposed methodology is conservative owing to its construction based on the Brewer-Hanif approximation. Also, it may be tedious to develop. For scenarios for which a components of variance analysis shows that the variability is confined to the lower levels of the sampling structure, as with the Georgia Strait Creel Survey, a jackknife estimator seems to work well as an alternative.

It is not known whether the results produced by methods involving restricted randomization in the overflight survey are, in general, better than those with randomization in both the access and overflight surveys using the more usual $\hat{C} = \widehat{CPUE} \cdot \hat{E}$ estimator. With the assumptions required for estimating catch rate generally less restrictive than those for estimating catch, given a simple effort profile, the latter is likely the method of choice, however, this is a matter of further research. Another area requiring further research is that relating to a minimum acceptable value for $\sum_{i=1}^{nI_{alb}} m_{albi}^{(A)}$, or equivalently, a sufficient size for the total of all counts made by all overflight surveys of episodes selected for sample. Also requiring further research is sensitivity to incorrect inclusion probabilities. Preliminary results appear to show that the "naïve" approach to forming inclusion probabilities, by using a simple product of initial selection probability and sample size, may be acceptable if the initial selection probabilities are not too different from $1/N$.

Because implementation of the DFO estimator requires, as a minimum, full randomization of the access survey, their estimates may not be unbiased, however, the magnitude of the difference in these estimates from those obtained using the proposed methods remains disturbing. One explanation could be use of overall rather than species specific effort profiles. As a first step in rationalizing the difference, it is proposed that a sample, or series of samples, from the simulation data (for which there is a known population catch) be run through the DFO programs and the programs for the proposed estimator.

4.9 Acknowledgments

Data from The Georgia Strait Creel Surveys were obtained from Department of Fisheries and Oceans, 3225 Stephenson Point Road, Nanaimo, B.C., V9T 1K3. The research was supported by Research Grants from the Natural Science and Engineering Research Council (NSERC) of Canada.

Chapter 5

Summary and Future Research

Over the years, complemented surveys with catch estimation based on $\mathit{Effort} \times CPUE$ has been one of the most widely used techniques in creel survey work. Under standard conditions where simple catch rate assumptions and simple access or roving sampling designs suffice, underlying theory is well developed (Pollock et al., 1994) and clear guidelines for the appropriate choice of estimators have been outlined (Pollock et al., 1997). This thesis has extended the theory to include other conditions which are apt to occur in many applications.

For applications in which it is reasonable to assume that the catch rate of each fishing episode declines (the standard assumption is a constant catch rate over the entire duration of each episode), Chapter 2 proposes an alternative catch estimator and a method of estimating its variance.

- **Future research.** Consider other catch rate models e.g. a non-homogeneous Poisson process. To be suitable for a live bait fishery, the decline in catch rate might more suitably be based on time fished. Also, a model involving more parameters might more aptly describe data patterns (Figure 2.3).

The variance of $\hat{R}$, the catch rate estimator (Equation 2.3), involves the expectation of the reciprocals of L_j and L_j^2 (*i.e.* time to interview of sampled episodes) and as such, does not exist unless bounded from zero. Also, with the possibility of small L_j, $\widehat{Var}(\hat{R})$ is extremely unstable. One method to deal with this problem is use of L', a minimum time to interview required for sample inclusion of the selected episodes (a technique also used by Hoenig et al., 1997 and Pollock et al., 1997). It was shown (Table 2.3) that variance does become more stable as L' increases, however, it was seen (and shown analytically) that a bias induced in $\hat{R}$ also increases as L' increases.

- **Future research.** Investigate an optimal value of L' and the conditions that determine this value.

For estimators to be unbiased, the starting time of the roving survey must be randomly determined in such a way that all times in the full fishing day have a positive probability of being included. In some applications it may be a practical necessity to restrict the window of opportunity for the roving survey. Such restrictions result in episodes overlapping the limits of the window and, in general, lower estimates of catch rate.

- **Future research.** Investigate correction procedures for window restrictions. One consideration is a restricted window that is not symmetric with respect to the effort profile *i.e.* an irregular fishing preference profile or restrictions such that the window is situated nearer one end of the fishing day.

- **Future research.** Investigate the effects of fishing episodes overlapping the current fishing day e.g. a gill net fishery in which nets may soak from one day into the next.

- **Future research.** Develop appropriate methods for night versus day fisheries.

In Chapter 3, methods were developed for applications in which randomization of the effort survey is not possible or practical. Ratio type estimators were developed for use with both access and roving ground surveys for estimating $CPUE$. For roving surveys it was assumed that fishers could accurately estimate the end time of their fishing episodes. Consequently, in the development of estimators, L_i^* was used in place of the usual estimator, $2L_i$. Single as well as multiple overflight counts were considered and the question of how best to use this auxiliary data (*i.e.* ratios of means or mean of ratios) was examined. A homogeneous Poisson process was used to model a constant catch rate.

- **Future research.** Investigate the relative advantages of using L_i^* estimates provided by interviewed fishers in place of the more conventional $2L_i$ estimator in roving survey estimation. Of interest are the potential gains in precision and the effects of recall bias.

- **Future research.** Investigate the appropriate value for an L', a minimum time to interview for sample inclusion in roving surveys.

- **Future research.** Consider the use of an overall catch rate λ, so that completed episode catch C_i^* can be estimated as λL_i^* where L_i^* is the estimate provided by the i^{th} fisher. Estimates and associated variance estimates would then be developed conditional on λ. The parameter λ also requires estimation, as does a component of extra variation due to its estimation which must be added to the variance estimate. Also examine the use of λ with $2L_i$.

- **Future research.** Consider non–constant alternatives for the homogeneous Poisson model for catch rate in roving surveys.

- **Future research.** Investigate optimal allocation of resources. These should address the number and times of overflights as well as the trade-offs between optimizing the ground $CPUE$ survey versus the overflight effort survey(s).

- **Future research.** Investigate the relative merits of the proposed methodology with those based on fully randomized effort and $CPUE$ surveys.

Chapter 4 illustrates how the results of Chapter 3 (*i.e.* methods for restricted randomization in the effort survey) can be applied when the ground survey for estimating $CPUE$ is complex and of the access variety. In particular, formulae were developed for a three–stage design and it was shown how construction of the estimator for variance follows its decomposition. For generality, results were developed assuming unequal probabilities of selection.

Sample inclusion probabilities for unequal initial selection probabilities are, in general, not a straightforward calculation. With equal initial selection probabilities, *i.e.* $1/N$, the sample inclusion probabilities are simply n/N. For unequal, but approximately equal initial selection probabilities, multiplication by n may still produce acceptable results for some range of sampling intensity. Opting for such an estimation is likely in practice and was incorporated in the simulation testing.

- **Future research.** Investigate the effects of using $n\times$ {initial selection probabilities} as the sample inclusion probabilities in unequal probability sampling. Alternative methods for estimating selection probabilities should also be investigated. These should be examined in conjunction with the amount of bias they introduce and an acceptable range of sampling intensity.

- **Future research.** Explore methods to determine optimal allocation of resources with respect to sampling intensities at the various levels of the access structure and number and timings of overflight surveys.

- **Future research.** Determine minimal acceptable values for the number of overflight counts made on the sampled episodes, *i.e.* for $\sum_{i=1}^{nI_{alb}} m_{albi}^{(A)}$. Because this quantity is found in the denominator of $\hat{R}$ and its variance, small values can have powerful leverage.

- **Future research.** Investigate the relative merits of the proposed methodology with those based on fully randomized effort and $CPUE$ access surveys.

- **Future research.** Investigate the use of regression estimators in place of ratio estimators. Note that Table 4.4 suggests that, while catch does appear to increase linearly with the number of times counted in the overflight surveys, the relationship may not be linear through the origin.

Issues related to survey design offer numerous opportunities for future research. An obvious extension to this thesis, but with fewer opportunities for application, is roving surveys in complex designs. Less obvious might be applications to aquatic catch other than fish e.g. shell fish. Here the units observed in overflight may differ from those sampled by the ground survey e.g. parties versus individuals. Perhaps creel methods can have non-aquatic applications. Large scale and small scale surveys have their own particular sets of problems each requiring separate solutions. Also, the recent developments in bus route surveys to estimate catch may have triggered a shift away from the traditional use of aerial surveys.

Creel surveys, with their focus on catch, are only one aspect of the more encompassing activity of angler surveys. Fisheries are multi–million dollar resources involving political, economic, social, and environmental issues and fisheries management has come to involve more than inventory management of fish stocks. Surveys will be required to provide the necessary information about angler needs and habits, and the socioeconomic and demographic trends that affect the fisheries. Work on bringing the collection of this type of data together with the more traditional creel survey task of collecting biological information and catch has only just begun.

Appendix A

Catch Distributions for the Markov Model

This appendix outlines the derivation of a number of results that follow from modelling catch rate as the continuous time Markov process given by

$$\lambda_k = \lambda \left(1 - \frac{k}{N}\right) \quad ; \quad k = 0, 1, ..., N$$

where

λ_k is the occurrence rate of an event after k occurrences,

λ is the initial occurrence rate, and

N is the maximum possible number of occurrences.

The *pmf* for the number of occurrences at time t from such a model is given by the probability that the process takes on the value n at time t, or notationally,

$$P_n(t) = P_n(X(t) = n).$$

To simplify notation when finding an expression for this probability, reparameterize the model as $\lambda_k = \lambda' (N-k)$ and use λ for λ'. Finding the expression for this probability then follows from a solution to

$$\begin{aligned} P_n(t) &= P_n\left(X(t) = n \,|\, X(0) = 0\right) \\ &= \lambda_0 \cdot ... \cdot \lambda_{n-1}\left[B_{0,n}e^{-\lambda_0 t} + B_{1,n}e^{-\lambda_1 t} + ... + B_{n,n}e^{-\lambda_n t}\right] \text{ for } n > 1 \end{aligned}$$

where the components are given as follows:

$$\begin{aligned} \lambda_0 &= N\,\lambda && = N\lambda \\ \lambda_1 &= [N-1]\,\lambda && = N\lambda - \lambda \\ &\vdots && \vdots \\ \lambda_{i-1} &= [N-(i-1)]\,\lambda && = N\lambda - i\lambda + \lambda \\ \lambda_i &= [N-i]\,\lambda && = N\lambda - i\lambda \\ \lambda_{i+1} &= [N-(i+1)]\,\lambda && = N\lambda - i\lambda - \lambda \\ &\vdots && \vdots \\ \lambda_{N-1} &= [N-(N-1)]\,\lambda && = \lambda \\ \lambda_N &= [N-N]\,\lambda && = 0 \end{aligned}$$

and

$$
\begin{aligned}
B_{i,n} &= \frac{1}{(\lambda_0 - \lambda_i)(\lambda_1 - \lambda_i) \cdot \ ... \ \cdot (\lambda_{i-1} - \lambda_i)} \\
&\quad \times \frac{1}{(\lambda_{i+1} - \lambda_i)(\lambda_{i+2} - \lambda_i) \cdot \ ... \ \cdot (\lambda_n - \lambda_i)} \\
&= \frac{1}{[(N\lambda) - (N\lambda - i\lambda)][(N\lambda - \lambda) - (N\lambda - i\lambda)] \cdot \ ... \ \cdot [(N\lambda - i\lambda + \lambda) - (N\lambda - i\lambda)]} \\
&\quad \times \frac{1}{[(N\lambda - i\lambda - \lambda) - (N\lambda - i\lambda)] \cdot \ ... \ \cdot [(N\lambda - n\lambda) - (N\lambda - i\lambda)]} \\
&= \frac{1}{[i\lambda][(i-1)\lambda] \cdot \ ... \ \cdot [\lambda][-\lambda][-2\lambda] \cdot \ ... \ \cdot [-(n-i)\lambda]} \, .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
P_n(t) &= \{\lambda_0 \cdot \lambda_1 \cdot ... \cdot \lambda_{n-2} \cdot \lambda_{n-1}\} \\
&\times \left\{ \frac{e^{-\lambda_0 t}}{[-\lambda][-2\lambda] \cdot \ ... \ \cdot [-n\lambda]} \right. \\
&\quad + \frac{e^{-\lambda_1 t}}{[\lambda][-\lambda][-2\lambda] \cdot \ ... \ \cdot [-(n-1)\lambda]} \\
&\quad + \frac{e^{-\lambda_2 t}}{[2\lambda][\lambda][-\lambda][-2\lambda] \cdot \ ... \ \cdot [-(n-2)\lambda]} \\
&\quad \vdots \\
&\quad + \frac{e^{-\lambda_{n-1} t}}{[(n-1)\lambda][(n-2)\lambda] \cdot \ ... \ \cdot [\lambda][-\lambda]} \\
&\quad + \left. \frac{e^{-\lambda_n t}}{[n\lambda][(n-1)\lambda] \cdot \ ... \ \cdot [2\lambda][\lambda]} \right\} .
\end{aligned}
$$

Expressing λ_i in terms of λ and factoring $\frac{1}{\lambda}$ and $e^{-N\lambda t}$ from each term then gives

$$\begin{aligned}
P_n(t) &= \{N\lambda \cdot [N-1]\lambda \cdot [N-2]\lambda \cdot \ldots \cdot [N-n]\lambda\}\, e^{-N\lambda t} \\
&\times \left\{ \frac{1}{\lambda} \frac{1}{[-1][-2] \cdot \ldots \cdot [-n]} \right. \\
&+ \frac{1}{\lambda} \frac{e^{\lambda t}}{[1][-1][-2] \cdot \ldots \cdot [-(n-1)]} \\
&+ \frac{1}{\lambda} \frac{e^{2\lambda t}}{[2][1][-1][-2] \cdot \ldots \cdot [-(n-2)]} \\
&\vdots \\
&+ \frac{1}{\lambda} \frac{e^{(n-1)\lambda t}}{[n-1][n-2] \cdot \ldots \cdot [1][-1]} \\
&+ \left. \frac{1}{\lambda} \frac{e^{n\lambda t}}{[n][n-1] \cdot \ldots \cdot [2][1]} \right\}.
\end{aligned}$$

Finally, cancelling the free λ's and extracting $\frac{1}{n!}$ from each term and combining with the first factor, $P_n(t)$ can be written as

$$\begin{aligned}
P_n(t) &= \frac{N!}{(N-n)!\, n!} e^{-N\lambda t} \\
&\times \left\{ \frac{e^{\lambda t[n-n]}}{0!\, n!} \right. \\
&\pm \frac{e^{\lambda t[n-(n-1)]}}{1!\, (n-1)!} \\
&\pm \frac{e^{\lambda t[n-(n-2)]}}{2!\, (n-2)!} \\
&\vdots \\
&\pm \frac{e^{\lambda t[n-1]}}{(n-1)!\, 1!} \\
&\pm \left. \frac{e^{\lambda t[n-0]}}{n!\, 0!} \right\}
\end{aligned}$$

$$= \frac{N\,!}{(N-n)\,!\,n\,!} e^{-N\lambda t} \sum_{i=0}^{n} \begin{pmatrix} n \\ i \end{pmatrix} (-1)^i \left(e^{\lambda t}\right)^{n-i}$$

which, by the binomial theorem, can be written as

$$\begin{aligned} P_n(t) &= \frac{N\,!}{(N-n)\,!\,n\,!} e^{-N\lambda t} \left(e^{\lambda t}-1\right)^n \\ &= \begin{pmatrix} N \\ n \end{pmatrix} \left(e^{-\lambda t}\right)^N \left(e^{\lambda t}-1\right)^n \frac{\left(e^{-\lambda t}\right)^n}{\left(e^{-\lambda t}\right)^n} \\ &= \begin{pmatrix} N \\ n \end{pmatrix} \left(e^{-\lambda t}\right)^N \left(e^{\lambda t}-1\right)^{-n} \left[\left(e^{\lambda t}-1\right)\left(e^{-\lambda t}\right)\right]^n \\ &= \begin{pmatrix} N \\ n \end{pmatrix} \left(e^{-\lambda t}\right)^{N-n} \left(1-e^{-\lambda t}\right)^n . \end{aligned}$$

Thus, expressed in the original parameterization,

$$P_n(t) = \frac{N\,!}{(N-n)\,!\,n\,!} \left(1-e^{-\frac{\lambda t}{N}}\right)^n \left(e^{-\frac{\lambda t}{N}}\right)^{N-n}$$

which is seen to be binomial in $\left(1-e^{-\frac{\lambda t}{N}}\right)$ and N.

By modelling the catch rate as this continuous time Markov process and using the original parameterization, it has been shown that for any fishing episode, the *pmf* for $C(l)$, the catch at a given time l, is binomial *i.e.*

$$P(C(l)) = \begin{pmatrix} C_0 \\ C(l) \end{pmatrix} \left(1-e^{-\frac{\lambda l}{C_0}}\right)^{C(l)} \left(e^{-\frac{\lambda l}{C_0}}\right)^{C_0-C(l)} .$$

It then follows immediately that for catch at times L_j and L_j^*,

$$P(C_j|\,L_j,\delta=1) = \begin{pmatrix} C_0 \\ C_j \end{pmatrix} \left(1-e^{-\frac{\lambda L_j}{C_0}}\right)^{C_j} \left(e^{-\frac{\lambda L_j}{C_0}}\right)^{C_0-C_j}$$

and

$$P\left(C_j^* \middle| L_j^*, \delta = 1\right) = \begin{pmatrix} C_0 \\ C_j^* \end{pmatrix} \left(1 - e^{-\frac{\lambda L_j^*}{C_0}}\right)^{C_j^*} \left(e^{-\frac{\lambda L_j^*}{C_0}}\right)^{C_0 - C_j^*}.$$

To find the probability that catch equals C_j^* at time L_j^* given that a catch of $C_j < C_j^*$ has been realized at time $L_j < L_j^*$ requires an adjustment to account for the non-randomness in the time and catch that have already occurred. The *pmf* at time L_j^* remains binomial, but only after the first C_j catches and after time L_j. Effectively, the randomness is $C_0 - C_j$ possible values of catch for the variable $C_j^* - C_j$. The initial catch rate, λ, is accordingly modified as $\lambda(L_j^* - L_j)$ rather than remaining as the λL_j^* that would have been used in the unconditioned case. This results in

$$P\left(C_j^* \middle| C_j, L_j, L_j^*, \delta = 1\right) = \begin{pmatrix} C_0 - C_j \\ C_j^* - C_j \end{pmatrix} \left(1 - e^{-\frac{\lambda\left(L_j^* - L_j\right)}{C_0}}\right)^{C_j^* - C_j} \left(e^{-\frac{\lambda\left(L_j^* - L_j\right)}{C_0}}\right)^{C_0 - C_j^*}.$$

This result can also be derived directly by using a continuous time Markov process where again $\lambda_k = \lambda\left(1 - \frac{k}{C_0}\right)$ but now for $k = C_j,\, C_j + 1,\, \ldots, C_0$.

Finally, observe that

$$P\left(C_j \middle| L_j, C_j^*, L_j^*, \delta = 1\right) = \frac{P\left(C_j^* \middle| C_j, L_j, L_j^*, \delta = 1\right) P\left(C_j \middle| L_j, \delta = 1\right)}{P\left(C_j^* \middle| L_j^*, \delta = 1\right)}.$$

Using the previous results and simplifying the components,

$$\frac{\begin{pmatrix} C_0 - C_j \\ C_j^* - C_j \end{pmatrix} \begin{pmatrix} C_0 \\ C_j \end{pmatrix}}{\begin{pmatrix} C_0 \\ C_j^* \end{pmatrix}} = \begin{pmatrix} C_j^* \\ C_j \end{pmatrix}$$

and, with the aid of a symbolic computational package such as MAPLE,

$$\frac{\left(1-e^{-\frac{\lambda\left(L_j^*-L_j\right)}{C_0}}\right)^{C_j^*-C_j}\left(e^{-\frac{\lambda\left(L_j^*-L_j\right)}{C_0}}\right)^{C_0-C_j^*}\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)^{C_j}\left(e^{-\frac{\lambda L_j}{C_0}}\right)^{C_0-C_j}}{\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)^{C_j^*-C_j}\left(e^{-\frac{\lambda L_j^*}{C_0}}\right)^{C_0-C_j^*}}$$

$$= \left(\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)^{C_j}\left(\frac{e^{-\frac{\lambda L_j}{C_0}}-e^{-\frac{\lambda L_j^*}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)^{C_j^*-C_j}$$

$$= \left(\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)^{C_j}\left(1-\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)^{C_j^*-C_j}.$$

Combining,

$$P\left(C_j\,|\,L_j,C_j^*,L_j^*,\delta=1\right)=\begin{pmatrix} C_j^* \\ C_j \end{pmatrix}\left(\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)^{C_j}\left(1-\frac{1-e^{-\frac{\lambda L_j}{C_0}}}{1-e^{-\frac{\lambda L_j^*}{C_0}}}\right)$$

which is seen to be binomial in C_j^* and $\left(1-e^{-\frac{\lambda L_j}{C_0}}\right)\Big/\left(1-e^{-\frac{\lambda L_j^*}{C_0}}\right)$.

Appendix B

Results for the Markov Model as $C_0 \rightarrow \infty$

This appendix outlines how, when $N \rightarrow \infty$, the continuous time Markov process considered in appendix A approaches a homogeneous Poisson process and also develops some of the basic results that follow as an immediate consequence of using these processes to model catch rate.

In appendix A was shown that, under the original parameterization,

$$P_n(t) = \frac{N!}{(N-n)!\,n!}\left(1 - e^{-\frac{\lambda t}{N}}\right)^n \left(e^{-\frac{\lambda t}{N}}\right)^{N-n}.$$

Let $\theta = E[n] = N\left(1 - e^{-\frac{\lambda t}{N}}\right)$. It is well known that for a binomial distribution, the limiting distribution when $N \rightarrow \infty$ in such a way that θ remains constant is a Poisson distribution in θ. To show that $\theta = \lambda t$ when $N \rightarrow \infty$, write

$$\begin{aligned} \theta &= N\left(e^{-\frac{\lambda t}{N}}\right) \\ &= \frac{1 - e^{-\frac{\lambda t}{N}}}{N^{-1}}. \end{aligned}$$

Applying l'Hopital's rule gives

$$
\begin{aligned}
\theta &= \lim_{N\rightarrow\infty} N\left(1-e^{-\frac{\lambda t}{N}}\right) \\
&= \lim_{N\rightarrow\infty} \frac{\frac{d}{dN}\left(1-e^{-\frac{\lambda t}{N}}\right)}{\frac{d}{dN}N^{-1}} \\
&= \frac{\lim\limits_{N\rightarrow\infty} -\lambda t e^{-\frac{\lambda t}{N}} N^{-2}}{\lim\limits_{N\rightarrow\infty} -N^{-2}} \\
&= \lambda t
\end{aligned}
$$

where λ is interpreted as a rate or intensity.

From the above it can be seen that, in the limit as $C_0 \rightarrow \infty$, modelling the catch rate as a continuous time Markov process results in a homogeneous Poisson process with parameter λ. It is also clear that the initial catch rate used in the Markov process is the same rate used in the Poisson process where the *pmf* of catch at time l is now given by

$$
P(C(l)) = \frac{(\lambda l)^{C(l)} e^{-\lambda l}}{C(l)\,!}.
$$

It immediately follows that

$$
\begin{aligned}
P(C_j | L_j, \delta = 1) &= \frac{(\lambda L_j)^{C_j} e^{-\lambda L_j}}{C_j\,!} \\
P\left(C_j^* \middle| L_j^*, \delta = 1\right) &= \frac{\left(\lambda L_j^*\right)^{C_j^*} e^{-\lambda L_j^*}}{C_j^*\,!}
\end{aligned}
$$

and

$$
P\left(C_j^* \middle| C_j, L_j, L_j^*, \delta = 1\right) = \frac{\left[\lambda\left(L_j^* - L_j\right)\right]^{C_j^* - C_j} e^{-\lambda\left(L_j^* - L_j\right)}}{\left(C_j^* - C_j\right)!}.
$$

The above can be used to derive

$$
\begin{aligned}
P\big(C_j |\, L_j, C_j^*, L_j^*, \delta = 1\big) &= \frac{P\Big(C_j \cap C_j^*\Big)}{P\Big(C_j^*\Big)} \\
&= \frac{P\Big(C_j^* \Big| \, C_j, L_j, L_j^*, \delta = 1\Big)\, P(C_j |\, L_j, \delta = 1)}{P\Big(C_j^* \Big| \, L_j^*, \delta = 1\Big)}.
\end{aligned}
$$

Substituting,

$$
\begin{aligned}
P\big(C_j |\, L_j, C_j^*, L_j^*, \delta = 1\big) &= \frac{\frac{\left[\lambda\left(L_j^* - L_j\right)\right]^{C_j^* - C_j} e^{-\lambda\left(L_j^* - L_j\right)}}{\left(C_j^* - C_j\right)!} \cdot \frac{(\lambda L_j)^{C_j} e^{-\lambda L_j}}{C_j\,!}}{\frac{\left(\lambda L_j^*\right)^{C_j^*} e^{-\lambda L_j^*}}{C_j^*\,!}} \\
&= \frac{C_j^*\,!}{\Big(C_j^* - C_j\Big)\,!C_j\,!} \cdot \frac{\Big(L_j^* - L_j\Big)^{C_j^* - C_j} L_j^{C_j}}{L_j^{*\,C_j^*}} \\
&= \begin{pmatrix} C_j^* \\ \\ C_j \end{pmatrix} \left(\frac{L_j}{L_j^*}\right)^{C_j} \left(1 - \frac{L_j}{L_j^*}\right)^{C_j^* - C_j}
\end{aligned}
$$

i.e. binomial in $\left(\frac{L_j}{L_j^*}\right)$ and C_j^*.

Appendix C

Approximating the True Inclusion Probabilities

This appendix outlines a defensible method of approximating sample inclusion probabilities using unequal initial selection probabilities. Results are developed for τ_l and π_{lb}, *i.e.* inclusion probabilities for *Launch* and *Block* respectively. Because *Episodes* are selected as a simple random sample with equal initial selection probability, their inclusion probabilities have been correctly calculated as initial selection probability $\times$ sample size $= n/N$.

For any particular sampling unit, calculation of its inclusion probability in a sample of size n with unequal initial selection probabilities p_j for $j = 1, \ldots, N$ sampling units involves summing all $\begin{pmatrix} N-1 \\ n-1 \end{pmatrix}$ possible samples that contain that unit. The probability of each of these samples is the sum of the probabilities of each of their $n!$ permutations. To illustrate the calculation required for each permutation, consider a sample of size $n = 3$ where units

1, 3, and 7 with respective initial probabilities p_1, p_3, and p_7 have been selected. Then

$$p_{\langle 1,3,7\rangle} = p_1 \cdot \frac{p_3}{1-p_1} \cdot \frac{p_7}{1-(p_1+p_3)} \cdot$$

A method of approximating the inclusion probability for any unit is then to randomly generate m such permutations and average the resulting permutation probabilities. This average is then expanded by $\binom{N-1}{n-1} n!$

Table C.1 demonstrates the error in using $n\times$ initial selection probability for *Block*. Note that for sample size $n = 3$, $\pi_2 > 1$. With $N = 4$, it is possible to perform the 192

Table C.1: Block inclusion probabilities: Estimates (initial selection probabilities × sample size).

Sample Size	Unit Inclusion Probabilities π_1	π_2	π_3	π_4	$\sum_{j=1}^{N} \pi_j$
1	0.10	0.40	0.25	0.25	1
2	0.20	0.80	0.50	0.50	2
3	0.30	1.20	0.75	0.75	3
4	0.40	1.60	1.00	1.00	4

calculations required to construct Table C.2, the table of correct inclusion probabilities for all possible sample sizes. Tables C.3, C.4, and C.5 use the proposed method to approximate the inclusion probabilities and show the usefulness of increasing the number of random permutations.

Table C.2: Block inclusion probabilities: Exact values.

Sample Size	Unit Inclusion Probabilities π_1	π_2	π_3	π_4	$\sum_{j=1}^{N} \pi_j$
1	0.1000000	0.4000000	0.2500000	0.2500000	1
2	0.2333333	0.7111111	0.5277778	0.5277778	2
3	0.4380952	0.9196581	0.8211233	0.8211233	3
4	1.0000000	1.0000000	1.0000000	1.0000000	4

Table C.3: Block inclusion probabilities: Approximation (based on samples of size m=30).

Sample Size	Unit Inclusion Probabilities π_1	π_2	π_3	π_4	$\sum_{j=1}^{N} \pi_j$
1	0.1000000	0.4000000	0.2500000	0.2500000	1.000000
2	0.2433333	0.6755556	0.4633333	0.6111111	1.993333
3	0.4638462	0.8777289	0.9997436	0.7294505	3.070769
4	0.9568742	1.0334554	0.9418803	1.0832723	4.015482

Table C.4: Block inclusion probabilities: Approximation (based on samples of size m=100).

Sample Size	Unit Inclusion Probabilities				
	π_1	π_2	π_3	π_4	$\sum_{j=1}^{N} \pi_j$
1	0.1000000	0.4000000	0.2500000	0.2500000	1.000000
2	0.2346667	0.6966667	0.5190000	0.5270000	1.977333
3	0.4531648	0.9258901	0.7315495	0.7970989	2.907703
4	0.9039707	0.9798535	0.8469158	0.9770989	3.707839

Table C.5: Block inclusion probabilities: Approximation (based on samples of size m=500).

Sample Size	Unit Inclusion Probabilities				
	π_1	π_2	π_3	π_4	$\sum_{j=1}^{N} \pi_j$
1	0.1000000	0.4000000	0.2500000	0.2500000	1.000000
2	0.2315333	0.7284000	0.5114667	0.4966667	1.968067
3	0.4565670	0.8602286	0.8088352	0.8051165	2.930747
4	0.9957158	0.9985729	0.9892806	0.9818784	3.965448

Inclusion probabilities for *Launch* using the simplistic approach are given in Tables C.6 to C.9. Results using the proposed technique are given in Tables C.10 to C.13.

Table C.6: Launch inclusion probabilities Part I: Estimates (initial selection probabilities × sample size).

Sample Size	Unit Inclusion Probabilities τ_1	τ_2	τ_3	τ_4	τ_5
1	0.0675	0.0625	0.0550	0.0500	0.0500
2	0.1350	0.1250	0.1100	0.1000	0.1000
3	0.2025	0.1875	0.1650	0.1500	0.1500
4	0.2700	0.2500	0.2200	0.2000	0.2000
5	0.3375	0.3125	0.2750	0.2500	0.2500
6	0.4050	0.3750	0.3300	0.3000	0.3000
7	0.4725	0.4375	0.3850	0.3500	0.3500
8	0.5400	0.5000	0.4400	0.4000	0.4000
9	0.6075	0.5625	0.4950	0.4500	0.4500
10	0.6750	0.6250	0.5500	0.5000	0.5000
11	0.7425	0.6875	0.6050	0.5500	0.5500
12	0.8100	0.7500	0.6600	0.6000	0.6000
13	0.8775	0.8125	0.7150	0.6500	0.6500
14	0.9450	0.8750	0.7700	0.7000	0.7000
15	1.0125	0.9375	0.8250	0.7500	0.7500
16	1.0800	1.0000	0.8800	0.8000	0.8000
17	1.1475	1.0625	0.9350	0.8500	0.8500
18	1.2150	1.1250	0.9900	0.9000	0.9000
19	1.2825	1.1875	1.0450	0.9500	0.9500
20	1.3500	1.2500	1.1000	1.0000	1.0000

Table C.7: Launch inclusion probabilities Part II: Estimates (initial selection probabilities $\times$ sample size).

Sample Size	Unit Inclusion Probabilities				
	τ_6	τ_7	τ_8	τ_9	τ_{10}
1	0.0450	0.0450	0.0450	0.0400	0.0400
2	0.0900	0.0900	0.0900	0.0800	0.0800
3	0.1350	0.1350	0.1350	0.1200	0.1200
4	0.1800	0.1800	0.1800	0.1600	0.1600
5	0.2250	0.2250	0.2250	0.2000	0.2000
6	0.2700	0.2700	0.2700	0.2400	0.2400
7	0.3150	0.3150	0.3150	0.2800	0.2800
8	0.3600	0.3600	0.3600	0.3200	0.3200
9	0.4050	0.4050	0.4050	0.3600	0.3600
10	0.4500	0.4500	0.4500	0.4000	0.4000
11	0.4950	0.4950	0.4950	0.4400	0.4400
12	0.5400	0.5400	0.5400	0.4800	0.4800
13	0.5850	0.5850	0.5850	0.5200	0.5200
14	0.6300	0.6300	0.6300	0.5600	0.5600
15	0.6750	0.6750	0.6750	0.6000	0.6000
16	0.7200	0.7200	0.7200	0.6400	0.6400
17	0.7650	0.7650	0.7650	0.6800	0.6800
18	0.8100	0.8100	0.8100	0.7200	0.7200
19	0.8550	0.8550	0.8550	0.7600	0.7600
20	0.9000	0.9000	0.9000	0.8000	0.8000

Table C.8: Launch inclusion probabilities Part III: Estimates (initial selection probabilities × sample size).

Sample Size	Unit Inclusion Probabilities τ_{11}	τ_{12}	τ_{13}	τ_{14}	τ_{15}
1	0.0675	0.0625	0.0550	0.0500	0.0500
2	0.1350	0.1250	0.1100	0.1000	0.1000
3	0.2025	0.1875	0.1650	0.1500	0.1500
4	0.2700	0.2500	0.2200	0.2000	0.2000
5	0.3375	0.3125	0.2750	0.2500	0.2500
6	0.4050	0.3750	0.3300	0.3000	0.3000
7	0.4725	0.4375	0.3850	0.3500	0.3500
8	0.5400	0.5000	0.4400	0.4000	0.4000
9	0.6075	0.5625	0.4950	0.4500	0.4500
10	0.6750	0.6250	0.5500	0.5000	0.5000
11	0.7425	0.6875	0.6050	0.5500	0.5500
12	0.8100	0.7500	0.6600	0.6000	0.6000
13	0.8775	0.8125	0.7150	0.6500	0.6500
14	0.9450	0.8750	0.7700	0.7000	0.7000
15	1.0125	0.9375	0.8250	0.7500	0.7500
16	1.0800	1.0000	0.8800	0.8000	0.8000
17	1.1475	1.0625	0.9350	0.8500	0.8500
18	1.2150	1.1250	0.9900	0.9000	0.9000
19	1.2825	1.1875	1.0450	0.9500	0.9500
20	1.3500	1.2500	1.1000	1.0000	1.0000

Table C.9: Launch inclusion probabilities Part IV: Estimates (initial selection probabilities × sample size).

Sample Size	Unit Inclusion Probabilities τ_{16}	τ_{17}	τ_{18}	τ_{19}	τ_{20}	$\sum_j^N \tau_{j=1}$
1	0.0450	0.0450	0.0450	0.0400	0.0400	1.0
2	0.0900	0.0900	0.0900	0.0800	0.0800	2.0
3	0.1350	0.1350	0.1350	0.1200	0.1200	3.0
4	0.1800	0.1800	0.1800	0.1600	0.1600	4.0
5	0.2250	0.2250	0.2250	0.2000	0.2000	5.0
6	0.2700	0.2700	0.2700	0.2400	0.2400	6.0
7	0.3150	0.3150	0.3150	0.2800	0.2800	7.0
8	0.3600	0.3600	0.3600	0.3200	0.3200	8.0
9	0.4050	0.4050	0.4050	0.3600	0.3600	9.0
10	0.4500	0.4500	0.4500	0.4000	0.4000	10.0
11	0.4950	0.4950	0.4950	0.4400	0.4400	11.0
12	0.5400	0.5400	0.5400	0.4800	0.4800	12.0
13	0.5850	0.5850	0.5850	0.5200	0.5200	13.0
14	0.6300	0.6300	0.6300	0.5600	0.5600	14.0
15	0.6750	0.6750	0.6750	0.6000	0.6000	15.0
16	0.7200	0.7200	0.7200	0.6400	0.6400	16.0
17	0.7650	0.7650	0.7650	0.6800	0.6800	17.0
18	0.8100	0.8100	0.8100	0.7200	0.7200	18.0
19	0.8550	0.8550	0.8550	0.7600	0.7600	19.0
20	0.9000	0.9000	0.9000	0.8000	0.8000	20.0

Table C.10: Launch inclusion probabilities Part I: Approximation (based on samples of size m=500).

Sample Size	Unit Inclusion Probabilities τ_1	τ_2	τ_3	τ_4	τ_5
1	0.0675000	0.0625000	0.0550000	0.05000000	0.05000000
2	0.1346493	0.1245176	0.1108029	0.09924387	0.09954335
3	0.1959769	0.1867371	0.1646648	0.14953696	0.14752081
4	0.2629198	0.2477580	0.2195224	0.20332355	0.20578856
5	0.3281954	0.3029183	0.2758950	0.25374610	0.24875815
6	0.3774486	0.3585680	0.3305658	0.30049576	0.30450734
7	0.4469081	0.4242398	0.3728706	0.36017878	0.35386192
8	0.4949651	0.4821595	0.4243410	0.40456261	0.40499865
9	0.5374499	0.5365399	0.4833915	0.45603660	0.46324712
10	0.6302676	0.5858602	0.5292155	0.49630734	0.50641742
11	0.6743451	0.6210828	0.5774209	0.57557255	0.57415287
12	0.7170988	0.6994098	0.6255683	0.58059688	0.61646668
13	0.7532376	0.7117767	0.6592466	0.65449762	0.65597756
14	0.8033414	0.7922664	0.7235615	0.71114748	0.68715393
15	0.8419091	0.8264320	0.7688638	0.71986671	0.70427442
16	0.9003853	0.8477645	0.8214963	0.78716461	0.78504976
17	0.8860276	0.9426887	0.9142763	0.84977488	0.86629850
18	0.9691974	0.9318554	0.9581405	0.91463731	0.90258405
19	0.9719204	0.9798450	0.9606950	0.94445742	0.95093686
20	1.0176409	1.0527602	1.0106806	1.00291996	0.94037269

Table C.11: Launch inclusion probabilities Part II: Approximation (based on samples of size m=500).

Sample Size	Unit Inclusion Probabilities τ_6	τ_7	τ_8	τ_9	τ_{10}
1	0.0450000	0.0450000	0.04500000	0.04000000	0.04000000
2	0.0896518	0.0901465	0.08984819	0.08034797	0.07962864
3	0.1361961	0.1359559	0.13730069	0.12065753	0.12206792
4	0.1788189	0.1816934	0.18005582	0.15578200	0.16324841
5	0.2292446	0.2276714	0.22277442	0.20875861	0.20422774
6	0.2841982	0.2713462	0.27923416	0.24986449	0.24535700
7	0.3261092	0.3240168	0.31479728	0.28735546	0.30211323
8	0.3718857	0.3756389	0.36357195	0.33879679	0.32501651
9	0.4181641	0.4215374	0.42099432	0.37718203	0.38261930
10	0.4572962	0.4702630	0.45951477	0.42274847	0.42771507
11	0.5263391	0.5164458	0.53805328	0.46323904	0.46623802
12	0.5441679	0.5418895	0.56574482	0.53326098	0.52796854
13	0.6300473	0.6207567	0.62245582	0.58285258	0.61280555
14	0.6629983	0.6422980	0.67195800	0.62779072	0.63428308
15	0.7566360	0.7622972	0.70130597	0.63953448	0.66705342
16	0.7546106	0.7561330	0.78718814	0.75343682	0.72780452
17	0.8477016	0.8404843	0.82744326	0.81279276	0.80036706
18	0.8651557	0.8697812	0.91472182	0.84988022	0.91895031
19	0.8949400	0.9900792	0.87945584	0.92385089	0.94470532
20	0.9814962	1.0215923	0.99404229	0.97488078	1.01319523

Table C.12: Launch inclusion probabilities Part III: Approximation (based on samples of size m=500).

Sample Size	Unit Inclusion Probabilities τ_{11}	τ_{12}	τ_{13}	τ_{14}	τ_{15}
1	0.0675000	0.0625000	0.0550000	0.05000000	0.0500000
2	0.1329987	0.1242675	0.1105155	0.09982692	0.1005485
3	0.1993074	0.1879401	0.1656474	0.14953166	0.1501808
4	0.2629347	0.2529937	0.2236341	0.20213867	0.2031855
5	0.3205978	0.2888352	0.2693552	0.24436359	0.2508030
6	0.3832685	0.3559710	0.3305106	0.30571411	0.3062322
7	0.4414147	0.4312900	0.3812751	0.35168300	0.3531653
8	0.5027901	0.4746572	0.4492888	0.41322801	0.4147091
9	0.5442168	0.5364258	0.4991249	0.45025113	0.4233271
10	0.6189088	0.5847700	0.5435932	0.51515792	0.5080231
11	0.6417237	0.6556217	0.5970652	0.55032574	0.5463011
12	0.7209494	0.7077335	0.6423451	0.59841360	0.6036795
13	0.7629902	0.7299963	0.6955751	0.68777631	0.6365514
14	0.7966527	0.7363406	0.7674146	0.68362148	0.7171547
15	0.8708451	0.8576469	0.7759921	0.74617133	0.7486741
16	0.8988033	0.9028806	0.8152420	0.79337823	0.8146582
17	0.9172304	0.9185908	0.8714121	0.86684129	0.8489400
18	0.9497322	0.9196450	0.9332090	0.93641358	0.9009972
19	1.0012602	1.0359072	0.9085764	0.94310995	0.9612150
20	0.9852716	1.0173743	0.9865965	0.98229742	1.0114769

Table C.13: Launch inclusion probabilities Part IV: Approximation (based on samples of size m=500).

Sample Size	Unit Inclusion Probabilities					
	τ_{16}	τ_{17}	τ_{18}	τ_{19}	τ_{20}	$\sum_{j}^{N} \tau_{j=1}$
1	0.04500000	0.04500000	0.04500000	0.04000000	0.04000000	1.0000
2	0.09090157	0.09102988	0.09012941	0.08037868	0.08014218	1.9992
3	0.13695313	0.13484320	0.13835225	0.12345900	0.12208495	3.0049
4	0.18116614	0.18377876	0.18433681	0.16534431	0.16191717	4.0203
5	0.23164525	0.23825735	0.22926204	0.20369743	0.20328612	4.9823
6	0.27855694	0.27047779	0.27527685	0.25064024	0.24721538	6.0054
7	0.31424076	0.31900970	0.31484456	0.29386346	0.29306868	7.0063
8	0.36737707	0.36937764	0.36852207	0.34281330	0.32719904	8.0159
9	0.41803170	0.41488454	0.41028914	0.38234167	0.36554519	8.9416
10	0.48303003	0.45977287	0.46296382	0.42968171	0.42158397	10.0131
11	0.50797904	0.52587655	0.50280402	0.48547239	0.47359236	11.0197
12	0.58926059	0.57894483	0.58301270	0.49685986	0.50528354	11.9787
13	0.64112348	0.60683516	0.62402958	0.56787504	0.58018614	13.0366
14	0.68837489	0.67111492	0.67728493	0.64718778	0.62312858	13.9651
15	0.72596764	0.71171117	0.72996601	0.66488326	0.69167439	14.9117
16	0.76260520	0.80841871	0.76243165	0.73650549	0.73000351	15.9460
17	0.80554324	0.77724962	0.79269985	0.83473967	0.77721580	16.9983
18	0.91131142	0.91465121	0.86978748	0.86249195	0.84173711	18.1349
19	0.88679060	0.86514645	0.93206295	0.88093604	0.94168881	18.7976
20	0.95082976	1.00817357	1.01828613	1.02841223	0.99777377	19.9961

Simulations were rerun using results of the proposed methodology. Table C.14 is a revised Table 4.3 while Table C.15 gives revised results for Table 4.4.

Table C.14: Simulated data: Performance of the $\hat{R}$ estimator for differing sampling rates on Launch Site, Interview Blocks and fishing Episodes (L=25%; H=75%). Catch rates vary by Episode in Scenario I, vary by Block in Scenario II, vary by Launch in Scenario III and vary by Block and Launch in Scenario IV. (500 replicated simulations.)

Sampling Percentages nL	nB	nI		Scenario I	Scenario II	Scenario III	Scenario IV
			R (population)	5.676	6.328	5.834	9.608
L	L	L	$Mean_s(\hat{R})$	5.678	6.273	5.856	9.451
			Bias (%)	+0.0	-0.9	+0.4	-1.7
			c.v. (%)	7.8	9.0	17.8	11.2
L	L	H	$Mean_s(\hat{R})$	5.682	6.271	5.852	9.452
			Bias (%)	+0.1	-0.9	+0.3	-1.7
			c.v. (%)	5.7	7.6	16.4	10.4
L	H	L	$Mean_s(\hat{R})$	5.661	6.265	5.809	9.513
			Bias (%)	-0.3	-1.0	-0.4	-1.0
			c.v. (%)	5.7	6.3	17.1	9.4
L	H	H	$Mean_s(\hat{R})$	5.648	6.261	5.792	9.484
			Bias (%)	-0.5	-1.1	-0.7	-1.3
			c.v. (%)	3.7	4.7	16.2	8.6
H	L	L	$Mean_s(\hat{R})$	5.644	6.298	5.704	9.402
			Bias (%)	-0.6	-0.5	-2.3	-2.2
			c.v. (%)	5.6	6.7	10.2	7.3
H	L	H	$Mean_s(\hat{R})$	5.654	6.268	5.691	9.444
			Bias (%)	-0.4	-1.0	-2.5	-1.7
			c.v. (%)	3.2	4.2	6.2	4.3
H	H	L	$Mean_s(\hat{R})$	5.650	6.283	5.708	9.448
			Bias (%)	-0.5	-0.7	-2.2	-1.7
			c.v. (%)	3.3	3.3	6.2	4.1
H	H	H	$Mean_s(\hat{R})$	5.645	6.281	5.710	9.448
			Bias (%)	-0.6	-0.8	-2.2	-1.7
			c.v. (%)	1.9	2.5	5.6	3.6

Table C.15: Simulated data: Comparison of variance and variance estimates of $\hat{R}$ for differing sampling rates on Launch Site, Interview Blocks and fishing Episodes (L=25%; H=75%). $V_s(\hat{R})$ denotes the variance of the simulated $\hat{R}$ values, $\widehat{V}(\hat{R})$ the mean of formula estimates and $\widehat{V}_j(\hat{R})$ the mean of jackknife estimates (without a finite population correction factor). Catch rates vary by Episode in Scenario I, vary by Block in Scenario II, vary by Launch in Scenario III and vary by Block and Launch in Scenario IV. (500 replicated simulations.)

Sampling Percentages nL	nB	nI		Scenario I	Scenario II	Scenario III	Scenario IV
L	L	L	$V_s(\hat{R})$	0.198	0.317	1.082	1.117
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	2.22	2.24	1.12	1.54
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.21	1.10	1.19	1.23
L	L	H	$V_s(\hat{R})$	0.105	0.224	0.918	0.971
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.86	2.09	1.04	1.27
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.14	1.07	1.25	1.18
L	H	L	$V_s(\hat{R})$	0.105	0.156	0.984	0.796
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	2.04	2.03	1.01	1.45
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.07	1.11	1.18	1.41
L	H	H	$V_s(\hat{R})$	0.042	0.085	0.878	0.671
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.83	2.04	0.98	1.29
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.12	1.12	1.25	1.44
H	L	L	$V_s(\hat{R})$	0.101	0.176	0.341	0.471
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.78	1.76	1.32	1.45
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.03	0.96	1.87	2.44
H	L	H	$V_s(\hat{R})$	0.033	0.069	0.122	0.161
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.36	1.68	1.14	1.42
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.12	1.10	3.24	2.39
H	H	L	$V_s(\hat{R})$	0.035	0.042	0.125	0.148
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.46	1.83	1.13	1.34
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.06	1.26	3.10	2.46
H	H	H	$V_s(\hat{R})$	0.011	0.024	0.103	0.114
			$\widehat{V}(\hat{R})/V_s(\hat{R})$	1.55	1.75	1.05	1.15
			$\widehat{V}_J(\hat{R})/V_s(\hat{R})$	1.36	1.25	3.53	2.80

Bibliography

Brewer, K. R. W. and Hanif, M. (1983). *Sampling with unequal probabilities*. New York: Springer-Verlag.

Brown, T. L. (1991). Use and abuse of mail surveys in fisheries management. *American Fisheries Society Symposium* **12**, 255-261.

Chen, S. X. and Woolcock, J. L. (1999). A condition for designing bus-route type access site surveys to estimate recreational fishing effort. *Biometrics* **55**, 799-804.

Crone, P. R. and Malvestuto, S. P. (1991). A comparison of five estimators of fishing success from creel survey data on three Alabama reservoirs. *American Fisheries Society Symposium* **12**, 61-66.

Cochran, W. G. (1977). *Sampling techniques*, 3rd edition. New York: John Wiley.

Dahiya, R. C. (1981). An improved method of estimating an integer-parameter by maximum likelihood. *The American Statistician* **35**, 34-37.

Dent, R. J. and Wagner, B. (1991). Changes in sampling design to reduce variability in selected estimates from a roving creel survey conducted on Pomme de Terre Lake. *American Fisheries Society Symposium* **12**, 88-96.

English, K. K., Shardlow, T. F. and Webb, T. M. (1986). Assessment of Strait of Georgia sport fishing statistics, sport fishing regulations and trends in chinook catch using creel survey data. Can. Tech. Rep. Fish. Aquat. Sci.; **1375** : 54 p.

Eschmeyer, R. W. (1942). The catch, abundance and migration of game fishes in Norris Reservoir, Tenn., 1940. *Proc. Tenn. Acad. Sci.* **17(1)** : 90-115.

Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association* **55**, 708-713.

Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *J. R. Statist. Soc* B **46**, 1949-192.

Gunderson, D. R. (1993). *Surveys of fisheries resources.* New York: John Wiley.

Guthrie, D., Hoenig, J. M., Holliday, M., Jones, C. M., Mills, M. J., Moberly, S. A., Pollock, K. H., and Talhelm, D. R., editors (1991). Creel and angler surveys in fisheries management. *American Fisheries Society Symposium* **12**.

Hayne, D. W. (1991). The access point creel survey: Procedures and comparisons with the roving clerk creel survey. *American Fisheries Society Symposium* **12**, 123-138.

Hardie, D. C., Nagtegaal, D. A. and Nagy, L. (1999). Strait of Georgia sport fishery creel survey statistics for salmon and groundfish, 1998. Can. Manuscr. Rep. Fish. Aquat. Sci.; (Draft).

Hoenig, J. M., Jones, C. M., Pollock, K. H., Robson, D. S. and Wade, D. L. (1997). Calculation of catch rate and total catch in roving surveys of anglers. *Biometrics* **53**, 306-317.

Hoenig, J. M., Robson, D. S., Jones, C. M. and Pollock, K. H. (1993). Scheduling counts in the instantaneous and progressive count methods for estimating sport-fishing effort. *North American Journal of Fisheries Management* **13**, 723-736.

Malvestuto, S. P., Davies, W. D. and Shelton, W. L. (1978). An evaluation of the roving creel survey with nonuniform probability sampling. *Trans. Amer. Fish. Soc.* **107(2)** : 255-262.

Malvestuto, S. P. (1996). Sampling the recreational creel. Pages 591-623 in B. R. Murphy and D. W. Willis. *Fisheries techniques,* 2nd edition. American Fisheries Society, Bethesda, Maryland.

McNeish, J. D. and Trial, J. G. (1991). A cost-effective method for estimating angler effort from interval counts. *American Fisheries Society Symposium* **12**, 236-243.

Olkin, I. (1958). Multivariate ratio estimation for finite populations. *Biometrika* **45**, 154-165.

Palermo, V. and Ennevor, B. (1997). Catch & effort surveys conducted on the first nation sockeye fisheries in the Fraser river between Sawmill Creek and Kelly Creek, 1995 and 1996. Unpublished manuscript, Department of Fisheries and Oceans, Government of Canada.

Paker, R. A. (1956). Discussion. Pages 59-62 in K. D. Carlander, editor. Proceedings of Iowa State Creel Survey Symposium. Ames, Iowa.

Pippen, K. W. and Bergersen, E. P. (1991). Accuracy of a roving creel survey's harvest estimate and evaluation of possible sources of bias. *American Fisheries Society Symposium* **12**, 51-60.

Pollock, K. H., Hoenig, J. M.,Jones, C. M., Robson, D. S. and Greene, C. J. (1997). Catch rate estimation for roving and access point surveys. *North American Journal of Fisheries Management* **17**, 11-19.

Pollock, K. H., Jones, C. M. and Brown, T. L. (1994). *Angler survey methods and their applications in fisheries management.* Special Publication 25, American Fisheries Society, Bethesda, Maryland.

Pollock, K. H. and Kendall, W. L. (1979). Visibility bias in aerial surveys: a review of estimation procedures. *J. Wildl. Manage.* **51(2)** : 502-510.

Robson, D. S. (1960). An unbiased sampling and estimation procedure for creel censuses of fishermen. *Biometrics* **16**, 261-275.

Robson, D. S. (1961). On the statistical theory of a roving creel census of fishermen. *Biometrics* **17**, 415-437.

Robson, D. S. and Jones, C. M. (1989). The theoretical basis of an access site angler survey design. *Biometrics* **45**, 83-98.

Rose, C. D. and Hassler, W. W. (1969). Application of survey techniques to the dolphin, *Coryphaena hippurus*, fishery of North Carolina. *Trans. Amer. Fish. Soc.* **98**, 94-103.

Särndal, C. E., Swensson, B. and Wretman, J. (1992). *Model assisted survey sampling.* New York: Springer-Verlag.

Siegler, W. F. and Siegler, J. F. (1990). *Recreational fisheries: Management, theory, and application.* Reno: University of Nevada Press.

Shardlow, T. F. and Collicutt, L. D. (1989). Strait of Georgia sport fishery creel survey statistics for salmon and groundfish, 1984. Can. Man. Rep. Fish. Aquat. Sci.; **2032** : vii 61 p.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Asoke, C. (1984). *Sampling theory of surveys with applications*, 3rd edition. Ames, Iowa: Iowa State University Press.

Tarzwell, C. M. and Miller, L. F. (1943). The measurement of fishing intensity on the lower T. V. A. reservoirs. *Trans. Amer. Fish. Soc.* **72**, 246-256.

Taylor, H. M. and Karlin, S. (1984). *An introduction to stochastic modeling.* San Diego, California: Academic Press, Inc.

Wade, D. L., Jones, C. M., Robson, D. S. and Pollock, K. H. (1991). Computer Simulation techniques to assess bias in the roving-creel-survey estimator. *American Fisheries Society Symposium* **12**, 123-138.