STATISTICAL THEORY FOR HYDROACOUSTIC AND MARK-RECAPTURE ESTIMATION OF FISH POPULATIONS

by

B.M. Sarath Gamini Banneheka
B.Sc.(Hons.), University of Sri Jayewardenepura, Sri Lanka, 1985
M.Sc.(Mathematics), King's College London, University of London, England, 1987
M.Sc.(Statistics), Simon Fraser University, Canada, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department
of
Mathematics and Statistics

SIMON FRASER UNIVERSITY
May 1995

APPROVAL

Name: B.M. Sarath Gamini Banneheka

Degree: DOCTOR OF PHILOSOPHY

Title of thesis: STATISTICAL THEORY FOR HYDROACOUSTIC AND MARK-RECAPTURE ESTIMATION OF FISH POPULATIONS

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

Dr. Richard D. Routledge
Senior Supervisor

Dr. Charmaine B. Dean

Dr. Richard A. Lockhart

Dr. Carl J. Schwarz

Dr. John R. Skalski, University of Washington
External Examiner

Date Approved: May 19, 1995

PARTIAL COPYRIGHT LICENSE

Title of Thesis/~~Project/Extended Essay~~

Statistical Theory for Hydroacoustic and Mark-Recapture

Estimation of Fish Populations

Author:_
(signature)

B.M. Sarath Gamini Banneheka
(name)

May 19, 1995.
(date)

Abstract

Hydroacoustic sampling is a widely used technique for the estimation of fish population sizes. The need to estimate the volume of the hydroacoustic beam makes this method difficult. Recent research has focused on overcoming this difficulty. Existing techniques, called 'duration-in-beam methods' estimate the volume of the beam assuming it is cone shaped.

We present a hydroacoustic sampling technique that captures the advantages of the duration-in-beam method, but avoids the cone-shaped assumption. The method was specifically developed for making in-river estimates of the abundance of migrating fish, and has also been adapted to estimate resident fish populations. The population sizes in these cases can be written as sums of products of pairwise means. The problem of constructing approximate confidence intervals for such quantities is examined. Specific theory for normal data are provided and suggestions are made for other situations.

Tag recovery methods are also commonly used to estimate fish (and other animal) populations. Two sample tag recovery experiments when the tagging and recovering are stratified have been studied by several authors, who have provided maximum likelihood estimates and moment type estimators. We enrich this class of estimators by introducing the least squares estimates which are applicable when the numbers of tagging and recovery strata are unequal. The asymptotic variances of these estimates are also provided.

A common practice in stratified tag recovery experiments is to pool the strata before or after the experiment. This can produce inconsistent estimates. Sufficient conditions for the consistency of the estimates in the case of complete pooling are already known. We examine sufficient conditions for the case of partial pooling.

Acknowledgments

I would like to express my sincere appreciation and gratitude to my senior supervisor Prof. Richard Routledge for providing constant encouragement and invaluable guidance with much patience throughout the course of my study. I feel very fortunate to have him as my supervisor. My thanks also go to Prof. Carl Schwarz who suggested the problems in Chapters 5 and 6, and provided suggestions for their solution.

I will always remember the Department of Mathematics and Statistics of Simon Fraser University as a friendly place where I received help whenever needed. Specifically, I consider the opportunity to have known and worked with Professors Richard Routledge, Charmaine Dean, David Eaves, Richard Lockhart, Michael Stephens, Carl Schwarz, Tim Swartz, and Larry Weldon as an asset. I immensely value the education that I received from them and gratefully acknowledge the considerable help provided to me. I also wish to thank Sylvia Holmes, Maggie Fankboner, Dianne Pogue, and other department staff. The financial support provided by the Department of Mathematics and Statistics, and from research grants received by Prof. Routledge is greatly appreciated.

I would like to extend my thanks to Ian Guthrie and Jim Woodey of the Pacific Salmon Commission, for helping me understand the practical side of the problem discussed in Chapter 3 and for providing necessary information to solve the problem.

Finally I thank all my colleagues whose help and friendship made my stay at Simon Fraser University comfortable. A special thanks to my dear friend and colleague Chandanie Perera, without whose kind, caring support, the study and stay in Canada would have been much difficult.

Dedication

To my parents

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Estimation of the sizes of animal populations is required for many purposes such as wildlife management, fisheries and pest control. Recently, there has been a growing interest in this branch of statistics. Advances in general statistical theory and instrumentation have created opportunities for substantive improvements.

Hydroacoustic sampling and tag-recovery studies are two widely used methods in estimating fish populations. For example, these methods are used for the estimation of sockeye (*Oncorhynchus nerka*) and pink (*Oncorhynchus gorbuscha*) salmon in the Fraser River, Canada. In hydroacoustic sampling, a beam of sound waves is transmitted through water. The number of fish detected by the beam is then extrapolated to estimate the fish population. In-river hydroacoustic sampling provides quick daily estimates of abundance in the Fraser. Quick estimation is essential for efficient management of fisheries. Tag-recovery experiments in spawning areas provide estimates of numbers of fish reaching spawning grounds. These estimates are useful in post-season evaluation of the management programme, and in monitoring the dynamics of the stocks.

With the development of new data collection techniques such as hydroacoustic sampling methods, radio telemetry, remote sensing etc., the need arises for suitable statistical methods to extract the information from such data. The first half of this thesis focuses on hydroacoustics. Chapter 2 provides a brief introduction to hydroacoustic sampling techniques.

Our interest in this area was initiated by a request made by the Pacific Salmon Commission. They needed to gauge the variance of a hydroacoustic abundance estimator developed by the International Pacific Salmon Fisheries Commission (IPSFC) in 1977. This method had been used by the IPSFC (from 1977 until 1986) and by its successor, the Pacific Salmon Commission (since 1986) for estimating the abundance of Fraser River sockeye and pink salmon that pass Mission, British Columbia, during their upstream migration to spawning grounds.

We refined the formula for the estimator itself, with the aim of improving its statistical behavior while maintaining the practical advantages of the existing method. In Chapter 3, we describe the derivation of this estimator. As well, formulae are derived for the bias and the variance of the estimator. This method is now being tested for full-scale implementation by the Pacific Salmon Commission. In Chapter 3, we also suggest suitable modifications of the method to estimate resident fish populations.

The quantities we estimated can be written in terms of sums of products of pairwise means. We consider the problem of constructing confidence intervals for such quantities in Chapter 4. The theoretical results are derived formally for constructing approximate confidence intervals for normally distributed data. However, often the sample sizes in these kind of experiments are large, and a central limit theorem can be used to adapt the normal-based results for non-normal distributions. The validity of the theoretical results derived in this chapter is examined by a Monte Carlo study. We end Chapter 4 with a description of the Monte Carlo study.

Tag recovery experiments are widely used in estimating animal populations. In the simplest kind of tag-recovery experiments, a random sample of animals is tagged and released. After allowing for the tagged animals to disperse, another random sample is taken from the population and the number of tagged animals in the sample is recorded. These experiments are therefore often called two-sample tag-recovery experiments. When the population is stratified geographically or temporally, this feature can be exploited to obtain more precise estimates of the population size. In this method, a known number of tagged animals is released in each stratum, using a different tag for each stratum. After allowing for the tagged animals to disperse, a random sample is taken from each recovery stratum. The numbers of untagged animals and the numbers of tagged animals of each type in each sample are then recorded. These experiments are often called stratified tag-recovery experiments. Maximum likelihood estimators and moment-type estimators have already been developed for estimating population sizes using stratified tag-recovery data. When the number of tagging strata and recovery strata are unequal, these require imposing restrictions on unknown survival probabilities to obtain the estimates. In Chapter 5, we derive least squares estimates that avoid this requirement. We also prove the consistency of the derived estimates and provide formulae for the variances of the estimates.

In certain situations, the experimenter is unable to use different marks for different strata, is unaware of how the population is stratified, or, after the collection of the recovery data, finds only a very small number of a certain type of tags. A common practice in such situations is to pool the strata before or after the experiment. This can produce inconsistent estimates. Sufficient conditions for the consistency of the estimates in the case of complete pooling are already known. In Chapter 6, we discuss the estimation problem and provide conditions for consistent estimates to be produced after partial pooling.

Chapter 2

Use of Echo Sounding Techniques for Detecting Fish

In hydroacoustic sampling, a beam of sound waves is transmitted through water. The number of fish detected by the beam is used to estimate the total number of fish in a given volume, or the total number of fish that pass a given point. A hydroacoustic sampling method for the detection of fish was first reported in the scientific literature in 1929 (Kimura [19]). In this experiment, continuous waves at a frequency of 200kHz were directed across ponds containing goldfish. Crittenden [11] provided a review of the historical background up to 1970.

The purpose of this chapter is to introduce the terminology used in the next chapter, and to provide a brief summary of the 'duration in beam method,' that is used to estimate the volume of the hydroacoustic beam. Further details can be found in Burczynski [6], Thorne [39], [40], Johannesson and Mitson [16], and Crittenden [11].

2.1 Terminology

2.1.1 Echo Sounder

Generally speaking, an apparatus used for obtaining information about underwater objects and events with the use of sound waves is called a 'sonar system'[1]. Sonar systems used in fisheries work produce ultrasounds, i.e. sounds with a frequency usually ranging from 12 to 500 kHz (which are not detectable by a human ear). According to Burczynski [6], a sonar system that transmits vertically is usually called an 'echo-sounder' (Figure 2.1 a) while a sonar system that transmits horizontally is called 'a sonar' (Figure 2.1 b).

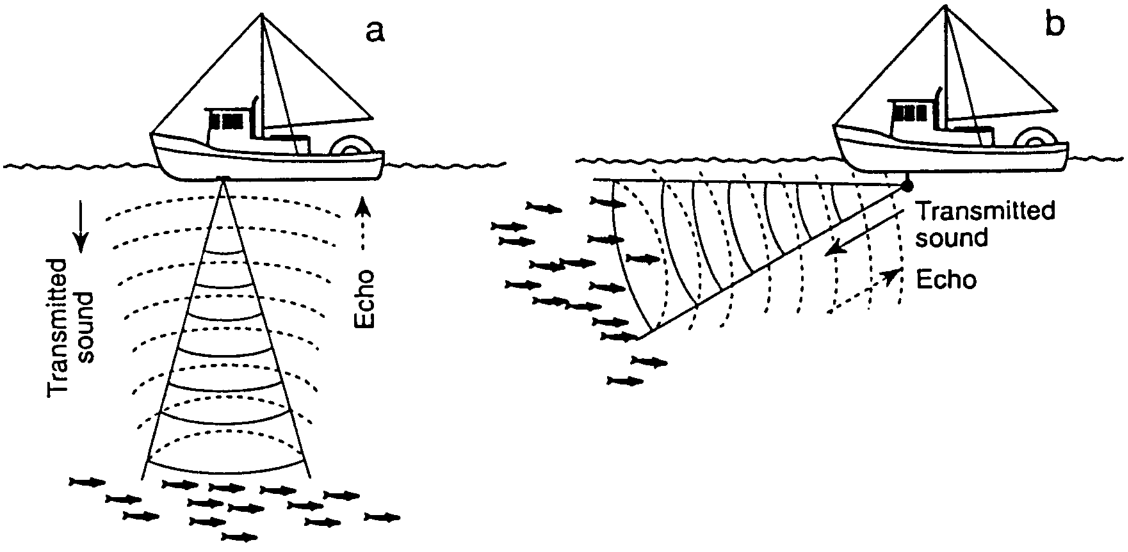


Figure 2.1: (a) Echosounder (b) Sonar

---

[1] sonar is an abbreviation for Sound Navigation and Ranging

An echo-sounder consists of four main components (Figure 2.2), the transmitter, the transducer, the receiver, and the recorder. The function of the transmitter is to produce energy in the form of pulses of electrical oscillations. A pulse is generated when a timer activates the electrical transmitter at a known frequency for a fixed period of time. Usually this frequency is 38kHz to 120 kHz. The time interval during which the transducer actually vibrates in generating each pulse, is called the 'pulse duration'. Typically this is about one millisecond. The electrical oscillations thus generated are then converted mechanically into sound waves in the water at the vibrating face of the transducer, which continues to generate sound until the timer switches off the transmitter. The result is a sound pulse of the same frequency, traveling through the water away from the face of the transducer. The number of pulses (or transmissions) sent out per unit time is called the 'pulse repetition rate'.

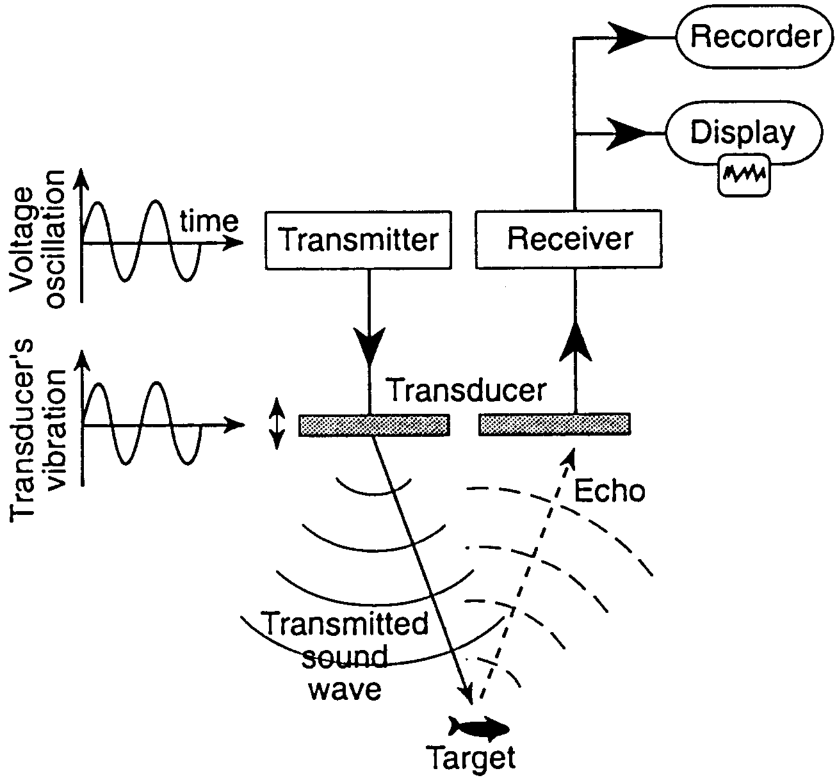


Figure 2.2: Functioning of an echo-sounder

After each pulse the system waits for a certain period to receive echoes from targets (e.g. fish) in the ensonified volume of water. For example, when a transducer is operating at a pulse repetition rate of sixty transmissions per minute, it generates sound for one millisecond in the transmitting mode, then waits for 999 milliseconds in the receiving mode, then generates the next one-millisecond pulse and so on. The transducer, when in the waiting mode, performs the reverse of its function in the transmitting mode, i.e. it converts pressure oscillations produced at its face by the echo, into electrical oscillations. The function of the receiver is to amplify these oscillations so that they can be recorded or displayed.

2.1.2 Echo Trace (Target), and Echogram.

The detectable sign on the display unit is called the 'echo trace'. The record of echo traces is called the 'echogram' (see Figures 3.2 and 3.3). A trained technician can usually distinguish echo traces of fish, from those of other objects. In the next chapter, we loosely use the term 'target' for an echo trace of a fish.

2.1.3 Range

The distance between the fish and transducer can be calculated by multiplying half the time interval between transmitting the pulse and receiving its echo by the sound velocity. The average distance of the fish from the transducer at the first detection and the last detection is called the 'range' of the fish.

2.1.4 Directivity Pattern

A transducer can be regarded as an array of point sources of sound. A point source radiates sound uniformly in all directions. However if two point sources that generate the same sound are located fairly close to one another, the resulting sound intensity will not be the same in all directions. In some directions, the waves from the two sources will be in phase or nearly so, and will reinforce each other to produce a high density; in other directions, they will cancel each other. This phenomenon, called interference, produces a regular but non-uniform distribution of sound intensity according to direction. The intensity distribution of a transducer has a maximum on its axis (a line passing through its centre perpendicular to the surface) and lower values elsewhere. It is convenient to describe the intensity distribution, or 'directivity pattern', of a transducer as the relative intensity $I_{\phi,\theta}$, at each bearing (ϕ,θ) (see Figure 2.3) with respect to intensity $I_{0,0}$, along the axis. The ratio of the two intensities at a given range is constant. This ratio is denoted by $b(\phi,\theta) = \frac{I_{\phi,\theta}}{I_{0,0}}$.

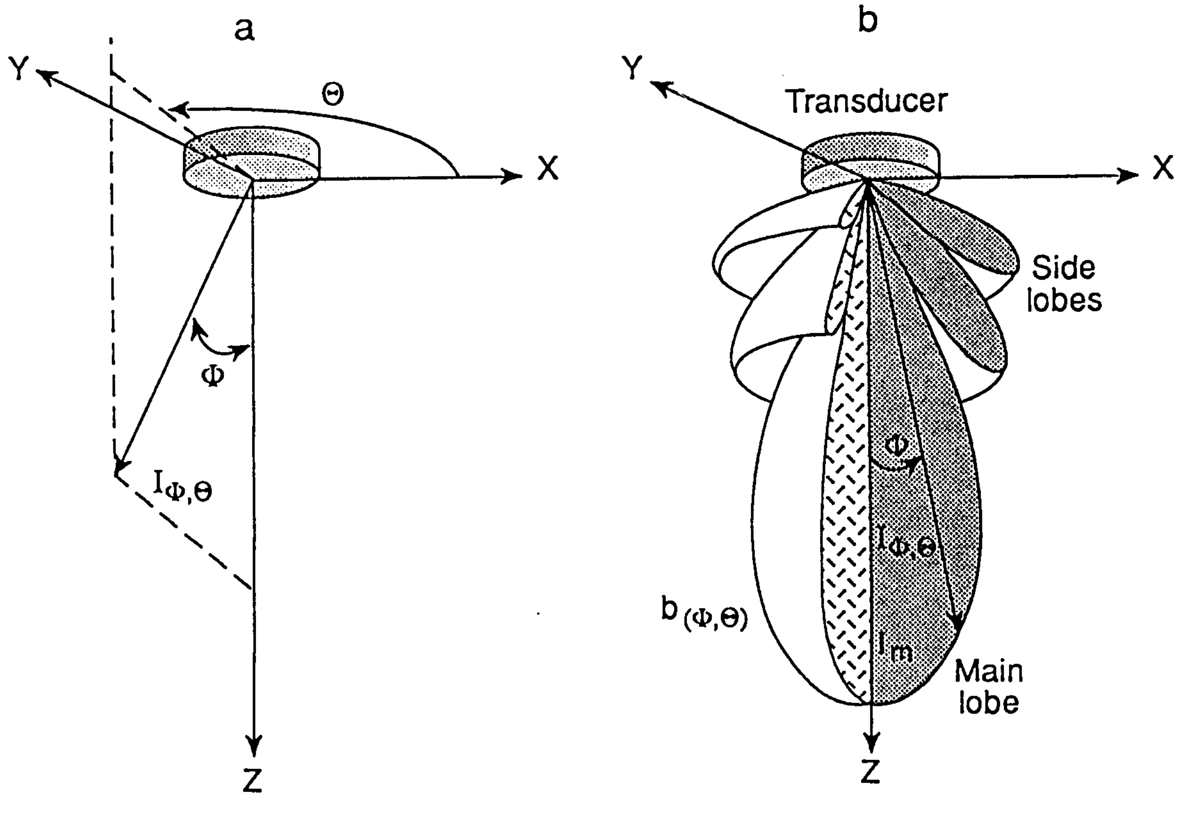


Figure 2.3: Three-dimensional view of the directivity

According to Clay and Medwin [8], a cylindrical transducer produces a directivity pattern that is symmetrical about the acoustic axis. Figure 2.3 shows a typical directivity pattern of a circular transducer. Intensity is highest along the axis. With increasing deviation from the axis, intensity first decreases to a minimum, then increases again to a value much lower than the maximum on the axis, then decreases and so on. The region of high values near the axis is called the main lobe of the beam and the successively smaller peaks around the main lobe are called side lobes.

2.1.5 Target Strength

Experiments have shown that fish species with swim bladders reflect about 85 percent of the sound energy by the swim bladder. The ratio of echo intensity from a target to the incident intensity is called the 'target strength'. The target strength (or scattering cross section) of a fish has a directivity pattern quite similar to that of a transducer. In general the directivity pattern depends on the anatomy of the fish, its overall size, and the dimensions of the swim bladder relative to the wave length. For a given species of fish, and for a given wavelength, there is a close relationship between the target strength and the size of the fish; the larger the fish, the larger its target strength. The relative intensity of sound reflected by a fish back to the source also depends on its orientation relative to the source.

2.1.6 Effective Beam

As a sound signal propagates through water, its intensity decreases with distance. There are two reasons for this: the dispersal of energy due to spreading effect, which is called the geometric loss, and the temperature and frequency-dependent power loss, which is called the absorption loss. The sum of these two types of energy loss is called the transmission loss. The intensity of the signal is further reduced according to reflective properties of the target and by the subsequent transmission loss of the echo. According to Thorne [40], the intensity of the signal (I_e) when it returns to the transducer can be described as

$$I_e = \frac{10^{k/10} I_i \, 10^{-2\alpha R} b^2(\phi, \theta)}{R^4} \tag{2.1}$$

where

k = target strength,

I_i = intensity of the signal at a unit range from the transducer,

α = absorption coefficient (a function of signal frequency, water temperature and salinity),

R = distance to the target, and

$b(\phi, \theta)$ = directivity pattern.

The absorption coefficient α is negligible for short or moderate ranges in fresh water. Therefore, if the receiving intensity threshold of the transducer is T, the maximum detection range for a fish of given target strength k, at bearing (ϕ, θ) can be approximated by

$$R = \left[\frac{10^{k/10} I_i \, b^2(\phi, \theta)}{T} \right]^{1/4} \tag{2.2}$$

The volume defined by this (R, ϕ, θ) relationship given by (2.2) is the 'effective beam' for fish of target strength k. It is usually hoped that the targets encountered in side lobes will return echoes that are below the detection threshold. Then, for a cylindrical transducer the effective beam can be approximated by a single lobe that is radially symmetrical about the acoustic axis.

2.2 Estimation of Beam Volume: Duration-in-Beam Method

Most hydroacoustic surveys attempt to estimate the number of fish in a population by extrapolating the number of fish counted in the acoustic beam to the entire survey region according to the ratio between the volume of water in the survey region and the volume sampled by the acoustic beam (e.g., Thorne and Dawson [38]). A similar method is used by

Skalski *et al.* [36] to estimate fish passage at a dam using a weighted sum of the numbers of fish ensonified by a beam at different depth ranges. The weights are determined by the dimensions of the beam and the dam passage ways. A major difficulty with these methods is that the volume of the acoustic beam must be estimated. Not only is the shape of the beam complicated by interference patterns that can produce side lobes (Clay and Medwin [8], pp.144-146; MacLennan and Simmonds [24], pp. 13-20; and references therein), but also the effective dimensions of an acoustic beam vary with the size, orientation and swimming movements of fish that encounter the beam (Clay and Medwin [8], 245; MacLennan and Simmonds [24], 137; and references therein). Dependence on estimates of beam width also creates problems with enumerating populations that have a wide distribution of body sizes.

Usually the effective beam is assumed to be a simple cone with an unknown angle. One popular method for estimating the effective sampling beam volume is the empirical technique called the 'duration-in-beam method'. This method was first conceived by R. Thorne and H. Lahore in 1970 and first applied to estimation of juvenile sockeye salmon in Lake Washington in 1971. (Thorne [39], Thorne and Dawson [38]). Since then several authors have derived alternative estimates based on the same principle (e.g. Nunnallee and Mathisen [27], Crittenden *et al.* [12], and Kieser and Ehrenberg [18]).

According to Thorne [39], the effective diameter of the beam at a given depth is estimated from the number of successive echoes received from individual fish ('duration-in-beam') as the transducer moves over the fish at a known speed. Any movement of fish is assumed to be negligible compared to the boat speed. His estimate of the diameter at range R is calculated as

$$d_R = \frac{4\bar{e}(R)u}{\pi\nu},$$

where $\bar{e}(R)$ is the average number of echoes from individual fish at range R, ν is the pulse repetition rate and u is the boat speed. The beam angle and the volume can now be estimated by simple trigonometry from the beam diameter and the range.

Nunnallee and Mathisen [27] proposed an unbiased estimator for the sine of the half angle of the assumed cone-shaped beam. This estimator is based on the relationship between the expected value of the number of echoes from individual fish and the range R from the transducer to the fish. They found that

$$E(e) = \frac{\psi \pi \nu}{2u} R, \tag{2.3}$$

where ψ is the sine of the half beam angle. They considered several range strata, and for each stratum, ψ was estimated by $2\bar{e}u/\pi\nu\bar{R}$. Then, taking the average of these estimates over strata (say a strata), ψ was estimated by

$$\hat{\psi} = \frac{2u}{\pi\nu} \frac{\sum_{i=1}^{a} (\bar{e}_i/\bar{R}_i)}{a}.$$

Crittenden *et al.* [12] suggested an alternative estimator based on weighted regression. They noticed that equation (2.3) can be expressed in the form of a model that is linear in R, has slope $\beta = \frac{\psi\pi\nu}{2u}$, and passes through the origin,

$$e_i = \beta R_i + \epsilon_i.$$

Finding that the variance of e is proportional to R^2, they regressed e on R with weights $w = 1/R^2$. Then, based on the estimated slope of this regression, they estimated ψ by

$$\hat{\psi} = \frac{2u}{\pi\nu} \frac{\sum w_i e_i R_i}{\sum w_i R_i^2} = \frac{2u}{\pi\nu} \text{Average}(e/R).$$

These methods represent a substantial reduction in the information needed to estimate fish abundance, because the calibration and enumeration are essentially done simultaneously using real targets. Nunnallee and Mathisen's [27] estimate for sine of the cone angle leads to an unbiased estimate of beam volume. Crittenden *et al.*'s [12] estimate for sine of the cone angle leads to a minimum variance unbiased estimate of the beam volume. However, these estimates do not lead to a direct unbiased estimate of fish population size because the population size is not a linear function of this quantity. It is a function of the the reciprocal beam volume. In the next chapter, we present an alternative method that requires only that the beam be circular symmetric and does not require to estimate the volume by duration-in-beam method.

Chapter 3

Estimation of Fish Populations Using Echo Sounding Techniques

3.1 Estimation of Migrating Fish Populations

The purpose of this section[1] is to present a hydroacoustic sampling technique that captures the advantages of the duration-in-beam method, but that is particularly suitable for making in-river estimates of the abundance of migrating fish. The technique was developed by the International Pacific Salmon Fisheries Commission (IPSFC) in 1977, and used by the IPSFC from then until 1986, and by its successor, the Pacific Salmon Commission, since then. The application was the estimation of abundance of Fraser River sockeye (*Oncorhynchus nerka*) and pink (*Oncorhynchus gorbuscha*) salmon that pass Mission, British Columbia, during their upstream migration to spawning grounds in the Fraser River watershed (Woodey [42]). We have refined the calculation methods to improve the statistical properties of the estimator and have developed formulae for estimating the variance and bias of the estimate. With this method, a boat-mounted echo sounder that conducts shore-to-shore transects

---

[1] A paper based on the contents of this section has been submitted to the Canadian Journal of Fisheries and Acquatic Sciences as: Banneheka, S.G., R.D. Routledge, I.C. Guthrie, and J. C. Woodey. 1995. Estimation of In-River Fish Passage Using a Combination of Transect and Stationary Hydroacoustic Sampling.

across the river is used to estimate the density of fish targets per depth interval. Stationary soundings provide an estimate of the amount of time it takes fish in each depth interval to swim through the beam. Combining these estimates yields estimates of the number of fish that pass the transect location per unit time. The method eliminates the need to estimate beam dimensions directly, as long as the same hydroacoustic equipment and settings are used for both the transect and stationary soundings. The sampling techniques used by the Pacific Salmon Commission are further described in the following section.

3.1.1 Methods

The hydroacoustic program conducted by the Pacific Salmon Commission in 1993 at Mission, British Columbia, is briefly described below. Minor technical aspects of the methods varied over time as, for example, the abating spring freshet led to declining river depths. The specific methods described below were applied on August 31. Data from this day are later used in a numerical example illustrating the calculations.

Equipment

The equipment that was used for both the transect and stationary soundings was installed on a 5.8 m aluminum vessel. A BioSonics Model 105 echo-sounder operating at a frequency of 50 kHz was used with a 34 degree (full angle) circular-beam transducer. The transducer was mounted in an aluminum stabilizer and towed from a davit 1.1 m from the starboard gunnel at a depth of 0.5 m. Settings for the equipment included a pulse rate of 15.3 ensonifications per second and, to compensate for two-way spreading loss, a time-varied gain of 40 Log R (where R = range of target). To record the soundings, a BioSonics Model 111 Thermal Chart Recorder was used. It was set to record a 15 m depth range with 5 m depth intervals marked. This depth range encompassed the maximum river depth at the site, except during spring freshet, when the depth range was increased to 20 m. Paper speed for the recorder was 0.478 mm s^{-1}. The equipment was serviced annually, and each day the crew checked the settings to ensure that they had not been accidentally changed.

Site

The data collection site was about 1 km upstream of the railway bridge across the Fraser River at Mission, British Columbia. River width at this site is about 437 m. This location has been used by the IPSFC and the Pacific Salmon Commission for echo sounding since 1977 because it satisfies several criteria. First, fish actively migrate at this location: the site is in a straight stretch of the river with no eddies or deep pools where fish can hold or mill about. Second, it has a rapid drop-off to deep water on both shores: there are only a few meters along each shore that cannot be sampled, and few sockeye salmon are believed to migrate this close to shore. Third, the smooth bottom is unlikely to acoustically obscure fish. Fourth, the flow of the river here is relatively even and moderately slow so that a small boat can cross the river at suitable speeds for echo sounding without being swept downstream. Finally, the site is close to a road, dock and fuel source.

Field Sampling

Two data collection activities were performed (Figure 3.1). The first consisted of 215 shore-to-shore transect soundings across the river, perpendicular to the direction of fish movement. All transects used the same transect line, boat speed, and echo-sounder settings. These transect soundings took about five minutes each to complete.

Second, nine stationary soundings were conducted along the transect line, using the same equipment and settings used for the transect soundings. These soundings were taken at scheduled times, with one located in each of the south, center and north sections of the transect line during each 8-hour shift. The boat was anchored during these soundings.

Hydroacoustic data were collected over a twenty-four hour cycle, from 5:00 a.m. to 4:59 a.m. the next morning. Transect soundings were conducted for about eighteen hours, and stationary soundings for about three hours. Other activities such as refuelling, changing crews and setting anchors for the stationary soundings took up the remaining three hours.

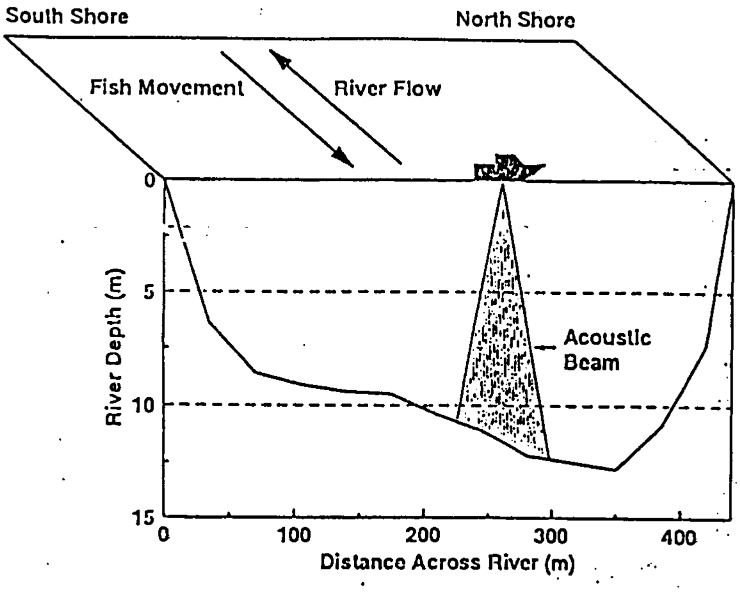


Figure 3.1: Schematic diagram showing hydroacoustic sampling activities on the Fraser River at Mission (courtesy of the Pacific Salmon Commission, Vancouver).

Data processing

After the 24-hour recording session, the resulting echogram was interpreted by a trained hydroacoustics biologist. For each transect, the number of fish targets in each depth stratum were counted (see Figure 3.2). For each stationary sounding, the widths of each echo traces from fish (target widths) in each depth stratum were measured using a micrometer (see Figure 3.3). Periodically, the paper speed of the recorder was verified, using marks that the echo-sounder recorded on the echogram each minute.

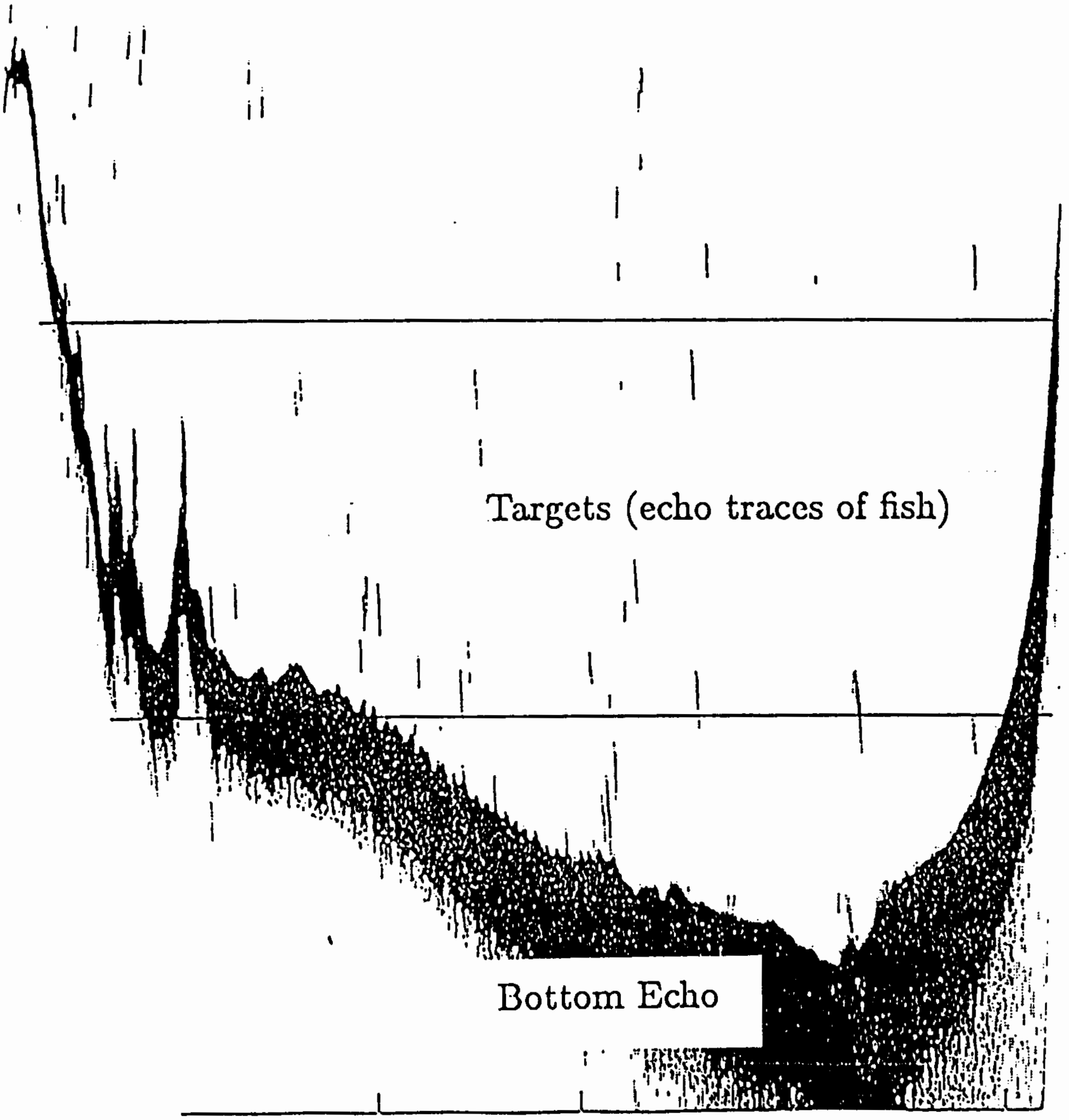


Figure 3.2: An echogram from a transect sounding (courtesy of the Pacific Salmon Commission, Vancouver)

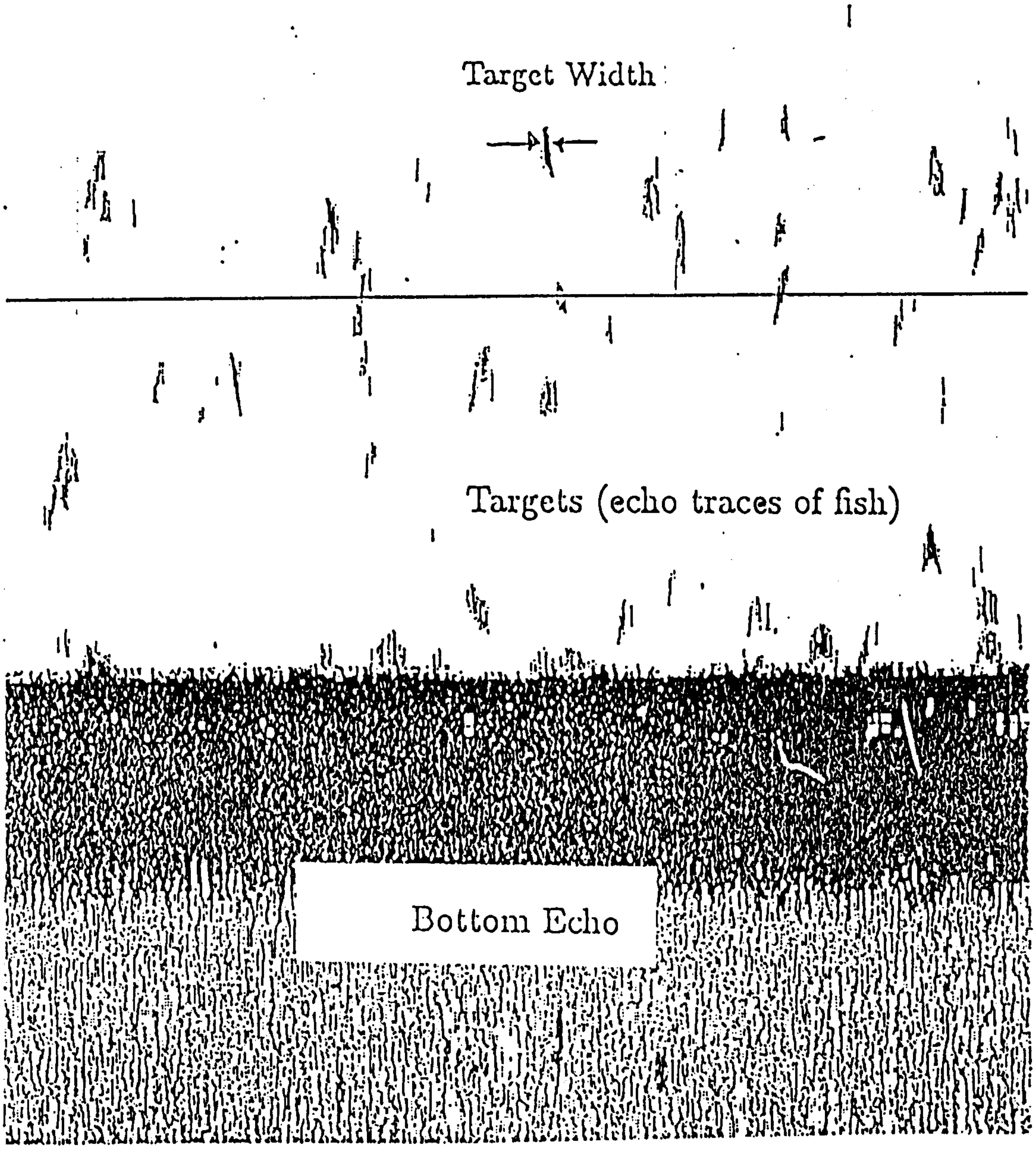


Figure 3.3: An echogram from a stationary sounding (courtesy of the Pacific Salmon Commission, Vancouver).

3.1.2 Assumptions

1. If there are fish present that are not part of the population of interest (e.g., other species), then these can either be distinguished on the record or otherwise removed from the estimates. (In the Fraser River programme, the number of resident [i.e., non-migratory] fish was estimated by sampling when few salmon were in the river [i.e., before or after salmon runs or after commercial gillnet fisheries downstream]. The target density associated with resident fish was then removed from the daily total target density to obtain an estimate of daily salmon density, which was used to estimate the salmon population. To determine the sockeye salmon component in this estimate, the species composition of these co-migrating salmon was obtained from a test fishery using a variable-mesh gillnet that was conducted downstream of the Mission site.)

2. The beam has circular horizontal cross sections. (This assumption is less stringent than the more common requirement that the beam be a cone, and is required because the presence of small side lobes near the transducer caused by local interference patterns (Clay and Medwin [8], 144-146; MacLennan and Simmonds [24], 13-20; and references therein) cannot be ignored, particularly in the shallow water near the shorelines at the sampling site.)

3. The hydroacoustic beam extends to the river bottom. (In the Mission echo-sounding programme, the depth range recorded by the echo-sounder encompassed the maximum depth encountered at the study site.)

4. The records of fish echos do not overlap on the record. (Schooling or heavy runs could generate high local abundances that would lead to substantial numbers of overlapping records. In extreme cases, fish density can be high enough to extinguish the bottom echo signal [Foote 1990]. In the present application, the local abundance of sockeye salmon was not large enough to cause significant overlapping of fish echoes.)

5. Fish echos can be distinguished from the background noise. (The ability to distinguish

fish from noise is determined by the experience of the interpreter and the selection of criteria that can be consistently applied to determine whether a target is "real" [e.g., three successive echoes].)

6. Fish are distributed randomly with a uniform probability density over depth within each stratum. (The validity of this assumption, along with the following one, can be promoted by choosing suitably narrow depth strata.)

7. Fish speeds are generated independently according to some common distribution which does not depend on depth within each stratum.

8. Fish swimming speed is negligible relative to the boat speed. (This assumption will later be relaxed.)

9. Fish behavior is not altered by the survey vessel. (Although research concerning this issue is lacking, sockeye salmon likely swim too deep in the turbid waters of the Fraser River to be very susceptible to boat avoidance. For example, few migrating sockeye at the study site swim at depths above 3 m (Levy *et al.* [23]), and few are caught near the surface in test fishing nets downstream of Mission.)

10. There are few fish swimming close enough to the surface and to the transducer to go undetected. (Since the large majority of sockeye at the study site migrate at depths below 3 m (Levy *et al.* [23]), and the minimum detection depth is likely between 1.5 and 2.5 m. [R. Kieser, Pacific Biological Station, Nanaimo, B. C., V9R 5K6, pers. comm.], it is likely that few sockeye salmon were undetected.)

11. The fish swim directly upriver. (A site should be chosen that minimizes opportunities for fish to mill about.)

12. Fish that cross the beam in a stationary sounding are independently and uniformly distributed across the beam. (This is essentially equivalent to a key assumption in Crittenden *et al.* [12].)

13. The observations from both transect and stationary soundings, though typically systematically sampled over time, can be viewed as being randomly sampled over time. (If, for example, there are substantial daily cycles, departures from this assumption could be minimized by stratifying over shorter time intervals. Failure to do so would inflate the estimated variance.)

14. The transect and stationary soundings sample fish with the same characteristics such as speed, size, depth, and orientation. (The validity of this assumption can be promoted by ensuring that the stationary soundings are taken at times that are interspersed spatially and temporally with the transect soundings.)

15. The target widths of individual fish in the stationary soundings are generated independently. (This assumption would be violated if, for example, fish speeds displayed a strong diurnal cycle, and the results were pooled over an entire day. The measured widths of adjacent targets, coming from roughly the same time of day, would then be positively correlated.)

3.1.3 The Estimator

Upstream Migration Rate

Suppose the water column is divided into several depth strata within each of which the number of fish per unit volume does not vary appreciably with depth. First, consider any depth stratum and assume that target strength is the same for each fish in the depth stratum.

Let

τ = number of fish migrating upstream per second in the depth stratum: the quantity to be estimated,

α = width of the river,

z_1 = depth at the top of the depth stratum,

z_2 = depth at the bottom of the depth stratum,

λ = average number of fish per unit volume in the depth stratum: unknown, and

γ = maximum speed of fish (distance per second): an unknown constant.

Next imagine a barrier across the river. At any given moment, which will be referred to as 'the start', fish that are capable of crossing the barrier within one second from the start will be within a distance γ downstream from the barrier. The volume of the river section that contains these fish is $\alpha(z_2 - z_1)\gamma$ and the expected number of fish in this volume is $\lambda\alpha(z_2 - z_1)\gamma$. However, only a fraction of these fish cross the barrier within one second from the start, because not all fish travel at the maximum speed γ. Let

β = probability that a fish will cross the barrier within one second from the start.

Then the expected number of fish that will cross the barrier within one second is

$$\tau = \lambda\alpha(z_2 - z_1)\gamma\beta. \tag{3.1}$$

Calculation of β

Let

D = distance from the fish to the barrier at the start: a random variable,

S = speed of the fish: a random variable, and

B = time taken by the fish to reach the barrier: a random variable.

Suppose that S has some probability density function $f_S(s)$ whose domain is $[0, \gamma]$. Over the small distance (γ) being considered, D will be essentially uniformly distributed in $[0, \gamma]$. Thus the probability density function of D is $f_D(d) = \frac{1}{\gamma}$. The variables, D and S, will also be essentially independent, with joint probability density function,

$$f_{D,S}(d, s) \;=\; f_D(d)f_S(s) = \frac{1}{\gamma}f_S(s), \text{for } d \text{ and } s \text{ both in } [0, \gamma].$$

Note that $B = \frac{D}{S}$, with B then restricted to the interval, [0,1]. Let $U = S$. Then $D = BU$. The joint probability density function of B and U is

$$\begin{aligned} f_{B,U}(b,u) &= f_{D,S}(bu,u)J, \text{ where } J = u \text{ is the Jacobian of the transformation,} \\ &= \frac{1}{\gamma} f_S(u)u; \quad (b,u) \in A, \\ &\quad \text{where } A \text{ is the region with boundaries, } b = 0, u = 0, u = \gamma \text{ and } bu = \gamma. \end{aligned}$$

Therefore, the marginal probability density function of B in region $[0,1]$ is

$$f_B(b) = \int_0^\gamma f_{B,U}(b,u)du = \int_0^\gamma \frac{1}{\gamma} f_S(u)udu = \frac{1}{\gamma}E(S) = \frac{\mu_s}{\gamma},$$

where $\mu_s = E(S)$ is the mean fish speed.

Now,

$$\beta = Pr(B \leq 1) = \int_0^1 f_B(b)db = \int_0^1 \frac{\mu_s}{\gamma} db = \frac{\mu_s}{\gamma}.$$

Substituting $\frac{\mu_s}{\gamma}$ for β in formula (3.1) yields

$$\tau = \lambda\alpha(z_2 - z_1)\mu_s. \tag{3.2}$$

This formula shows that the number of fish passing the barrier per second is equal to the number of fish per unit upstream distance $[\lambda\alpha(z_2 - z_1)]$ times the mean fish speed (μ_s). The next step is to calculate the mean fish speed in terms of quantities associated with stationary soundings.

Calculation of Mean Fish Speed, μ_s

Consider a stationary sounding. Let Z be the random variable representing the depth of a fish relative to the transducer. Assume that the horizontal cross section of the beam at depth z is a circle with radius $r(z)$. Now consider the fish that swim through the beam at depth Z. Suppose that S is the speed of the fish, C is the distance that a fish swims through the beam (chord length) and T is the time taken to do so. Speed is distance over time. However, mean fish speed is not equal to mean distance over mean time. Because C

and S are independent, the mean speed can be obtained as

$$\mu_s = E(S) = \frac{E(\frac{1}{T})}{E(\frac{1}{C})}. \tag{3.3}$$

This formula can also be written as $\mu_s = \frac{1/E(\frac{1}{C})}{1/E(\frac{1}{T})}$. This implies that mean fish speed is equal to the ratio of harmonic mean chord length to an harmonic mean time.

The next two steps are devoted to obtaining expressions for the numerator and the denominator of formula (3.3).

Calculation of Mean Reciprocal Time, $E(\frac{1}{T})$

Let

W = width of the target in the echogram: a random variable and

p = paper speed of the recorder.

Then, $W = Tp$. For notational convenience in the subsequent calculations, let $M = \frac{1}{W}$. Then,

$$E\left(\frac{1}{T}\right) = pE\left(\frac{1}{W}\right) = pE(M). \tag{3.4}$$

This formula can also be viewed as $1/E(\frac{1}{W}) = p[1/E(\frac{1}{T})]$, which means that the harmonic mean target width is equal to the paper speed times the harmonic mean time.

Calculation of Mean Reciprocal Chord Length, $E(\frac{1}{C})$

Let Θ be the half angle that is formed in the center of the horizontal cross section at depth Z by the chord that a fish travels in the beam (Figure 3.4). Then, $C = 2r(Z)\sin(\Theta)$ and hence,

$$\begin{aligned} E\left(\frac{1}{C}\right) &= E\left[\frac{1}{2r(Z)\sin(\Theta)}\right] \\ &= E\left[\frac{1}{2r(Z)}\right] E\left[\frac{1}{\sin(\Theta)}\right]. \end{aligned} \tag{3.5}$$

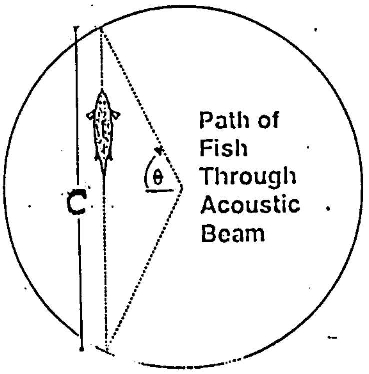


Figure 3.4: The distance that a fish swims to cross through the acoustic beam (C) can be calculated from the radius $r(z)$ of the beam at depth z and from the half angle θ formed in the center of the beam

Within each depth stratum, fish depths are distributed uniformly (Assumption 6). However, the sampling probability for a fish at depth z is proportional to $r(z)$. Hence for the sampled fish, the probability density function $f_Z(z)$ of Z is

$$f_Z(z) = \frac{1}{\int_{z_1}^{z_2} r(z)dz} r(z). \tag{3.6}$$

Let

$$\delta = \text{average diameter of the beam} = \frac{1}{(z_2 - z_1)} \int_{z_1}^{z_2} 2r(z)dz. \tag{3.7}$$

Formulae (3.6) and (3.7) lead to

$$f_Z(z) = \frac{2}{\delta(z_2 - z_1)} r(z), \quad \text{for} \quad z_1 \leq z \leq z_2. \tag{3.8}$$

Then,

$$E\left[\frac{1}{2r(Z)}\right] = \int_{z_1}^{z_2} \frac{1}{2r(z)} f_Z(z)dz = \frac{1}{\delta}. \tag{3.9}$$

Now let $f_\Theta(\theta)$ be the probability density function of Θ. It follows from Assumption 9 (that fish do not respond to the survey vessel) that fish that cross the beam do so along a line that is a uniformly distributed distance away from the beam centre. From this, it follows in turn through a change of variables that $f_\Theta(\theta)$ is proportional to $\sin(\theta)$. Then,

$$f_\Theta(\theta) = \frac{\sin(\theta)}{2}, \quad \text{for} \quad 0 \leq \theta \leq \pi, \text{ and}$$

$$E\left[\frac{1}{\sin(\Theta)}\right] = \int_0^\pi \frac{1}{\sin(\theta)} f_\Theta(\theta) d\theta = \frac{\pi}{2}. \tag{3.10}$$

Substituting formulae (3.9) and (3.10) in formula (3.5),

$$E\left(\frac{1}{C}\right) = \frac{\pi}{2}\frac{1}{\delta}. \tag{3.11}$$

Formula (3.11) can be written as $1/E(\frac{1}{C}) = \frac{2}{\pi}\delta$. That is, the harmonic mean chord length is equal to the product of (i) the average beam diameter and (ii) a correction factor, $2/\pi$, which adjusts for the shorter swimming distances for fish that do not cross the centre of the beam. The factor $2/\pi$ is valid for any beam with circular horizontal cross-sections.

Substituting formulae (3.3), (3.4) and (3.11) into formula (3.2) leads to

$$\tau = \frac{2}{\pi} p\lambda\alpha(z_2 - z_1)\delta E(M). \tag{3.12}$$

This formula still contains quantities which are not readily estimable. The next step shows that the quantity, $\lambda\alpha(z_2 - z_1)\delta$, is an estimable parameter arising from transect soundings.

Calculation of Expected Number of Fish per Transect

Let

N = number of fish detected per transect in the depth stratum.

Ignoring river bottom irregularities, edge effects near the shorelines, and fish less than the minimum detection distance from the transducer, the volume of the depth stratum that is ensonified by a transect sounding, v, can be approximated by

$$v = \alpha \int_{z_1}^{z_2} 2r(z)dz = \alpha(z_2 - z_1)\delta. \tag{3.13}$$

Now consider an ideal situation where the boat is moving at such a speed that a transect sounding can be considered a snapshot across the river (Assumption 8). Then, the expected number of fish that are ensonified during a transect sounding is

$$E(N) = \lambda v = \lambda \alpha (z_2 - z_1)\delta. \tag{3.14}$$

However, in practice, there is a maximum feasible speed of the boat. Therefore the assumption of high boat speed introduces a bias to the estimator of $E(N)$ and thereby to the final estimator of τ. This is discussed later in Subsection 3.1.4.

Calculation of Fish Passage per Second, τ

Now we can combine formulae (3.12) and (3.14) to obtain τ, the number of fish migrating upstream per second, in terms of estimable parameters as

$$\begin{aligned} \tau &= \frac{2p}{\pi}\lambda v E(M) && (3.15) \\ &= \frac{2p}{\pi}E(N)E(M). && (3.16) \end{aligned}$$

Extension to Unequal Target Strengths

So far, we have assumed that fish target strength is a constant. Now we allow it to be a random variable and denote it by K. For simplicity, assume that K is a discrete random variable that takes values in a countable set. The result for continuous K is similar, but the proof is technically more complex. Let $\tau(k), \lambda(k), v(k)$ and $N(k)$ refer to the corresponding quantities above, but for fish of target strength $K = k$. Then,

$$\begin{aligned} N &= \sum_k N(k), \quad \text{and} \\ E(N) &= E\left[\sum_k N(k)\right] && (3.17) \\ &= \sum_k E[N(k)]. \end{aligned}$$

Analogous to formula (3.14), $E[N(k)] = \lambda(k)v(k)$. Thus formula (3.17) can be written as $E(N) = \sum_k \lambda(k)v(k)$.

Let $E(M|K=k)$ be the conditional expectation of M given $K=k$. Then,

$$E(M) = E\left[E(M|K=k)\right] = \sum_k E(M|K=k)\frac{\lambda(k)v(k)}{\sum_k \lambda(k)v(k)}, \tag{3.18}$$

where $\frac{\lambda(k)v(k)}{\sum_k \lambda(k)v(k)}$ is the probability mass function of target strengths of fish in the stationary beam. Substitution of $E(N)$ for $\sum_k \lambda(k)v(k)$ in formula (3.18) yields

$$E(M) = \frac{1}{E(N)}\sum_k E(M|K=k)\lambda(k)v(k). \tag{3.19}$$

Analogous to formula (3.15),

$$\tau(k) = \frac{2p}{\pi}\lambda(k)v(k)E(M|K=k). \tag{3.20}$$

Then, formulae (3.19) and (3.20) lead to

$$E(M) = \frac{1}{E(N)}\sum_k \frac{\pi}{2p}\tau(k) = \frac{\pi}{2p}\frac{1}{E(N)}\tau,$$

which again gives formula (3.16). This shows that even with fish of unequal target strengths, τ can be written in terms of estimable quantities that do not depend on parameters relating to beam shape or target strengths of fish.

Now we attach subscript i to τ, N and M to denote the corresponding quantities for the ith depth stratum. Finally, summing over all depth strata, we can write the total number of fish migrating upstream per second, τ, as

$$\tau = \sum_i \tau_i = \frac{2p}{\pi}\sum_i E(N_i)E(M_i). \tag{3.21}$$

Estimation of Fish Passage per Second, τ

To estimate τ, we first need to estimate $E(N_i)$'s and $E(M_i)$'s. We use data from transect soundings to estimate $E(N_i)$'s and data from stationary soundings to estimate $E(M_i)$'s. Let

$\bar{N}_i$ = sample average of number of targets per transect in the ith depth stratum, and

$\bar{M}_i$ = sample average of inverse target widths from the ith depth stratum.

Then,

$$\hat{\tau} = \frac{2p}{\pi} \sum_i \bar{N}_i \bar{M}_i \tag{3.22}$$

can be used as an estimator for τ.

3.1.4 Bias

Assuming that a transect sounding is a snapshot across the river is equivalent to assuming that the boat is traveling infinitely faster than the fish. Since the boat speed is limited in practice, this assumption is false and introduces a bias to the estimate. If the boat speed is not substantially greater than the fish speed during transect soundings, then the beam will ensonify fish that are in the sampled volume when the boat arrives, as well as fish that move into the sampled volume as the boat passes. These latter fish are the source of the bias. As the beam detects more fish than the model predicts, this bias is positive.

Fish that deviate from the upstream direction when migrating upriver will also contribute a positive bias. This includes fish that swim downstream, across the flow, or even a few degrees from the true upstream direction. The fish that do not swim directly upstream are not moving upstream as rapidly as the method assumes. For fish that swim at an angle, Φ, off the upstream direction, the upstream component to the velocity will be the fraction, $\cos(\Phi)$, of their overall speed. This does no affect the calculation of average fish speed using stationary soundings, because of the circular symmetry of the beam. But, by taking this average fish speed as the average upstream fish speed, the upriver migration rate is overestimated. However, for Φ close to 0^o, $\cos(\Phi)$ is close to 1. Hence, as long as fish movement is strongly directed upriver, this bias will not be serious.

Formula (3.15) now becomes

$$\tau = \frac{2p}{\pi} \lambda v E(M) E\left[\cos(\Phi)\right]. \tag{3.23}$$

Since Φ is small,

$$\cos(\Phi) \approx \sqrt{1 - \Phi^2} \approx 1 - \Phi^2/2.$$

Suppose that $E(\Phi) = 0$ and $Var(\Phi) = \sigma_\Phi^2$. Then

$$E\left[\cos(\Phi)\right] \approx E\left[1 - \Phi^2/2\right] = 1 - \sigma_\Phi^2/2.$$

So, (3.23) yields

$$\tau \approx \frac{2p}{\pi}\lambda v E(M)(1 - \sigma_\Phi^2/2). \tag{3.24}$$

The relative bias (RB) of $\hat{\tau}$ is

$$RB(\hat{\tau}) = \frac{[E(\hat{\tau}) - \tau]}{\tau} = \frac{E(\hat{\tau})}{\tau} - 1.$$

Since from formula (3.22) $\hat{\tau}$ for one stratum is $\hat{\tau} = \frac{2p}{\pi}\bar{N}\bar{M}$, the above formula becomes

$$RB(\hat{\tau}) = \frac{E(\bar{N}\bar{M})}{\lambda v E(M)(1 - \sigma_\Phi^2/2)} - 1.$$

Recall that $\bar{N}$ is the average target count per transect sounding, and is based on mobile soundings, while $\bar{M}$ is the average of reciprocal target widths and is based on stationary soundings. These two quantities are therefore independent. Hence, the above equation simplifies to

$$RB(\hat{\tau}) = \frac{E(\bar{N})}{\lambda v(1 - \sigma_\Phi^2/2)} - 1 = \frac{E(N)}{\lambda v(1 - \sigma_\Phi^2/2)} - 1. \tag{3.25}$$

Now let

$$u = \text{speed of the boat.}$$

Then, the time taken for one transect is $t = \frac{\alpha}{u}$. Boat speed relative to a fish swimming upstream at speed S is $\sqrt{u^2 + S^2 + 2uS\sin(\Phi)}$. Therefore, the distance that the boat would have traveled relative to the fish during time t is $t\sqrt{u^2 + S^2 + 2uS\sin(\Phi)} = \frac{\alpha}{u}\sqrt{u^2 + S^2 + 2uS\sin(\Phi)}$. This corresponds to a volume,

$$V = \frac{\alpha}{u}\sqrt{u^2 + S^2 + 2uS\sin(\Phi)}\delta(z_2 - z_1) = v\sqrt{1 + \frac{S^2}{u^2} + 2\frac{S}{u}\sin(\Phi)}.$$

Therefore,

$$E(N) = E(\lambda V) = \lambda v E\left(\sqrt{1 + \frac{S^2}{u^2} + 2\frac{S}{u}\sin(\Phi)}\right). \tag{3.26}$$

Then, formulae (3.25) and (3.26) lead to

$$RB(\hat{\tau}) \approx \frac{E\left(\sqrt{1 + \frac{S^2}{u^2} + 2\frac{S}{u}\sin(\Phi)}\right)}{(1 - \sigma_\Phi^2/2)} - 1. \tag{3.27}$$

Now, note that

$$\begin{aligned}
E\left(\sqrt{1 + \frac{S^2}{u^2} + 2\frac{S}{u}\sin(\Phi)}\right) &\approx E\left[\sqrt{1 + \frac{S^2}{u^2} + 2\frac{S\Phi}{u}}\right] \\
&= E\left[\sqrt{\left(1 + \frac{S^2}{u^2}\right)\left(1 + 2\frac{S\Phi}{u}\frac{u^2}{S^2 + u^2}\right)}\right] \\
&= E\left[\sqrt{\left(1 + \frac{S^2}{u^2}\right)}\sqrt{\left(1 + 2\frac{S\Phi}{u}\frac{u^2}{S^2 + u^2}\right)}\right] \\
&\approx E\left[\sqrt{\left(1 + \frac{S^2}{u^2}\right)}\left(1 + \frac{S\Phi u}{S^2 + u^2}\right)\right] \\
&= E\left[\sqrt{1 + \frac{S^2}{u^2}}\right] + E\left[\sqrt{\left(1 + \frac{S^2}{u^2}\right)}\left(\frac{Su}{S^2 + u^2}\right)\right]E(\Phi) \\
&= E\left[\sqrt{1 + \frac{S^2}{u^2}}\right]
\end{aligned}$$

Therefore, (3.27) leads to

$$RB(\hat{\tau}) = \frac{E\left(\sqrt{1 + \frac{S^2}{u^2}}\right)}{(1 - \sigma_\Phi^2/2)} - 1. \tag{3.28}$$

As the bias is positive, the effect of low boat speeds is to overestimate τ. However, by moving the boat at a high speed relative to fish speed, the bias can be reduced. We can obtain bounds for the relative bias as follows.

From Jensen's inequality (see Appendix) and the fact that $\sqrt{1 + \frac{S^2}{u^2}}$ is convex in S but concave in S^2, one can show that

$$\sqrt{1+\frac{E^2(S)}{u^2}} \le E\left(\sqrt{1+\frac{S^2}{u^2}}\right) \le \sqrt{1+\frac{E(S^2)}{u^2}}.$$

Therefore,

$$\frac{\sqrt{1+\frac{E^2(S)}{u^2}}}{(1-\sigma_\Phi^2/2)} - 1 \le \frac{E\left(\sqrt{1+\frac{S^2}{u^2}}\right)}{(1-\sigma_\Phi^2/2)} - 1 \le \frac{\sqrt{1+\frac{E(S^2)}{u^2}}}{(1-\sigma_\Phi^2/2)} - 1.$$

That is,

$$\frac{\sqrt{1+\frac{\mu_s^2}{u^2}}}{(1-\sigma_\Phi^2/2)} - 1 \le RB(\hat{\tau}) \le \frac{\sqrt{1+\frac{\mu_s^2+Var(S)}{u^2}}}{(1-\sigma_\Phi^2/2)} - 1. \tag{3.29}$$

Hence, a lower bound can be placed on the bias if the mean fish speed and σ_Φ are known. An upper bound can be added if the variance is also known, and a proper bias adjustment factor can be computed from an estimate of $\frac{E\left(\sqrt{1+\frac{S^2}{u^2}}\right)}{(1-\sigma_\Phi^2/2)}$.

For example, assume that σ_Φ was 10 degrees, and the boat speed was close to 1.5 m s^{-1}. Quinn and terHart [29] report for sockeye salmon (*Oncorhynchus nerka*) a mean fish speed of 0.66 m s^{-1}, with a standard deviation of 0.21 m s^{-1}. With these Figures the relative bias would be between 10% and 12%.

3.1.5 Estimation of Variance

The variance of $\hat{\tau}$ as given in formula (3.22) can be written as

$$Var(\hat{\tau}) = \sum_i Var(\hat{\tau}_i) + 2\sum_{i<j} Cov(\hat{\tau}_i, \hat{\tau}_j),$$

where i and j are stratum numbers. Let $\hat{\tau}_i = \tau_i + \epsilon_i$ and $\hat{\tau}_j = \tau_j + \epsilon_j$. Assume that the errors ϵ_i and ϵ_j are uncorrelated. Then, $Cov(\hat{\tau}_i, \hat{\tau}_j) = Cov(\epsilon_i, \epsilon_j) = 0$, and

$$Var(\hat{\tau}) = \sum_i Var(\hat{\tau}_i) = \left[\frac{2p}{\pi}\right]^2 \sum_i Var(\bar{N}_i \bar{M}_i). \tag{3.30}$$

Recall that $\bar{N}_i$ and $\bar{M}_j$ are independent for all i and j.

According to Kendall and Stuart [17],

$$Var(\bar{N}_i\bar{M}_i) = Var(\bar{N}_i)Var(\bar{M}_i) + Var(\bar{N}_i)E^2(\bar{M}_i) + Var(\bar{M}_i)E^2(\bar{N}_i). \tag{3.31}$$

Let

l = the number of times the boat crossed the river, (3.32)

m_i = total number of targets in depth stratum i from the stationary soundings,

$s^2_{N_i}$ = sample variance of the number of targets per transect sounding in depth stratum i, and

$s^2_{M_i}$ = sample variance of the reciprocals of m_i target widths in depth stratum i from the stationary soundings.

Since $\frac{s^2_{N_i}}{l}$ and $\frac{s^2_{M_i}}{m_i}$ are unbiased estimators for $Var(\bar{N}_i)$ and $Var(\bar{M}_i)$, respectively, according to Kendall and Stuart [17] an unbiased estimate of the variance of $\hat{\tau}$ can be obtained by

$$\hat{V}ar(\hat{\tau}) = \left[\frac{2p}{\pi}\right]^2 \sum_i \left(\frac{s^2_{N_i}}{l}\bar{M}_i^2 + \frac{s^2_{M_i}}{m_i}\bar{N}_i^2 - \frac{s^2_{N_i}s^2_{M_i}}{lm_i}\right). \tag{3.33}$$

3.1.6 Numerical Example

Data from August 31, 1993, are used to demonstrate the methods described in this section. Table 3.1 contains the summary statistics of the data and estimates of the number of fish per second in each depth stratum that migrated upriver, along with the estimated variances. The paper speed of the recorder is 0.478 mm s^{-1} for these calculations.

The estimated number of fish migrating upriver per second is $\hat{\tau}$. To estimate total numbers migrating over a time period of duration H, we can use $H\hat{\tau}$ with estimated variance, $H^2\hat{V}ar(\hat{\tau})$.

The total daily abundance is estimated with $H = 86\,400$ s d^{-1} as described above.

		Transect Observations			Stationary Observations			Migration Rates	
Depth Stratum (i)	Depth Range (meters)	# Passes (l)	Ave. Count ($\bar{N}_i$)	Var. ($s^2_{N_i}$)	# Fish (m_i)	Ave. Recip. Width ($\bar{M}_i$)	Var. ($s^2_{M_i}$)	Est. ($\hat{\tau}_i$)	Var. $\left[\hat{V}ar(\hat{\tau}_i)\right]$
1	0 - 5	215	1.20	1.53	77	1.07	0.136	0.39	0.0107
2	5 - 10	215	6.38	17.4	145	0.76	0.066	1.47	0.0815
3	10 - 15	215	1.74	2.36	42	0.79	0.130	0.42	0.0164

Table 3.1: Summary statistics of the data collected by the Pacific Salmon Commission on August 31, 1993, at Mission, B.C. Also shown are the estimated numbers of fish migrating upriver per second and the corresponding estimated variances.

$$\begin{aligned}
\text{Estimated abundance} &= H\hat{\tau} = H\sum_{i=1}^{3}\hat{\tau}_i = 196,998, \\
\text{Estimated standard error} &= \widehat{se}(H\hat{\tau}) = H\sqrt{\hat{V}ar(\hat{\tau})} = H\sqrt{\sum_{i=1}^{3}\hat{V}ar(\hat{\tau}_i)} = 8,651, \\
\text{Estimated coefficient of variation} &= \frac{\widehat{se}(H\hat{\tau})}{H\hat{\tau}}\% = 4\%.
\end{aligned}$$

Fish abundance on this day (about 197,000 fish) was about half the largest daily abundance (about 403,000 fish) observed in 1993.

3.2 Estimation of Resident Fish Populations

In this section we adapt the ideas developed in Section 3.1 to derive an estimator for the total number of fish in a resident population; for example, in a lake. For this derivation, we make the assumptions 1-10 stated in Subsection 3.1.2, and in addition, the following assumptions.

11′. The fish swim around in random directions at slow speeds.

12′. The fish that are crossed by the moving beam are independently and uniformly distributed across the beam.

13′. The survey area is randomly sampled by transect soundings.

Here, we are not estimating the numbers of fish that pass a given point. Hence, one main difference of this situation from a migrating fish population in a river is that calculation of fish speed is not necessary. For this reason, stationary soundings no longer provide relevant information.

3.2.1 The Estimator

As in Subsection 3.1.3, suppose that the water body is divided into several depth strata within each of which the number of fish per unit volume does not vary appreciably with depth. First, we estimate the number of fish in a given depth stratum assuming that all the fish in the stratum have the same target strength. Later we relax this assumption. Then, the total number of fish is estimated by adding the stratum estimates.

Let

T = total number of fish in the depth stratum: the quantity to be estimated,

A = survey area: assumed known,

z_1 = depth at the top of the depth stratum,

z_2 = depth at the bottom of the depth stratum,

v_0 = total volume of the water in the stratum: assumed known,

λ = number of fish per unit volume in the depth stratum: unknown,

u = boat speed: assumed known,

α = length of a transect: assumed known,

n = expected number of fish to be ensonified in the depth stratum during a transect :unknown,

N = observed number of fish ensonified in the depth stratum during a transect, and

v = effective volume sampled in a transect: unknown.

By the definition of λ,

$$\lambda = \frac{T}{v_0}. \tag{3.34}$$

Under the random sampling assumption,

$$\lambda = E\left(\frac{N}{v}\right) = \frac{n}{v}. \tag{3.35}$$

Since $v_0 = A(z_2 - z_1)$, (3.34) and (3.35) lead to

$$T = \frac{A(z_2 - z_1)n}{v}.$$

Now let δ be the average diameter of the effective beam defined by (3.7). Since $v = \alpha(z_2 - z_1)\delta$ by (3.13), T can be written as

$$T = \frac{An}{\alpha\delta}. \tag{3.36}$$

From (3.11),

$$\frac{1}{\delta} = \frac{2}{\pi}E\left(\frac{1}{C}\right). \tag{3.37}$$

In (3.11), C was the distance that a fish had to swim to cross a stationary beam. This formula is still valid for our purpose, but now C is the distance that the boat must travel for the beam to pass over the fish.

Now let

$$\begin{aligned} \nu &= \text{the pulse rate of the transducer,} \\ e &= \text{the number of echoes received from a fish, and,} \\ e_r &= 1/e. \end{aligned}$$

Then, according to Crittenden *et al.* [12],

$$e = \frac{C\nu}{u}. \tag{3.38}$$

Using equations (3.36), (3.37) and (3.38), T can be written in term of estimable quantities as

$$T = \frac{2A\nu}{\pi u}\frac{n}{\alpha}E(e_r). \tag{3.39}$$

Extension to Unequal Target Strengths

Now suppose that the target strength of a fish is a random variable K. For simplicity, assume that K is discrete that takes values in a countable set. Let $n(k)$ and $v(k)$ refer to the corresponding quantities above, but for fish of target strength $K = k$. Then,

$$n = \sum_k n(k) \tag{3.40}$$

and the probability mass function $p(K = k)$ of fish of target strength k that are ensonified in the beam is

$$p(K = k) = \frac{n(k)}{n}. \tag{3.41}$$

Let $E(e_r|K = k)$ be the conditional expectation of e_r given $K = k$. Then,

$$E(e_r) = E\left[E(e_r|K = k)\right] = \sum_k E(e_r|K = k)\frac{n(k)}{n}, \tag{3.42}$$

from which we get

$$\sum_k E(e_r|K=k)n(k) = nE(e_r). \tag{3.43}$$

Furthermore, according to (3.39),

$$T(k) = \frac{2A\nu}{\pi\alpha u} n(k)E(e_r|K=k). \tag{3.44}$$

By summing (3.44) over k, we obtain the total population size under the new assumption that target strength is a random variable. Denote this total by T_1. Then,

$$\begin{aligned} T_1 &= \sum_k T(k) = \frac{2A\nu}{\pi\alpha u}\sum_k n(k)E(e_r|K=k) & (3.45)\\ &= \frac{2A\nu}{\pi\alpha u} nE(e_r), & (3.46)\end{aligned}$$

which is again equal to T given by (3.39). This implies that the population size can be written in terms of estimable quantities which are independent of the target strength distribution.

Estimation of T

Suppose that l transects are randomly sampled. For each transect, the number of fish ensonified is counted and the length of the transect is recorded. The number of fish per unit distance is then calculated. Also, the number of echoes per fish is determined for all fish observed or a random sample of them. Let

$\bar{N}$ = average number of fish observed per unit distance and

$\bar{e}_r$ = average reciprocal number of echoes per fish.

The total number of fish in stratum, T can be estimated by

$$\hat{T} = \frac{2A\nu}{\pi u}\bar{N}\bar{e}_r. \tag{3.47}$$

The total number of fish in the survey area can be estimated by adding the stratum estimates over the depth strata.

Note:

For a paper recorder, we can write e_r in terms of the paper speed of the recorder and the mean reciprocal target widths. Denoting the parer speed by p and the target widths by W, it is easy to see that $W = p(e/\nu)$. Hence, $e_r = 1/e = p/(\nu W) = (p/\nu)M$, where M is the reciprocal target width. Therefore, (3.39) can also be written as

$$T = \frac{2Ap}{\pi u} nE(M). \tag{3.48}$$

Consequently, T can be estimated by

$$\hat{T} = \frac{2Ap}{\pi u} \bar{N}\bar{M}, \tag{3.49}$$

where $\bar{M}$ is the sample average of reciprocal target widths. Note that this estimate has the same form as (3.22) except for the multiplicative constant.

3.2.2 Estimation of Variance

It is reasonable to assume that $\bar{N}$ and $\bar{M}$ for a given depth stratum are independent. Also, the total estimates in different strata are uncorrelated. Hence, similarly to (3.33), the variance of the total estimator can be estimated by

$$\hat{V}ar(\hat{T}) = \left[\frac{2Ap}{\pi u}\right]^2 \sum_i \left(\frac{s_{N_i}^2}{l}\bar{M}_i^2 + \frac{s_{M_i}^2}{m_i}\bar{N}_i^2 - \frac{s_{N_i}^2 s_{M_i}^2}{l m_i}\right). \tag{3.50}$$

3.3 Discussion

The described method for estimating the number of migrating fish in a river using hydroacoustic techniques has two major advantages over similar methods used in lakes and reservoirs. First, the method is not sensitive to beam shape or dimensions. Second, the exact calculation of the expected reciprocal of the chord length allows the user to avoid the use of the bootstrap to reduce bias caused by using an estimation formula with a random variable in its denominator.

We identified several potential sources of bias inherent in the method. The first, which stems from the speed of the boat being only moderately faster than fish migration speed, tends to result in overestimation of abundance. This bias can be reduced by increasing the boat's speed as much as practically possible, or by estimating a bias adjustment factor that can be extracted from formula (27).

Second, fish swimming off the upstream direction will introduce a positive bias. The stationary soundings estimate overall fish speed, and the upstream component to the velocity will be overestimated. This source of bias can sometimes be reduced by choosing a location, such as the Mission site, where there are limited opportunities for the fish to mill about.

Contrasting with these sources of positive bias are sources of negative bias. Fish that travel in areas of the river that cannot be acoustically sampled, such as close to the surface above the minimum detection depth and close to shore where the echo-sounding boat cannot travel, will be missed. Similarly, fish that sense and avoid the echo-sounding boat will lead to a negative bias. Finally, when the targets of two or more fish overlap on the echogram, so that only one target is apparent, the population will be underestimated. These biases are not considered to be significant for sockeye salmon in the Fraser River.

In determining depth strata, we have taken the distances determined by the echosounder as actual depths. This can also introduce a bias. This bias can be reduced by reducing the beam angle. Ways of reducing bias are also being investigated through the application of new technology (e.g., split beam).

The estimated variance for the worked example could possibly be reduced further if the day was divided into smaller units of time. This is because a daily cyclical pattern in the abundance of migrating sockeye salmon is ignored in our example. Stratifying each day such that the strata are homogeneous with respect to migration (e.g., into hourly strata) could result in more precise estimates. Daily abundance and variance can then be estimated as the sum of the stratum estimates.

The fish speed calculated through target widths obtained from stationary soundings plays an important role in the estimation. Fish speed can vary substantially over time and space (depth and location in the river). Therefore, the timing and location of stationary soundings should be re-examined, and alterations to the sampling design should be considered to address concerns about possible biases.

In the example shown, the 264 targets sampled in the stationary soundings provided estimates with reasonable precision. On days with fewer fish in the river, the estimates will have a larger coefficient of variation. A scheme for optimal allocation of sampling effort between the two types of soundings could substantially reduce the coefficient of variation, particularly on such days.

The estimators are sums of products of averages, and may have skewed distributions. Therefore, the usual confidence intervals, calculated by adding and subtracting standard errors times normal quantiles may not provide an accurate coverage. In Chapter 4, we investigate two alternative strategies for this purpose.

The method described in this chapter has a strong potential to give accurate and precise estimates of fish passage that are robust and comparatively easy to obtain. Methods for promoting and checking the validity of the underlying assumptions are discussed. The method provides a significant tool for managers under pressure to make timely and effective management decisions.

Chapter 4

Confidence Intervals for a Sum of Pairwise Products of Means

The purpose of this chapter is to develop methods to find confidence intervals for a sum of products of means. The motivation arose from the requirement to find confidence intervals for the migration rate estimated in Section 3.1 and the number of fish in a resident population in Section 3.2. Both estimators are sums of products of pairwise means. Products of means also arise in other practical situations. As Berger*et al.* [5] pointed out, they arise in determining area based on measurements of length and width. Furthermore, in gypsy moth studies, the hatching rate of larvae per unit area can be estimated as a product of the mean number of egg masses per unit area times the mean number of larvae hatching area per egg mass. Applications also arise in forest sampling (A. Ellingsen, personal communication). In the fish applications described here, the independence of the individual means follows naturally from the assumptions about the sampling procedure. In other applications also approximately independent samples can be obtained for each mean (Southwood [37]).

Berger *et al.* [5] considered the problem of obtaining confidence intervals for a product of two means from independent normal distributions with known variances when only one observation from each distribution is available. They proposed a Bayesian approach with reference priors. We consider the problem assuming that large samples are available from each distribution, and explore two methods that are derived from the likelihood ratio test and the Wald test. In all the applications described above, the variables involved are positive. For this reason, in our development we assume that the variables are positive. Possible modifications for other cases are also suggested.

In the area-estimation example, both variables, the length and the width, may be well-approximated by normal distributions. In other applications, one or both variables can be far from normal. In the fish problem, one variable is the number of fish detected per transect. The other variable in this case is the reciprocal target width of a fish on an echogram. The latter is proportional to fish speed. Figure 4.4 shows histograms of these variables based on data collected by the Pacific Salmon Commission on August 31, 1993 at Mission, B.C.. According to these histograms, the variables seem to have positively skewed distributions. However, when the samples are large, sample averages are approximately normally distributed. Hence, the methods used to construct confidence intervals for sum of products of normal means may be adapted in these situations as well.

So, we first develop methods for finding confidence intervals for a sum of pairwise products of means from independent normal distributions. These methods are based on the standard, general likelihood ratio and Wald tests. The methods are then extended to other distributions using the central limit theorem. In the case of normal means, Bartlett-type adjustments are derived to improve the confidence intervals obtained through the Wald test.

4.1 Notation and Assumptions

4.1.1 Notation

Let p be the number of pairs of products in the sum. When we have only one pair of products ($p = 1$), we use the following notation:

$$\begin{aligned}
j &= \begin{cases} 1 & \text{for the first component of the pair} \\ 2 & \text{for the second component of the pair} \end{cases} \\
\mu_j &= \text{the mean of the } j\text{th distribution,} \\
\sigma_j^2 &= \text{the variance of the } j\text{th distribution,} \\
n_j &= \text{size of the sample from the } j\text{th distribution,} \\
\mathbf{x}_j &= \text{the sample from the } j\text{th distribution} = \begin{pmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jn_j} \end{pmatrix}, \\
\bar{x}_j &= \text{the sample average from the } j\text{th distribution,} \\
s_j^2 &= \frac{1}{n_j}\sum_{k=1}^{n_j}(x_{jk} - \bar{x}_j)^2, \\
\tilde{s}_j^2 &= \frac{1}{(n_j - 1)}\sum_{k=1}^{n_j}(x_{jk} - \bar{x}_j)^2, \\
u_j &= \sigma_j^2/n_j, \\
v_j &= s_j^2/n_j, \text{ and} \\
\tilde{v}_j &= \tilde{s}_j^2/n_j.
\end{aligned}$$

To further simplify notation, also let

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{u} \end{pmatrix},$$

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \text{ and } \tilde{\mathbf{v}} = \begin{pmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{pmatrix}.$$

When $p > 1$, we attach an additional subscript to denote the pair. For example, for $i = 1, 2, \ldots, p$,

$$
\begin{aligned}
\mu_{ij} &= \text{the mean of the } j\text{th distribution of the } i\text{th pair,} \\
\sigma_{ij}^2 &= \text{the variance of the } j\text{th distribution of the } i\text{th pair,} \\
n_{ij} &= \text{size of the sample from the } j\text{th distribution of the } i\text{th pair, and} \\
\bar{x}_{ij} &= \text{the sample average from the } j\text{th distribution of the } i\text{th pair, etc..}
\end{aligned}
$$

In addition, let

$$
\begin{aligned}
n_0 &= \text{the smallest sample size} \\
\rho &= \sum_{i=1}^{p} \mu_{i1}\mu_{i2}, \\
S &= \text{-2 times the maximum log-likelihood ratio [as defined by (4.2)], and} \\
W &= \text{the Wald statistic [as defined by (4.30)] .}
\end{aligned}
$$

4.1.2 Assumptions

The only main assumptions we carry throughout the discussion are,

1. all the variables involved are independent of each other, and
2. all the observations within a sample are independently, identically distributed.

4.2 Normally Distributed Data

4.2.1 Likelihood Ratio Approach

One approach is to construct approximate confidence intervals by inverting the likelihood ratio test. Consider the null hypothesis, $H_0 : \rho = \rho_0$. The likelihood ratio for testing this hypothesis is

$$LR(\rho_0) = \frac{\sup_{\{\boldsymbol{\mu}|\rho=\rho_0\}} l(\boldsymbol{\theta};\mathbf{x})}{\sup_{\{\boldsymbol{\mu}\}} l(\boldsymbol{\theta};\mathbf{x})}, \tag{4.1}$$

where $l(\boldsymbol{\theta};\mathbf{x})$ is the likelihood function. According to Wilks [41], -2 times the log-likelihood ratio,

$$S(\rho_0) = -2\ln LR(\rho_0) \tag{4.2}$$

has a limiting chi-square distribution under H_0. The method proceeds by determining the likelihood ratio test of the null hypothesis for a specified value of ρ_0, and then inverting this test in the usual way to obtain a confidence interval for ρ as those values which do not lead to the rejection of the null hypothesis.

Product of Two Normal Means: Variances Known

Let us first consider the case of two independent normal means with known variances. The likelihood ratio for testing $H_0 : \rho = \rho_0$ is

$$LR(\rho_0) = \frac{\sup_{\{\boldsymbol{\mu}|\mu_1\mu_2=\rho_0\}} l(\boldsymbol{\mu};\mathbf{x})}{\sup_{\{\boldsymbol{\mu}\}} l(\boldsymbol{\mu};\mathbf{x})}, \tag{4.3}$$

where

$$\begin{aligned} l(\boldsymbol{\mu};\mathbf{x}) &= \frac{1}{[2\pi]^{\frac{(n_1+n_2)}{2}}\sigma_1^{n_1}\sigma_2^{n_2}} \exp\left[-\frac{1}{2\sigma_1^2}\sum_{j=1}^{n_1}(x_{1j}-\mu_1)^2 - \frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}(x_{2j}-\mu_2)^2\right] \\ &= \frac{1}{[2\pi]^{\frac{(n_1+n_2)}{2}}\sigma_1^{n_1}\sigma_2^{n_2}} \exp\left[\frac{-s_1^2-(\bar{x}_1-\mu_1)^2}{2\,u_1} + \frac{-s_2^2-(\bar{x}_2-\mu_2)^2}{2\,u_2}\right] \end{aligned} \tag{4.4}$$

is the likelihood function. It is easy to see that the logarithm of the denominator of $LR(\rho_0)$ is

$$L_1 = -\frac{(n_1+n_2)}{2}\ln(2\pi) - n_1\ln(\sigma_1) - n_2\ln(\sigma_2) - \frac{s_1^2}{2u_1} - \frac{s_2^2}{2u_2}. \tag{4.5}$$

To determine the numerator of $LR(\rho_0)$, we need to find the maximum of $l(\boldsymbol{\mu};\mathbf{x})$ subject to the constraint $\mu_1\mu_2 - \rho_0 = 0$. This is equivalent to finding the maximum of $\ln l(\boldsymbol{\mu};\mathbf{x})$ subject to $\mu_1\mu_2 - \rho_0 = 0$. We use the Lagrangian multiplier method to find the restricted maximum likelihood estimates. Let

$$\begin{aligned} h(\boldsymbol{\mu}, \lambda) &= \ln l(\boldsymbol{\mu}; \mathbf{x}) - \lambda(\mu_1\mu_2 - \rho_0) \\ &= L_1 - \frac{(\bar{x}_1 - \mu_1)^2}{2\,u_1} - \frac{(\bar{x}_2 - \mu_2)^2}{2\,u_2} - \lambda(\mu_1\mu_2 - \rho_0), \end{aligned} \tag{4.6}$$

where λ is a Lagrange multiplier. The critical point of $h(\boldsymbol{\mu}, \lambda)$ is found by solving the equations,

$$\begin{aligned} \frac{\partial h}{\partial \mu_1} &= \frac{\bar{x}_1 - \mu_1}{u_1} - \lambda\,\mu_2 = 0, \quad \text{and} \\ \frac{\partial h}{\partial \mu_2} &= \frac{\bar{x}_2 - \mu_2}{u_2} - \lambda\,\mu_1 = 0. \end{aligned} \tag{4.7}$$

The critical point thus obtained is

$$\hat{\boldsymbol{\mu}}(\lambda) = \begin{bmatrix} \hat{\mu}_1(\lambda) \\ \hat{\mu}_2(\lambda) \end{bmatrix}, \tag{4.8}$$

where

$$\hat{\mu}_1(\lambda) = \frac{\bar{x}_1 - \lambda u_1 \bar{x}_2}{1 - \lambda^2 u_1 u_2}, \quad \text{and} \tag{4.9}$$

$$\hat{\mu}_2(\lambda) = \frac{\bar{x}_2 - \lambda u_2 \bar{x}_1}{1 - \lambda^2 u_1 u_2}, \quad \text{for} \quad \lambda \neq \sqrt{\frac{1}{u_1 u_2}}, \tag{4.10}$$

such that λ satisfies the constraint equation,

$$\rho(\lambda) = \left(\frac{\bar{x}_1 - \lambda u_1 \bar{x}_2}{1 - \lambda^2 u_1 u_2}\right)\left(\frac{\bar{x}_2 - \lambda u_2 \bar{x}_1}{1 - \lambda^2 u_1 u_2}\right) = \rho_0. \tag{4.11}$$

Note that $\hat{\boldsymbol{\mu}}(0)$ is the unrestricted maximum likelihood estimate. The constraint (4.11) reduces to a fourth degree polynomial and gives up to four solutions for λ. Since we assume that both variables are positive, $\hat{\mu}_1$ and $\hat{\mu}_2$ should also be positive. Therefore, some solutions can be eliminated. From the other solutions, we should choose the one corresponding to the maximum of $h(\lambda)$. Then, substituting these values in (4.2), $S(\rho_0)$ can be calculated. If $S(\rho_0)$ is less than or equal to $\chi^2_{1,(1-\alpha)}$, then ρ_0 is included in the confidence interval. The procedure is systematically repeated for different ρ_0 to construct the confidence interval.

However, this can be tedious because we have to solve (4.11) for each ρ_0. Madansky [25] introduced a method which simplifies these computations. This method avoids the need to solve (4.11) for different values of ρ_0. Instead, it obtains the restricted maximum likelihood estimates in terms of the Lagrangian multiplier, λ [as in (4.9) and (4.9)]. Then, substituting these in (4.2), S is obtained in terms of λ. Let this be denoted by $S(\lambda)$. Madansky's method solves the equation,

$$S(\lambda) - \chi^2_{1,(1-\alpha)} = 0, \tag{4.12}$$

for λ. Usually, this equation has two real roots, and can be solved easily if $S(\lambda)$ behaves well around the unrestricted maximum likelihood estimate. The roots are then substituted in the constraint function (4.11) to get confidence limits for $\rho = \mu_1\mu_2$. The following result due to Madansky [25], provides a necessary and sufficient condition for $S(\lambda)$ to have such a behavior around zero (zero is the value of λ that corresponds to the unrestricted maximum likelihood estimate of ρ).

Theorem 1 (Madansky) *-2 times the logarithm of the likelihood ratio is a monotonically decreasing (increasing) function of λ for $\lambda < 0$ $(\lambda > 0)$ if and only if the constraint function, as a function of λ, is monotonically decreasing.*

When the likelihood ratio and the constraint function exhibit such behavior, it is easy to invert the likelihood ratio test to construct confidence intervals as described above. We now investigate the possibility of using Madansky's method for our problem. Figure 4.1 shows an example of the behavior of $\rho(\lambda)$ and $S(\lambda)$ when $\bar{x}_1\bar{x}_2 > 0$. This example was constructed using $n_1 = 30, n_2 = 30, \bar{x}_1 = 10, \bar{x}_2 = 20, s_1 = 3$ and $s_2 = 5$. Note that $\rho(\lambda)$ is positive and monotonically decreasing in an interval containing 0, and in this interval, $S(\lambda)$ is convex.

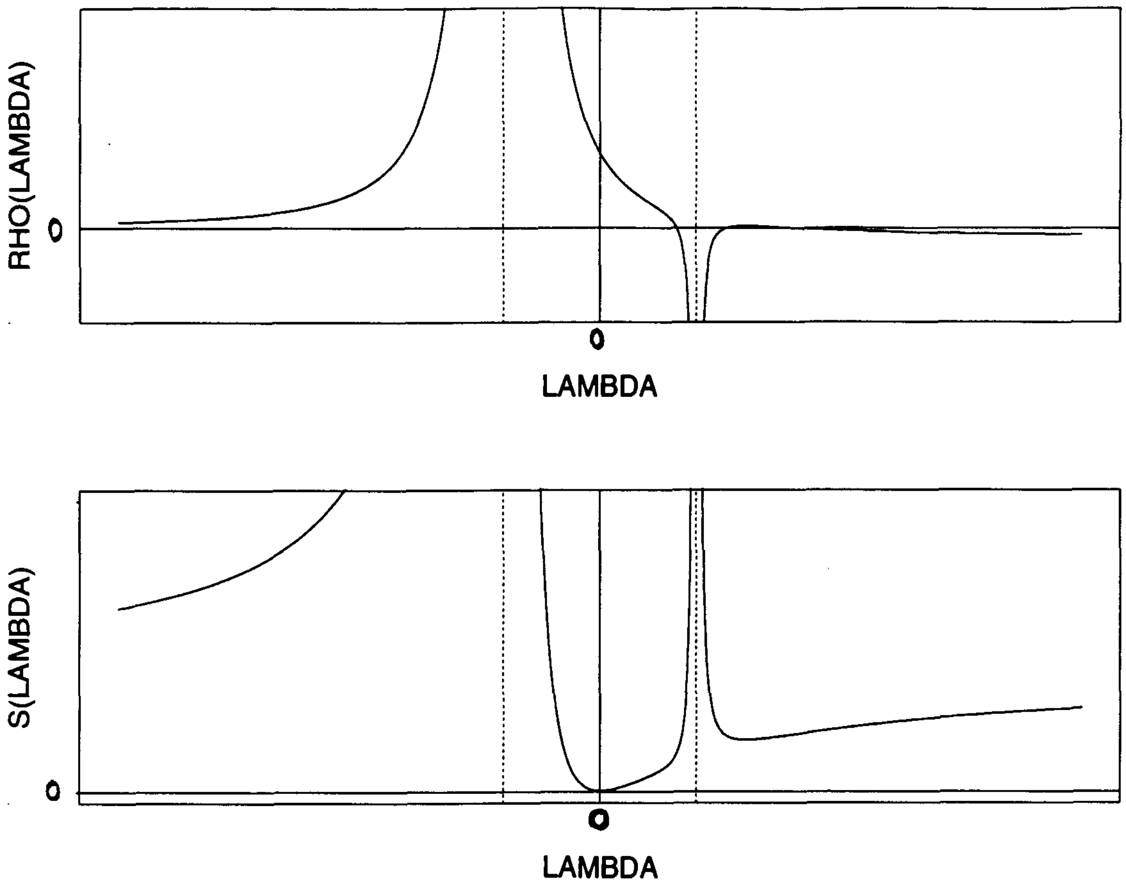


Figure 4.1: Behavior of $\rho(\lambda)$ and $S(\lambda)$ when $\bar{x}_1\bar{x}_2 > 0$ (example calculation using $n_1 = 30, n_2 = 30, \bar{x}_1 = 10, \bar{x}_2 = 20, s_1 = 3$ and $s_2 = 5$).

The following result identifies this interval in general, and shows that (4.11) has a unique solution λ_0 in this interval such that $\hat{\mu}_1(\lambda_0)$ and $\hat{\mu}_2(\lambda_0)$ are positive. Further, it shows that $\hat{\mu}(\lambda_0)$ is in fact a local maximum of l subject to the constraint, $\hat{\mu}_1(\lambda)\hat{\mu}_2(\lambda) - \rho_0 = 0$. So, the unique local maximum of the likelihood subject to the constraint can be easily found. In other words, this result enables us to pick up the unique set of local maxima that gives us a continuous relationship between $\hat{\mu}$ and ρ for ρ near ρ_0.

Result 4.1 *Let*

$$\begin{aligned}\gamma_1 &= \frac{\bar{x}_1}{u_1\bar{x}_2},\\ \gamma_2 &= \frac{\bar{x}_2}{u_2\bar{x}_1},\\ \beta_1 &= \sqrt{\gamma_1\gamma_2} = \sqrt{\frac{1}{u_1u_2}},\\ \beta_2 &= \min(\gamma_1,\gamma_2),\\ I_\lambda &= (-\beta_1,\beta_1), \text{ and}\\ J_\lambda &= (-\beta_1,\beta_2].\end{aligned}$$

Suppose that $\bar{x}_1 > 0$, *and* $\bar{x}_2 > 0$. *Then,*

(a) $\rho(\lambda)$ *is monotonically decreasing in* I_λ, *and* $S(\lambda)$ *is monotonically decreasing (increasing) for* $\lambda < 0$ $(\lambda > 0)$ *in* I_λ.

(b) $\hat{\mu}$ *is positive in* J_λ *and (4.11) has a unique solution* λ_0 *in* J_λ.

(c) $\mu = \hat{\mu}(\lambda_0)$ *is a local maximum of* $l(\mu;\mathbf{x})$ *subject to the constraint* $\mu_1\mu_2 - \rho_0 = 0$.

Proof

(a) Note that

$$\rho(\lambda) = \left(\frac{\bar{x}_1 - \lambda u_1\bar{x}_2}{1-\lambda^2u_1u_2}\right)\left(\frac{\bar{x}_2 - \lambda u_2\bar{x}_1}{1-\lambda^2u_1u_2}\right) = \frac{\bar{x}_1\bar{x}_2}{u_1u_2}\frac{(\lambda-\gamma_1)(\lambda-\gamma_2)}{\left(\lambda^2-{\beta_1}^2\right)^2}.$$

Let

$$g(\lambda) = \frac{(\lambda-\gamma_1)(\lambda-\gamma_2)}{\left(\lambda^2-{\beta_1}^2\right)^2}.$$

To prove part (a), it is sufficient to show that the derivative of g with respect to λ is negative in I_λ. This derivative is

$$\frac{dg}{d\lambda} = \frac{\lambda-\gamma_2}{\left(\lambda^2-{\beta_1}^2\right)^2} + \frac{\lambda-\gamma_1}{\left(\lambda^2-{\beta_1}^2\right)^2} - \frac{4(\lambda-\gamma_1)(\lambda-\gamma_2)\lambda}{\left(\lambda^2-{\beta_1}^2\right)^3}. \tag{4.13}$$

Now suppose without loss of generality that $\gamma_1 < \gamma_2$. Then it is clear that $\frac{dg}{d\lambda} < 0$ for $\lambda \in (-\beta_1, 0)$. To show that $\frac{dg}{d\lambda}$ is negative for $\lambda \in [0, \beta_1)$, we simplify (4.13) obtaining

$$\frac{dg}{d\lambda} = \frac{-2\lambda^3 + 3(\gamma_1 + \gamma_2)\lambda^2 - 6\gamma_1\gamma_2\lambda + \gamma_1\gamma_2(\gamma_1 + \gamma_2)}{\left(\lambda^2 - \beta_1{}^2\right)^3}. \tag{4.14}$$

Since the denominator of (4.14) is negative for $\lambda \in [0, \beta_1)$, we need to show that the numerator,

$$\eta(\lambda) = -2\lambda^3 + 3(\gamma_1 + \gamma_2)\lambda^2 - 6\gamma_1\gamma_2\lambda + \gamma_1\gamma_2(\gamma_1 + \gamma_2) \tag{4.15}$$

of (4.14) is positive in this interval. We show this by negating a contradiction. Suppose that $\eta(\lambda)$ is non-positive for some $\lambda \in [0, \beta_1)$. Now, note that $\eta(\lambda)$ is differentiable in $[0, \beta_1)$. Also note that

$$\eta(0) = \gamma_1\gamma_2(\gamma_1 + \gamma_2) > 0, \text{ and} \tag{4.16}$$

$$\eta(\beta_1) = 4\gamma_1\gamma_2\left[(\gamma_1 + \gamma_2) - 2\sqrt{\gamma_1\gamma_2}\right] > 0 \tag{4.17}$$

So, $\eta(\lambda)$ must have a non-positive minimum in $[0, \beta_1)$. In order to examine the critical points of $\eta(\lambda)$, consider the derivative

$$\frac{d\eta(\lambda)}{d\lambda} = -6\,(\lambda - \gamma_2)(\lambda - \gamma_1).$$

The derivative is zero at γ_1 and γ_2. But, $\eta(\gamma_1) = \gamma_1\,(\gamma_1 - \gamma_2)^2 > 0$, and $\beta_1 < \gamma_2$ (since $\gamma_1 < \sqrt{\gamma_1\gamma_2} < \gamma_2$). Therefore, $\eta(\lambda)$ does not have a non-positive minimum in $[0, \beta_1)$. This proves that $\eta(\lambda)$ is positive in $[0, \beta_1)$ and conseqently that $\rho(\lambda)$ is monotonically decreasing in I_λ. Then, by Madansky's theorem, $S(\lambda)$ is monotonically decreasing (increasing) for $\lambda < 0$ $(\lambda > 0)$ in I_λ.

(b) The denominators in (4.9) and (4.10) are positive as long as $|\lambda| < \sqrt{(1/u_1u_2)}$, and numerators are positive as long as $\lambda < \min(\gamma_1, \gamma_2)$. This implies that $\hat{\boldsymbol{\mu}}$ is positive in J_λ.

Now notice that $\rho(\lambda)$ approaches infinity as λ approaches $-\beta_1$ and $\rho(\beta_2) = 0$. Therefore, for any $\rho_0 > 0$, (4.11) has a unique real root λ_0 in J_λ.

(c) We have already established that $\hat{\boldsymbol{\mu}}(\lambda_0)$ is a critical point of h. According to Marsden and Tromba [26](pg. 274), in order to prove that $\hat{\boldsymbol{\mu}}(\lambda_0)$ is a maximum, it is enough to show that the value of the bordered Hessian determinant of h (see Appendix), when evaluated at $\hat{\boldsymbol{\mu}}(\lambda_0)$ is positive. The bordered Hessian determinant of h is

$$BH(\boldsymbol{\mu},\lambda) = -\mu_1\left(\lambda\,\mu_2 - \frac{\mu_1}{u_1}\right) - \mu_2\left(\lambda\,\mu_1 - \frac{\mu_2}{u_2}\right). \tag{4.18}$$

Let $BH_0 = BH(\hat{\boldsymbol{\mu}},\lambda_0)$. Then, using (4.9) and (4.10),

$$\begin{aligned} BH_0 &= -\frac{[2u_1^2u_2^2\bar{x}_1\bar{x}_2\lambda_0^3 - 3u_1u_2(u_2\bar{x}_1^2 + u_1\bar{x}_2^2)\lambda_0^2 + 6u_1u_2\bar{x}_1\bar{x}_2\lambda_0 - (u_2\bar{x}_1^2 + u_1\bar{x}_2^2)]}{(u_1\,u_2\lambda_0^2 - 1)^2\,u_1\,u_2} \\ &= \frac{u_1u_2\bar{x}_1\bar{x}_2\,[-2\,\lambda_0^3 + 3(\,\gamma_1 + \gamma_2\,)\lambda_0^2 - 6\gamma_1\gamma_2\lambda_0 + \gamma_1\gamma_2(\gamma_1 + \gamma_2)]}{(u_1\,u_2\lambda_0^2 - 1)^2} \\ &= \frac{u_1u_2\bar{x}_1\bar{x}_2\eta(\lambda_0)}{(u_1\,u_2\lambda_0^2 - 1)^2} \end{aligned}$$

Since $\bar{\mathbf{x}}, \mathbf{u}$ and $\eta(\lambda)$ are positive in J_λ (which is a subset of I_λ), $BH_0 > 0$ in J_λ. This proves the result.

According to Result 4.1, estimates (4.9) and (4.10) do indeed specify restricted maxima. Therefore, substituting these values in (4.4), the logarithm of the numerator of $l(\rho_0)$ can be written as

$$\begin{aligned} L_0 &= L_1 - \frac{1}{2}\left(\frac{\hat{\mu}_1(\lambda) - \bar{x}_1}{\sqrt{u_1}}\right)^2 - \frac{1}{2}\left(\frac{\hat{\mu}_2(\lambda) - \bar{x}_2}{\sqrt{u_2}}\right)^2 \\ &= L_1 - \left[\lambda^2u_2\left(\frac{\bar{x}_1 - \lambda u_1\bar{x}_2}{1 - \lambda^2u_1u_2}\right)^2 + \lambda^2u_1\left(\frac{\bar{x}_2 - \lambda u_2\bar{x}_1}{1 - \lambda^2u_1u_2}\right)^2\right] \\ &= L_1 - \left[\lambda^2u_2\hat{\mu}_1^2(\lambda) + \lambda^2u_1\hat{\mu}_2^2(\lambda)\right]. \end{aligned} \tag{4.19}$$

Hence, twice the log-likelihood ratio is

$$\begin{aligned} S(\lambda) &= -2L_0 - (L_1) \\ &= \lambda^2u_2\left(\frac{\bar{x}_1 - \lambda u_1\bar{x}_2}{1 - \lambda^2u_1u_2}\right)^2 + \lambda^2u_1\left(\frac{\bar{x}_2 - \lambda u_2\bar{x}_1}{1 - \lambda^2u_1u_2}\right)^2. \end{aligned} \tag{4.20}$$

Result 4.1 confirms that $S(\lambda)$ monotonically decreases (increases) for $\lambda < 0$ $(\lambda > 0)$ in I_λ while $\rho(\lambda)$ monotonically decreases in the interval. Therefore, S can be easily inverted to calculate an approximate $(1-\alpha)100\%$ confidence interval for ρ. Note that $S(0) = 0$ and $S(\lambda)$ approaches infinity when λ approaches β_1. Since $S(\lambda)$ is continuous in J_λ, there exist a $\lambda_1^*(< 0)$ such that $S(\lambda_1^*) = \chi^2_{1,(1-\alpha)}$. If $S(\beta_2) \geq \chi^2_\alpha(1)$, then there also exist a $\lambda_2^*(> 0)$ such that $S(\lambda_2^*) = \chi^2_{1,(1-\alpha)}$. Then the confidence interval is given by $[\rho(\lambda_2^*), \rho(\lambda_1^*)]$. If $S(\beta_2) < \chi^2_{1,(1-\alpha)}$, there does not exist a $\lambda_2^*(> 0)$ such that $S(\lambda_2^*) = \chi^2_{1,(1-\alpha)}$. Then the natural lower bound $\rho(\beta_2) = 0$ serves as the lower limit of the interval.

Remark

We have not imposed any sign restrictions on the means in the maximization procedure. Hence, the method can be used regardless of the sign of the averages. By an argument similar to that in the proof of Result 4.1, one can show that $\rho(\lambda)$ is monotonically decreasing in I_λ for all cases. Then, by Madansky's theorem, $S(\lambda)$ behaves well in I_λ. So, S can be inverted easily to construct the confidence intervals.

Sum of p Products of Pairwise Normal Means: Variances Known

Now we extend the method to construct confidence intervals for the sum of products,

$$\rho = \sum_{1=1}^{p} \mu_{i1}\mu_{i2}. \tag{4.21}$$

Here, $(\mu_{i1}\ , \mu_{i2})$ is the pair of means of the ith product. By direct extension of Result 4.1, it can be proved that the following result is true.

Result 4.2 *Let*

$$\begin{aligned}
u_{i1} &= \frac{s_{i1}^2}{n_{i1}}, \\
u_{i2} &= \frac{s_{i2}^2}{n_{i2}}, \\
\gamma_{i1} &= \frac{\bar{x}_{i1}}{u_{i1}\bar{x}_{i2}}, \\
\gamma_{i2} &= \frac{\bar{x}_{i2}}{u_{i2}\bar{x}_{i1}}, \\
\beta_1 &= \min_{i=1,..,p} [\sqrt{\gamma_{i1}\gamma_{i2}}] = \min_{i=1,..,p} \left[\sqrt{\frac{1}{u_{i1}u_{i2}}}\right], \\
I_\lambda &= (-\beta_1, \beta_1) \ \textit{and}, \\
S(\lambda) &= \sum_{i=1}^{p} \left[\lambda^2 u_{i2}\left(\frac{\bar{x}_{i1} - \lambda u_{i1}\bar{x}_{i2}}{1-\lambda^2 u_{i1}u_{i2}}\right)^2 + \lambda^2 u_{i1}\left(\frac{\bar{x}_{i2} - \lambda u_{i2}\bar{x}_{i1}}{1-\lambda^2 u_{i1}u_{i2}}\right)^2\right]
\end{aligned} \tag{4.22}$$

such that λ *satisfies*

$$\rho(\lambda) = \sum_{i=1}^{p} \left[\left(\frac{\bar{x}_{i1} - \lambda u_{i1} \bar{x}_{i2}}{1 - \lambda^2 u_{i1} u_{i2}} \right) \left(\frac{\bar{x}_{i2} - \lambda u_{i2} \bar{x}_{i1}}{1 - \lambda^2 u_{i1} u_{i2}} \right) \right] = \rho_0. \tag{4.23}$$

Then,

1. $\rho(\lambda)$ *is monotonically decreasing in* I_λ*, and* $S(\lambda)$ *is a monotonically decreasing (increasing) for* $\lambda < 0 \quad (\lambda > 0)$ *in* I_λ*.*
2. *Under the null hypothesis* $H_0 : \sum_{i=1}^{k} \mu_{i1}\mu_{i2} = \rho_0$*, the* S *is asymptotically distributed as* χ_1^2*.*

Therefore, an approximate $(1-\alpha)100\%$ level confidence interval can be calculated using Madansky's approach.

Product of Two Normal Means: Variances Unknown

Usually, in practical situations, the variances are unknown. We now develop likelihood ratio methods for the case of normal data with unknown variances.

The likelihood function (4.4) should now be viewed as a function of both $\boldsymbol{\mu}$ and $\mathbf{u}$. The log-likelihood function is

$$\begin{aligned} \ln l(\boldsymbol{\mu}, \mathbf{u}; \mathbf{x}) \; = \; & -\frac{(n_1+n_2)\ln(2\,\pi)}{2} - \frac{n_1 \ln(n_1 u_1)}{2} - \frac{n_2 \ln(n_2 u_2)}{2} \\ & - \frac{\left[s_1^2 + (\bar{x}_1 - \mu_1^2)^2\right]}{2\,u_1} - \frac{\left[s_2^2 + (\bar{x}_2 - \mu_2^2)^2\right]}{2\,u_2}. \end{aligned}$$

Method 1

It is very easy to find the unrestricted maximum likelihood. But it is not so easy to find the maximum subject to the constraint $\mu_1\mu_2 - \rho_0 = 0$. The critical points of the Lagrangian $h = \ln l(\boldsymbol{\mu}, \mathbf{u}; \mathbf{x}) - \lambda\,(\mu_1\mu_2 - \rho_0)$ can be found by solving the system of equations,

$$\frac{\partial h}{\partial \mu_1} = \frac{\bar{x}_1 - \mu_1}{u_1} - \lambda\,\mu_2 \;=\; 0,$$

$$\begin{aligned}
\frac{\partial h}{\partial \mu_2} = \frac{\bar{x}_2 - \mu_2}{u_2} - \lambda\,\mu_1 &= 0, \\
\frac{\partial h}{\partial u_1} = -\frac{n_1}{2\,u_1} + \frac{s_1^2 + (\bar{x}_1 - \mu_1)^2}{2\,u_1^2} &= 0, \text{ and} \\
\frac{\partial h}{\partial u_2} = -\frac{n_2}{2\,u_2} + \frac{s_2^2 + (\bar{x}_2 - \mu_2)^2}{2\,u_2^2} &= 0.
\end{aligned}$$

The Maple V solution of this system of equations can be simplified as follows:

$$\hat{\mu}_2(\lambda) = \text{ a root of the equation, } c_5\mu_2^5 + c_4\mu_2^4 + c_3\mu_2^3 + c_2\mu_2^2 + c_1\mu_2 + c_0 = 0, \tag{4.24}$$

where

$$\begin{aligned}
c_0 &= -\left({s_2}^2 + {x_2}^2\right) n_1 \left[-x_2\,n_2 + x_1\left({s_2}^2 + {x_2}^2\right)\lambda\right], \\
c_1 &= c_{12}\lambda^2 + c_{11}\lambda + c_{10}, \\
\text{with } c_{12} &= \left({s_2}^2 + \bar{x}_2^2\right)^2 \left({s_1}^2 + \bar{x}_1^2\right), \\
c_{11} &= 2\,\bar{x}_1\,\bar{x}_2 \left({s_2}^2 + \bar{x}_2^2\right)(2\,n_1 - n_2), \\
c_{10} &= -n_2\left(-\bar{x}_2^2 n_2 + n_1\,{s_2}^2 + 3\,n_1\,\bar{x}_2^2\right), \\
c_2 &= c_{22}\lambda^2 + c_{21}\lambda + c_{20}, \\
\text{with } c_{22} &= -4\,\bar{x}_2\left({s_2}^2 + \bar{x}_2^2\right)\left({s_1}^2 + \bar{x}_1^2\right), \\
c_{21} &= -2\,\bar{x}_1\left(3\,\bar{x}_2^2 + {s_2}^2\right)(n_1 - n_2), \\
c_{20} &= \bar{x}_2\,n_2\,(-2\,n_2 + 3\,n_1), \\
c_3 &= c_{32}\lambda^2 + c_{31}\lambda + c_{30}, \\
\text{with } c_{32} &= 2\left(3\,\bar{x}_2^2 + {s_2}^2\right)\left({s_1}^2 + \bar{x}_1^2\right), \\
c_{31} &= 2\,\bar{x}_1\,\bar{x}_2\,(2\,n_1 - 3\,n_2), \\
c_{30} &= -n_2\,(n_1 - n_2), \\
c_4 &= c_{42}\lambda^2 + c_{41}\lambda, \\
\text{with } c_{42} &= -4\,\bar{x}_2\left({s_1}^2 + \bar{x}_1^2\right), \\
c_{41} &= -\bar{x}_1\,(n_1 - 2\,n_2), \text{ and} \\
c_5 &= \left({s_1}^2 + \bar{x}_1^2\right)\lambda^2,
\end{aligned}$$

and

$$
\begin{aligned}
\hat{\mu}_1(\lambda) &= \frac{n_2\,[\bar{x}_2 - \hat{\mu}_2(\lambda)]}{\lambda\left[s_2^2 + (\bar{x}_2 - \hat{\mu}_2(\lambda))^2\right]}, \quad \lambda \neq 0, \\
\hat{u}_1 &= \frac{{s_1}^2 + [\bar{x}_1 - \hat{\mu}_1(\lambda)]^2}{n_1} \\
\hat{u}_2 &= \frac{{s_2}^2 + [\bar{x}_2 - \hat{\mu}_2(\lambda)]^2}{n_2}
\end{aligned}
$$

where λ satisfies $\rho(\lambda) = \hat{\mu}_1(\lambda)\hat{\mu}_2(\lambda) = \rho_0$.

In general, an algebraic solution of (4.24) is not possible. So it has to be solved numerically. This can easily be done with IMSL subroutine 'zplrc'. If more than one root is real, we have to choose the root that corresponds to the maximum of the likelihood. In the present case, we can ignore all the negative roots because we assume that the means are positive. We found that for negative λ, $\hat{\mu}_1(\lambda)$ is greater than $\bar{x}_1$ and $\hat{\mu}_2(\lambda)$ is greater than $\bar{x}_2$. Also, for positive λ, both estimates are smaller than the corresponding sample averages. Obviously, the case $\lambda = 0$ corresponds to the unrestricted maximum likelihood. We found that $\rho(\lambda)$ is monotonically decreasing in an interval including the origin. Therefore it is easy to invert the likelihood ratio using Madansky's approach to find confidence intervals.

This procedure has the advantage that it can be extended for a sum of products of means. However as the μ_2 has to be estimated numerically, and the valid root has to be selected from five roots, this procedure can be computationally intensive. We now suggest another method which is computationally more attractive, but can be used to construct confidence intervals for one positive product of means only.

Method 2

Suppose that $\rho = \mu_1\mu_2 > 0$. We can now formulate our restriction as $\ln(\mu_1\mu_2) = \ln(\rho_0)$. Hence, the critical points of the Lagrangian,

$$h = \ln l(\boldsymbol{\mu}, \mathbf{u}; \mathbf{x}) - \lambda\,[\ln(\mu_1\mu_2) - \ln(\rho_0)],$$

can be found by solving the following set of equations:

$$\begin{aligned}
\frac{\bar{x}_1 - \mu_1}{u_1} - \frac{\lambda}{\mu_1} &= 0,\\
\frac{\bar{x}_2 - \mu_2}{u_2} - \frac{\lambda}{\mu_2} &= 0,\\
-\frac{n_1}{2\,u_1} + \frac{{s_1}^2 + (\bar{x}_1 - \mu_1)^2}{2\,{u_1}^2} &= 0, \text{ and}\\
-\frac{n_2}{2\,u_2} + \frac{{s_2}^2 + (\bar{x}_2 - \mu_2)^2}{2\,{u_2}^2} &= 0.
\end{aligned}$$

The MAPLE V solution is as follows. $\hat{\mu}_1(\lambda)$ and $\hat{\mu}_2(\lambda)$ are roots of equations,

$$(n_1 + \lambda)\,{\mu_1}^2 - (\,n_1 + 2\,\lambda\,)\,\bar{x}_1\mu_1 + \left({s_1}^2 + \bar{x}_1^2\right)\lambda = 0, \text{ and} \tag{4.25}$$

$$(n_2 + \lambda)\,{\mu_2}^2 - (\,n_2 + 2\,\lambda\,)\,\bar{x}_2\mu_2 + \left({s_2}^2 + \bar{x}_2^2\right)\lambda = 0, \tag{4.26}$$

respectively. The estimates of variances are

$$\hat{u}_1(\lambda) = \frac{{s_1}^2 + (\bar{x}_1 - \hat{\mu}_1)^2}{n_1}, \text{ and} \tag{4.27}$$

$$\hat{u}_2(\lambda) = \frac{{s_2}^2 + (\bar{x}_2 - \hat{\mu}_2)^2}{n_2}. \tag{4.28}$$

Considering the facts that $\hat{\mu}_1(0) = \bar{x}_1$ and $\hat{\mu}_2(0) = \bar{x}_2$, valid roots can be chosen from (4.25) and (4.26) as

$$\begin{aligned}
\hat{\mu}_1(\lambda) &= \frac{\bar{x}_1\,(\,n_1 + 2\,\lambda) + \sqrt{\bar{x}_1^2{n_1}^2 - 4\,n_1\,{s_1}^2\,\lambda - 4{s_1}^2\,\lambda^2}}{2\,(\,n_1 + \lambda)}, \quad \text{for } \lambda \neq -n_1,\\
\hat{\mu}_1(-n_1) &= \bar{x}_1 + \frac{{s_1}^2}{\bar{x}_1},\\
\hat{\mu}_2(\lambda) &= \frac{\bar{x}_2\,(\,n_2 + 2\,\lambda) + \sqrt{\bar{x}_2^2{n_2}^2 - 4\,n_2\,{s_2}^2\,\lambda - 4{s_2}^2\,\lambda^2}}{2\,(\,n_2 + \lambda)}, \quad \text{for } \lambda \neq -n_2, \text{ and}\\
\hat{\mu}_2(-n_2) &= \bar{x}_2 + \frac{{s_2}^2}{\bar{x}_2}.
\end{aligned}$$

where λ satisfies $\hat{\mu}_1(\lambda)\hat{\mu}_2(\lambda) = \rho_0$. To ensure real roots, the appropriate interval for λ is $I_\lambda = [I_1, I_2]$, where

$$\begin{aligned} I_1 &= \max\left(-\frac{n_1}{2}\left[1+\sqrt{1+\frac{\bar{x}_1^2}{s_1^2}}\right], -\frac{n_2}{2}\left[1+\sqrt{1+\frac{\bar{x}_2^2}{s_2^2}}\right]\right) \\ I_2 &= \min\left(\frac{n_1}{2}\left[-1+\sqrt{1+\frac{\bar{x}_1^2}{s_1^2}}\right], \frac{n_2}{2}\left[-1+\sqrt{1+\frac{\bar{x}_2^2}{s_2^2}}\right]\right). \end{aligned}$$

We can show that $\hat{\mu}_1(\lambda)$ is non-increasing in I_λ. To see this, consider the derivative,

$$\frac{\partial \hat{\mu}_1(\lambda)}{\partial \lambda} = \frac{\left(\sqrt{\bar{x}_1^2 {n_1}^2 - 4\, n_1\, {s_1}^2\, \lambda - 4 {s_1}^2\, \lambda^2}\right) \bar{x}_1\, n_1 - \bar{x}_1^2 {n_1}^2 - 2\, n_1\, \lambda\, {s_1}^2 - 2\, {n_1}^2 {s_1}^2}{2\, (n_1 + \lambda)^2 \sqrt{\bar{x}_1^2 {n_1}^2 - 4\, n_1\, {s_1}^2\, \lambda - 4 {s_1}^2\, \lambda^2}}. \tag{4.29}$$

The denominator is positive in I_λ. We can show that the numerator is non-positive in I_λ. To end this, let

$$\begin{aligned} a &= \left(\sqrt{\bar{x}_1^2 {n_1}^2 - 4\, n_1\, {s_1}^2\, \lambda - 4\, {s_1}^2 \lambda^2}\right) \bar{x}_1\, n_1, \\ b &= \bar{x}_1^2 {n_1}^2 + 2\, n_1\, \lambda\, {s_1}^2 + 2\, {n_1}^2 {s_1}^2, \quad \text{and} \\ Q = a^2 - b^2 &= \left(\bar{x}_1^2 {n_1}^2 - 4\, n_1\, \lambda\, {s_1}^2 - 4\, \lambda^2 {s_1}^2\right) \bar{x}_1^2 {n_1}^2 - \left(\bar{x}_1^2 {n_1}^2 + 2\, n_1\, \lambda\, {s_1}^2 + 2\, {n_1}^2 {s_1}^2\right)^2 \\ &= -4\, {n_1}^2 {s_1}^2 (\lambda + n_1)^2 \left({s_1}^2 + \bar{x}_1^2\right). \end{aligned}$$

Suppose that both averages are positive (If both averages are negative, all the calculations can be done neglecting the signs). Then, a is positive in I_λ. When $\lambda > -\frac{n_1}{2}\left(2 + \frac{\bar{x}_1^2}{s_1^2}\right)$, b is positive. Since $I_1 > -\frac{n_1}{2}\left(2 + \frac{\bar{x}_1^2}{s_1^2}\right)$, it implies that b is positive in I_λ. So, the fact that Q is negative in I_λ except at $-n_1$, proves that (4.29) is negative in I_λ. This in turn shows that $\hat{\mu}_1(\lambda)$ is monotonically decreasing in I_λ except at $-n_1$, at which it has an inflection point. Similarly $\hat{\mu}_2(\lambda)$ is also monotonically decreasing in I_λ. Consequently, $\rho(\lambda) = \hat{\mu}_1(\lambda)\hat{\mu}_2(\lambda)$ is monotonically decreasing in I_λ. Hence, again the likelihood ratio test may be inverted using Madansky's approach to construct confidence intervals for ρ. It is possible that there does not exist a $\lambda_2^*(> 0)$ at which the likelihood ratio is equal to $\chi^2_{1,(1-\alpha)}$. For example, this can happen when I_2 is very close to zero, or in other words, when at least one of the coefficients of variation is very high (see Figure 4.2). If this happens, $\rho(I_2)$ may be taken as the lower confidence limit.

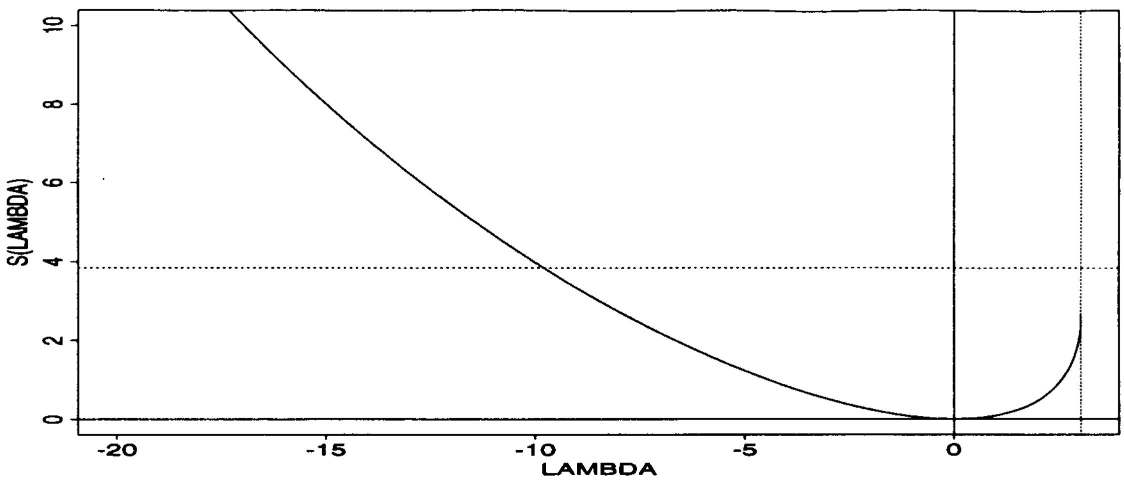


Figure 4.2: An example where $S(\lambda_2^*) < \chi^2_{1,(1-\alpha)}$: (example calculation using $n_1 = 30, n_2 = 30, \bar{x}_1 = 10, \bar{x}_2 = 20, s_1 = 15$, $s_2 = 5$ and $\alpha = .05$. Vertical dotted line represents $\lambda = I_2$. Horizontal dotted line represents $\chi^2_{1,.95}$.)

Sum of p Products of Pairwise Normal Means: Variances Unknown

Method 1 described above can be extended to construct confidence intervals for a sum of p products of pairwise means. The method proceeds by calculating the restricted maximum likelihood estimates of one of the means in each pair using the formulae provided.

As mentioned earlier, Method 2 cannot be extended for sums of products of means. However, it can be easily extended to construct confidence intervals for a product of $q(\geq 2)$ means. This could be especially useful in situations where volumes ($q = 3$) have to be estimated.

4.2.2 Wald Test Approach

Another way of constructing confidence intervals is to invert the Wald test. According to Silvey [35](pp. 115-117), the hypothesis $H_0 : g(\boldsymbol{\theta}) = 0$, can be tested by calculating the unrestricted maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ and basing our decision on the proximity to zero of $g(\hat{\boldsymbol{\theta}})$. This method has the advantage that we do not need to calculate the restricted maximum likelihood estimates. The general idea was first exploited by Wald (1943). Accordingly, the Wald statistic,

$$W = \frac{\left[g(\hat{\boldsymbol{\theta}})\right]^2}{\left[\frac{\partial g(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'} I^{-1}(\hat{\boldsymbol{\theta}}) \frac{\partial g(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right]} \tag{4.30}$$

has a limiting chi-square distribution with one degree of freedom when H_0 is true. Here, the prime denotes the transpose and I is the information matrix.

When the variances are known, for the one product case,

$$\boldsymbol{\theta} = \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \\ \mu_2 \end{pmatrix}, \quad \text{and } g(\boldsymbol{\theta}) = g(\boldsymbol{\mu}) = \mu_1\mu_2 - \rho_0.$$

The information matrix is

$$I(\boldsymbol{\mu}) \;=\; -E\left[\frac{\partial^2 \ln l(\boldsymbol{\mu}; x)}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}'}\right] = \begin{bmatrix} \frac{1}{u_1} & 0 \\ 0 & \frac{1}{u_2} \end{bmatrix},$$

and the derivative of g is

$$\frac{\partial g(\hat{\boldsymbol{\mu}})}{\partial \boldsymbol{\mu}} = \begin{bmatrix} \bar{x}_2 \\ \\ \bar{x}_1 \end{bmatrix}.$$

Therefore, the Wald statistic in this case is

$$W = \frac{(\bar{x}_1\bar{x}_2 - \rho_0)^2}{u_1\bar{x}_2^2 + u_2\bar{x}_1^2}. \tag{4.31}$$

Because of the simple form of this statistic, $\sqrt{W} = \frac{\bar{x}_1\bar{x}_2 - \rho_0}{\sqrt{u_1\bar{x}_2^2 + u_2\bar{x}_1^2}}$, can be easily used to construct confidence intervals. Since $\sqrt{W}$ has a limiting standard normal distribution under

H_0, an approximate $(1-\alpha)100\%$ confidence interval can be calculated as

$$\bar{x}_1\bar{x}_2 \pm z_{\alpha/2}\sqrt{u_1\bar{x}_2^2 + u_2\bar{x}_1^2}.$$

In general, for the p product case, the interval is

$$\sum_{i=1}^{p} \bar{x}_{i1}\bar{x}_{i2} \pm z_{\alpha/2}\sqrt{\sum_{i=1}^{p}(u_{i1}\bar{x}_{i2}^2 + u_{i2}\bar{x}_{i1}^2)}. \tag{4.32}$$

Similarly, when the variances are unknown, the corresponding Wald statistic leads to the approximate confidence interval,

$$\sum_{i=1}^{p} \bar{x}_{i1}\bar{x}_{i2} \pm z_{\alpha/2}\sqrt{\sum_{i=1}^{p}(v_{i1}\bar{x}_{i2}^2 + v_{i2}\bar{x}_{i1}^2)}. \tag{4.33}$$

Remark

The usual confidence interval obtained by adding and subtracting a multiple of the standard error is

$$\sum_{i=1}^{p} \bar{x}_{i1}\bar{x}_{i2} \pm z_{\alpha/2}\sqrt{\sum_{i=1}^{p}(\tilde{v}_{i1}\bar{x}_{i2}^2 + \tilde{v}_{i2}\bar{x}_{i1}^2 - \tilde{v}_{i1}\tilde{v}_{i2})}. \tag{4.34}$$

For large samples, the intervals (4.33) and (4.34) are almost the same because v_{ij} and $\tilde{v}_{ij}$ are asymptotically equivalent and the terms, $\tilde{v}_{i1}\tilde{v}_{i2}$, are of smaller order.

Bartlett-Type Adjustments

Bartlett [4] introduced the idea of adjustment factors to improve the chi-squared approximation to the null hypothesis distribution of the log-likelihood ratio statistic. Suppose that under the null hypothesis, the expected value of twice the log-likelihood ratio statistic S is

$$E(S) = d\{1 + b/n + O(n^{-3/2})\}, \tag{4.35}$$

where either b is known or can be estimated consistently. Then,

$$\tilde{S} = (1 + b/n)^{-1}S \tag{4.36}$$

has an expected value closer to that of χ^2_d than has S. Bartlett showed that for the test of homogeneity of variances, the first three moments of $\tilde{S}$ agree with those of χ^2_d with error $O(n^{-3/2})$, giving strong grounds for thinking that the density of $\tilde{S}$ is better approximated by the χ^2_d density. Barndorff-Nielsen and Cox [3] gave a proof of this useful result, namely that $\tilde{S}$ has the χ^2_d distribution with error $O(n^{-3/2})$.

It may be possible to use this idea to improve the confidence interval (4.33) by calculating an adjustment factor. Since the Wald statistic is asymptotically equivalent to the likelihood ratio, it is reasonable to expect that a Bartlett-type adjustment will improve the confidence intervals calculated through the Wald statistic as well.

Consider the statistic,

$$\tilde{W} = \frac{\left(\sum_{i=1}^{p} \bar{x}_{i1}\bar{x}_{i2} - \rho_0\right)^2}{\sum_{i=1}^{p} \tilde{v}_{i1}\bar{x}_{i2}^2 + \tilde{v}_{i2}\bar{x}_{i1}^2}, \tag{4.37}$$

obtained by respectively replacing v_{i1} and v_{i2} by $\tilde{v}_{i1}$ and $\tilde{v}_{i2}$ in the Wald statistic corresponding to the interval (4.33). Because the $\tilde{v}$'s are unbiased estimators of the u's, the expected values of many terms in the Taylor expansion of $\tilde{W}$ become zero. This simplifies the calculation of adjustment factors. Noting that both W and $\tilde{W}$ have the same asymptotic distribution under the null hypothesis, $H_0 : \sum_{i=1}^{p} \mu_{i1}\mu_{i2} = \rho_0$, we suggest the use of $\tilde{W}$ instead of W to calculate the adjustments.

We calculated the derivatives using Maple V functions. The first order derivatives, when evaluated at $\boldsymbol{\theta}$[1] are zero. The second order derivatives are $O(1)$ and contribute to the approximate distribution of $\tilde{W}$. The expected value of these terms is 1. The third order derivatives, when evaluated at $\boldsymbol{\theta}$ are zero. We found that in order to obtain the terms of order $1/n_0$[2], $\tilde{W}$ should be expanded in a Taylor expansion up to fourth order derivatives. The higher order derivatives are $O(n_0^{-3/2})$. The following Conjecture condenses the derivatives produced by Maple V and suggests an adjustment factor to $\tilde{W}$.

---

[1] θ is the vector of all parameters defined as in Subsection 4.1.1

[2] $n_0 = \min(n_{11}, n_{12}, \ldots, n_{p1}, n_{p2})$

Conjecture 4.1 *Let*

$$\delta = \sum_{i=1}^{p}\sum_{l=1}^{2} u_{il}\mu_{i(3-l)}^2.$$

Then, the fourth order derivatives of $\tilde{W}$, evaluated at $\boldsymbol{\theta}$, are

$$\begin{aligned}
\frac{\partial^4\tilde{W}(\boldsymbol{\theta})}{\partial\bar{x}_{il}^4} &= \frac{96\mu_{il}^2\mu_{i(3-l)}^2u_{i(3-l)}^2}{\delta^3} - \frac{24\mu_{i(3-l)}^2u_{i(3-l)}}{\delta^2}, \\
\frac{\partial^4\tilde{W}(\boldsymbol{\theta})}{\partial\bar{x}_{il}^2\partial\bar{x}_{i(3-l)}^2} &= \frac{4}{\delta} - \frac{20\left[\mu_{il}^2u_{i(3-l)} + \mu_{i(3-l)}^2u_{il}\right]}{\delta^2} \\
&\quad + \frac{16\left[\mu_{il}^4u_{i(3-l)}^2 + \mu_{i(3-l)}^4u_{il}^2\right]}{\delta^3} \\
&\quad + \frac{64\left[\mu_{il}^2\mu_{i(3-l)}^2u_{il}u_{i(3-l)}\right]}{\delta^3}, \\
\frac{\partial^4\tilde{W}(\boldsymbol{\theta})}{\partial\bar{x}_{il}^2\partial\bar{x}_{jm}^2} &= \frac{16\left[\mu_{il}^2\mu_{j(3-m)}u_{i(3-l)}^2 + \mu_{jm}^2\mu_{i(3-l)}u_{j(3-m)}^2\right]}{\delta^3} \\
&\quad - \frac{4\left[\mu_{i(3-l)}^2u_{j(3-m)} + \mu_{j(3-m)}^2u_{i(3-l)}\right]}{\delta^2} \\
&\quad + \frac{64\mu_{il}\mu_{i(3-l)}\mu_{jm}\mu_{j(3-m)}u_{il}u_{i(3-l)}u_{jm}u_{j(3-m)}}{\delta^3}, \text{ and} \\
\frac{\partial^4\tilde{W}(\boldsymbol{\theta})}{\partial\bar{x}_{il}^2\partial\tilde{v}_{jm}^2} &= \frac{4\mu_{i(3-l)}^2\mu_{j(3-m)}^4}{\delta^3}, \\
\frac{\partial^4\tilde{W}(\boldsymbol{\theta})}{\partial\tilde{v}_{jm}^4} &= 0.
\end{aligned}$$

The fourth order expected values are

$$\begin{aligned}
E\left[(\bar{x}_{il} - \mu_{il})^4\right] &= 3u_{il}^2, \text{ and} \\
E\left[(\bar{x}_{il} - \mu_{il})^2(\tilde{v}_{jm} - u_{jm})^2\right] &= E(\bar{x}_{il} - \mu_{il})^2E(\tilde{v}_{jm} - u_{jm})^2 \\
&= u_{il}\frac{2u_{jm}^2}{(n_{jm} - 1)}.
\end{aligned}$$

Let

$$\begin{aligned}
A &= \frac{1}{4}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{2}\sum_{m=1}^{2} E\left[(\bar{x}_{il} - \mu_{il})^2(\bar{x}_{jm} - \mu_{jm})^2\right]\frac{\partial^4\tilde{W}(\boldsymbol{\theta})}{\partial\bar{x}_{il}^2\partial\bar{x}_{jm}^2}, \\
B &= \frac{1}{4}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{2}\sum_{m=1}^{2} E\left[(\bar{x}_{il} - \mu_{il})^2(\tilde{v}_{jm} - u_{jm})^2\right]\frac{\partial^4\tilde{W}(\boldsymbol{\theta})}{\partial\bar{x}_{il}^2\partial\tilde{v}_{jm}^2},
\end{aligned}$$

$C = (1 + A + B)^{-1}$,

$\hat{C}$ = *estimated C obtained by replacing means and variances by the corresponding sample averages and unbiased sample variances, and* (4.38)

$n_0 = \min(n_{11}, n_{12}, \ldots, n_{p1}, n_{p2}).$ (4.39)

Then,

1. $E(\tilde{W}) = 1 + A + B + O(n_0^{-3/2})$, *and*
2. $\tilde{\tilde{W}} = \hat{C}\tilde{W}$ *has an approximate* χ_1^2 *distribution with error* $O(n_0^{-3/2})$.

So, the confidence interval,

$$\sum_{i=1}^{p} \bar{x}_{i1}\bar{x}_{i2} \pm z_{\alpha/2}\sqrt{\frac{1}{\hat{C}}\sum_{i=1}^{p}(\tilde{v}_{i1}\bar{x}_{i2}^2 + \tilde{v}_{i2}\bar{x}_{i1}^2)}, \tag{4.40}$$

is expected to perform better than (4.33). Some numerical comparisons are presented in Section 4.4.

4.3 Adaptation to Distributions other than Normal

Now, we address the problem for non normal data with the particular application of the fish problem in mind. Recall that in this problem, one variable is the reciprocal target width (M) of a fish on an echogram, which is proportional to the fish speed (stem and leaf plots of target widths W_i are shown in Figure 4.3). The other variable is the number of fish detected per transect (N). Figure 4.4 shows histograms of these variables based on data collected by the Pacific Salmon Commission on August 31, 1993 at Mission, B.C.. According to these histograms, the variables seem to have positively skewed distributions. In the following table, we have reproduced the summary data in Table 4.1.

One way to construct confidence intervals in this case is to model the data by suitable probability distributions and then to invert the likelihood ratio in the usual way. But this

```
Variable = W1
No. of observations = 77
Median = 1
Quartiles = 0.8, 1.3

Decimal point is at the colon

  0 : 55
  0 : 66666666666777
  0 : 8888888888899999999999
  1 : 0001111111111
  1 : 223333333
  1 : 4444444555
  1 : 66666
  1 : 8
  2 : 0
```

```
Variable = W2
No. of observations = 145
Median = 1.5
Quartiles = 1.1, 1.8

Decimal point is at the colon

  0 : 66
  0 : 8888888888888899
  1 : 000000000111111111111111111
  1 : 2222222333333333333
  1 : 4444444555555555
  1 : 66666666666666667777777777
  1 : 888888888888888889999
  2 : 000000011
  2 : 2333
  2 : 4455
  2 : 6
```

```
Variable = W3
No. of observations = 42
Median = 1.45
Quartiles = 1, 1.9

Decimal point is at the colon

  0 : 66678889
  1 : 0001133333344
  1 : 55566688899
  2 : 001123
  2 : 5588
```

Figure 4.3: Stem and leaf plots of target widths, W_i's (based on hydroacoustic data collected by the Pacific Salmon Commission on August 31, 1993 at Mission, B.C.).

Depth stratum (i)	n_i	$\bar{N}_i$	s_{N_i}	$CV(N)\%$	m_i	$\bar{M}_i$	s_{M_i}	$CV(M)\%$
1	215	1.20	1.24	103	77	1.07	0.37	34
2	215	6.38	4.17	65	145	0.76	0.26	34
3	215	1.74	1.54	88	42	0.79	0.36	46

Table 4.1: Summary statistics of the data collected by the Pacific Salmon Commission on August 31, 1993, at Mission, B.C.

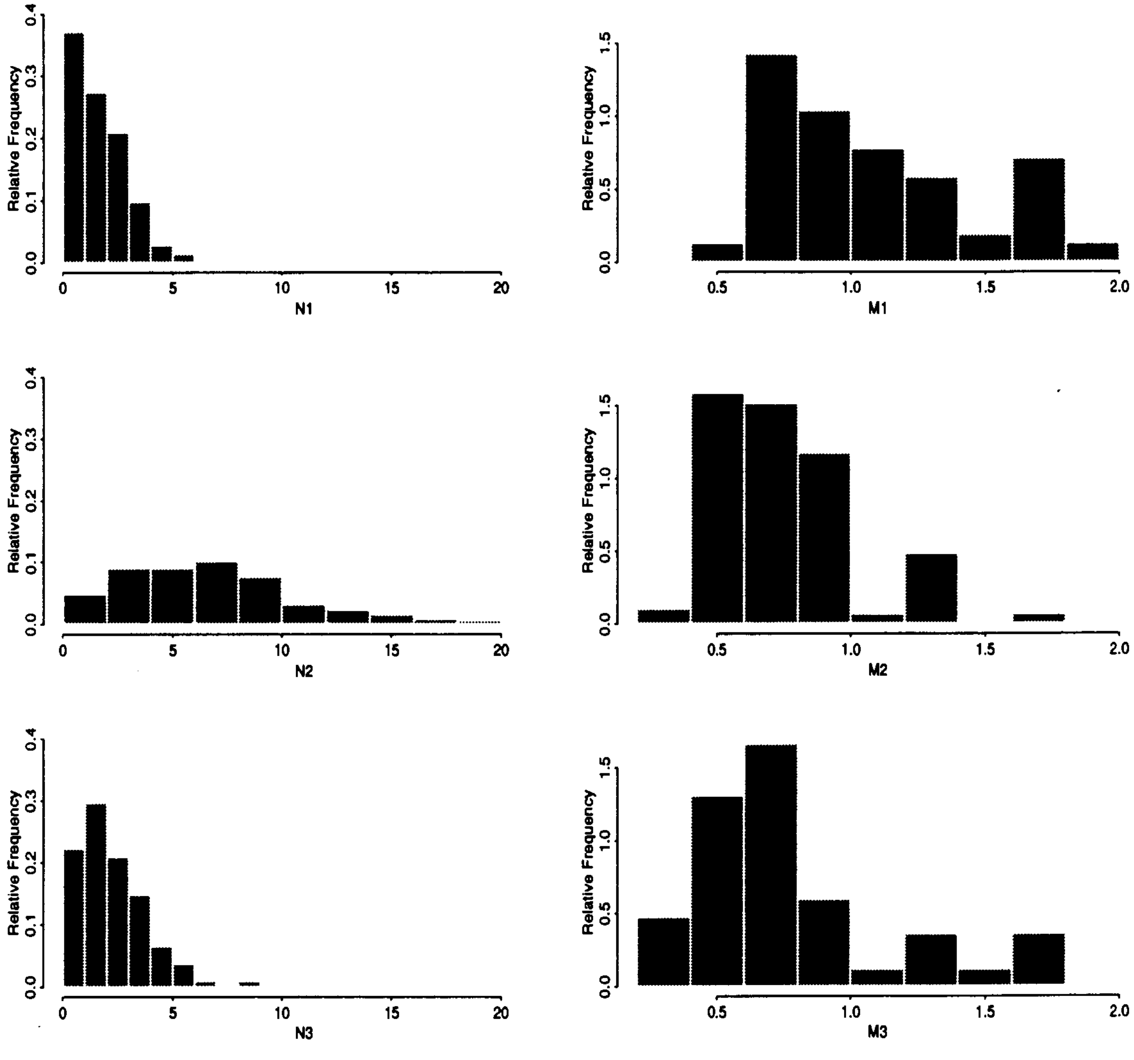


Figure 4.4: Histograms of N_i's and M_i's (based on hydroacoustic data collected by the Pacific Salmon Commission on August 31, 1993 at Mission, B.C.). The height of each bar is the proportion of observations in the corresponding interval on the horizontal axis.

method requires a parametric assumption which may be difficult to justify. We recommend the following alternative.

With large samples[3], by the central limit theorem, we can expect the sample averages to be fairly normal. Also, the unknown variances can be estimated with high accuracy. So, the unbiased estimates of variances may be treated as true population variances. Therefore, the methods developed for normal data with known variances may be adapted for this problem.

First we adapt the likelihood ratio methods. Now, the maximum likelihood estimates of variances cannot be derived. Hence exact likelihood ratios cannot be calculated. However, an approximate likelihood ratio may be calculated using the the approximate likelihood function,

$$l(\mu; \bar{\mathbf{x}}) \;=\; \exp[-\ln(\pi) - \ln(u_1)/2 - \ln(u_2)/2 - (\bar{x}_1 - \mu_1)^2/2u_1 - (\bar{x}_2 - \mu_2)^2/2u_2].$$

It is easy to see that this (approximate) likelihood is proportional to that given by (4.4). Therefore, when the samples are large enough, the likelihood ratio methods developed for the case of known variances may still be used.

Similarly, approximate confidence intervals may also be calculated using (4.33). For samples of moderate sizes, one may replace the normal quantiles by t quantiles. As a conservative approach, the size of the smallest sample may be taken as the corresponding degrees of freedom. Some comparisons and further discussion are presented in Section 4.4.

4.4 Worked Example and Monte Carlo Studies

4.4.1 Worked Example

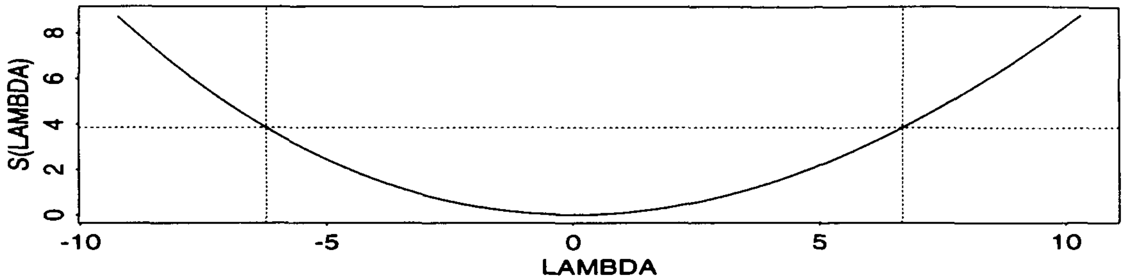
To illustrate the methods suggested for constructing confidence intervals in the case of non normal means, we constructed approximate 95% confidence intervals for the migration rate for the Fraser River fish problem. We first constructed confidence intervals for $\rho =$

---

[3] This is usually the case in the Fraser River hydroacoustic application. See Table 4.1.

$\sum_{i=1}^{3} E(N_i)E(M_i)$ using statistics S given by (4.22).

(a) -2 times the log likelihood ratio



(b) The sum of pairwise products of means

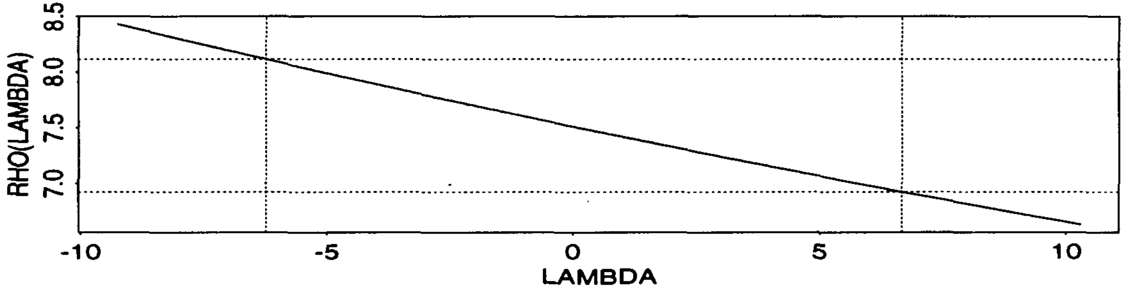


Figure 4.5: Calculation of 95% confidence interval for $\rho = \sum_{i=1}^{3} E(N_i)E(M_i)$.

To explain the calculation of confidence intervals through S, let $\bar{x}_1 = \bar{N}, \bar{x}_2 = \bar{M}, s_1^2 = (s_{N_i}^2), s_2^2 = (s_{M_i}^2), n_1 = n$ and $n_2 = m$. Treating the sample variances as true population variances, let $u_i = s_1^2/n_i$. The resulting $S(\lambda)$ and $\rho(\lambda)$ as given by (4.22) and (4.23), and the calculation of confidence intervals are schematically shown in Figure 4.5. The dotted horizontal line in (a) represents the value $\chi^2_{.95,1} = 3.841459$. From this figure, it is clear that $S(\lambda)$ cuts this line at two values between -10 and 10. Using the bisecton method, we found the corresponding λ values to be λ = -6.242009 and 6.678391. Substituting them in $\rho(\lambda)$, an approximate 95% confidence interval for ρ was calculated. Then, by multiplying the limits by $\frac{2p}{\pi} \times 86,400$, a confidence interval for the total abundance can be calculated. We also calculated confidence intervals for those quantities through the Wald statistic with normal and t quantiles. These confidence intervals are presented in Table 4.2. In this Table, W_z and W_t denote the confidence intervals obtained with normal and t quantiles respectively.

	95% confidence intervals		
	S	W_z	W_t
$\rho = \sum_{i=1}^{3} E(M_i)E(N_i)$	(6.92, 8.11)	(6.91,8.10)	(6.89, 8.12)
Daily abundance	(181,940, 213,227)	(181,745, 213,023)	(181,281, 213,486)

Table 4.2: 95% confidence intervals for the fish problem

4.4.2 Monte Carlo Studies

Next we describe several Monte Carlo studies that were performed to assess the appropriateness of the suggested methods.

First, to assess the appropriateness of the methods suggested for non normal data, we generated large samples from different distributions with the means and variances that are equal to sample averages and sample variances in the above example (in the case of Poisson, only the mean can be matched). Treating the sample variances as true population variances, we conducted 10,000 Monte Carlo simulations. For each simulation, we calculated confidence intervals based on S given by (4.22) and W given by (4.33). Since the usual confidence intervals calculated by adding and subtracting the standard errors are almost the same as those calculated using W (traditional intervals are slightly narrower; see (4.34)), in this way we can compare the the likelihood ratio intervals and the traditional intervals. The W intervals calculated using normal quantiles and t quantiles are denoted by W_z and W_t respectively. The results are presented in Tables 4.5 and 4.6. To conserve space in the tables, we give a group number for each combination of distributions and sample sizes as identified in Tables 4.3 and 4.4.

Group of Distributions	Dist of $\mathbf{x}_1$	Dist of $\mathbf{x}_2$
1	Poisson Log-Normal	Gamma
2	Poisson	Gamma
3	Gamma	Gamma
4	Poisson Log-Normal	Log-Normal
5	Poisson	Log-Normal
6	Gamma	Log-Normal

Table 4.3: Group numbers for combinations of distributions

Group of	Sample sizes of $\mathbf{x}_1$			Sample sizes of $\mathbf{x}_2$		
Sample sizes	n_{11}	n_{12}	n_{13}	n_{21}	n_{22}	n_{23}
1	30	30	30	30	30	30
2	40	40	40	40	40	40
3	60	60	60	40	40	40
4	100	100	100	40	40	40
5	200	200	200	70	140	40

Table 4.4: Group numbers for combinations of sample sizes

Group of Distns.	Group of sam. sizes	Nominal Coverage=90%			Nominal Coverage=95%			Nominal Coverage=99%		
		Observed Coverage			Observed Coverage			Observed Coverage		
		S	W_z	W_t	S	W_z	W_t	S	W_z	W_t
1	1	0.891	0.891	0.900	0.942	0.940	0.949	0.987	0.984	0.989
1	2	0.891	0.890	0.899	0.946	0.944	0.950	0.984	0.982	0.986
1	3	0.897	0.896	0.903	0.946	0.944	0.953	0.985	0.984	0.988
1	4	0.893	0.892	0.902	0.946	0.945	0.953	0.986	0.986	0.981
1	5	0.899	0.899	0.906	0.953	0.952	0.958	0.987	0.987	0.991
2	1	0.895	0.892	0.902	0.946	0.945	0.956	0.989	0.988	0.991
2	2	0.896	0.896	0.904	0.949	0.948	0.955	0.987	0.987	0.989
2	3	0.898	0.899	0.906	0.945	0.945	0.953	0.990	0.988	0.992
2	4	0.895	0.894	0.903	0.948	0.946	0.954	0.988	0.987	0.991
2	5	0.896	0.896	0.904	0.951	0.951	0.959	0.989	0.989	0.992
3	1	0.885	0.884	0.893	0.938	0.936	0.945	0.984	0.981	0.986
3	2	0.893	0.894	0.901	0.945	0.945	0.951	0.988	0.986	0.990
3	3	0.893	0.893	0.902	0.946	0.946	0.952	0.989	0.986	0.990
3	4	0.904	0.902	0.910	0.947	0.946	0.953	0.989	0.986	0.991
3	5	0.898	0.898	0.905	0.951	0.950	0.957	0.990	0.990	0.994

Table 4.5: Observed coverage in 10,000 simulations: $E(\mathbf{x}_1)=(1.20, 6.38, 1.74)$, $V(\mathbf{x}_1)=(1.53, 17.4, 2.36)$, $E(\mathbf{x}_2)=(1.07, 0.76, 0.79)$, $V(\mathbf{x}_2)=(0.136, 0.066, 0.130)$.

Group of Distns.	Group of sam. sizes	Nominal Coverage=90%			Nominal Coverage=95%			Nominal Coverage=99%		
		Observed Coverage			Observed Coverage			Observed Coverage		
		S	W_z	W_t	S	W_z	W_t	S	W_z	W_t
4	1	0.893	0.894	0.903	0.936	0.934	0.945	0.986	0.983	0.988
4	2	0.893	0.892	0.899	0.946	0.945	0.952	0.987	0.985	0.989
4	3	0.888	0.887	0.896	0.944	0.941	0.949	0.986	0.984	0.988
4	4	0.897	0.896	0.905	0.948	0.948	0.954	0.986	0.986	0.989
4	5	0.893	0.892	0.900	0.946	0.946	0.953	0.988	0.987	0.992
5	1	0.899	0.898	0.910	0.944	0.943	0.953	0.989	0.989	0.9928
5	2	0.897	0.897	0.905	0.945	0.945	0.952	0.989	0.987	0.992
5	3	0.899	0.900	0.908	0.948	0.948	0.955	0.989	0.989	0.992
5	4	0.899	0.898	0.905	0.949	0.950	0.956	0.988	0.987	0.990
5	5	0.903	0.902	0.909	0.949	0.950	0.956	0.987	0.988	0.992
6	1	0.883	0.882	0.893	0.942	0.940	0.950	0.983	0.980	0.986
6	2	0.894	0.893	0.901	0.941	0.940	0.947	0.985	0.982	0.987
6	3	0.894	0.895	0.903	0.947	0.946	0.953	0.986	0.985	0.989
6	4	0.902	0.901	0.907	0.944	0.944	0.951	0.988	0.986	0.991
6	5	0.901	0.902	0.908	0.951	0.952	0.959	0.988	0.987	0.991

Table 4.6: Observed coverage in 10,000 simulations: $E(\mathbf{x}_1)$=(1.20, 6.38, 1.74), $V(\mathbf{x}_1)$=(1.53,1 7.4, 2.36), $E(\mathbf{x}_2)$=(1.07, 0.76, 0.79), $V(\mathbf{x}_2)$=(0.136, 0.066, 0.130).

Sample size groups 4 and 5 contain the largest sample sizes. From these rows in Tables 4.5 and 4.6, there is evidence that for large samples, all three methods work well. For small samples (sample size groups 1-3) S and W_z coverages are slightly lower than the nominal value. However, from W_t column, it is apparent that the W intervals calculated using t quantiles with smallest sample size degrees of freedom, work fairly well even for small samples.

To compare S and W confidence intervals for normal data, we considered several combinations of variance-to-mean ratios. First, we considered the one product case. The results are presented in Tables 4.7 and 4.8. In these tables, Adjusted W is denoted by W_a.

Means		Standard Deviations		Sample Sizes		Coeff. of Var. of		Observed Coverage (Nominal Coverage=.95)		
μ_1	μ_2	σ_1	σ_2	n_1	n_2	$\bar{x}_1$	$\bar{x}_2$	S	W_z	W_a
10	20	3	5	10	10	9.49	7.91	0.948	0.933	0.956
10	20	3	5	20	20	6.71	5.60	0.949	0.942	0.953
10	20	3	5	30	30	5.48	4.56	0.949	0.945	0.954
10	20	3	5	40	40	4.74	3.95	0.950	0.946	0.953

Continued on next page

Table 4.7: Observed coverage in 10,000 simulations for normal data. (Nominal coverage =95%)

These simulated results suggest that likelihood ratio intervals are superior to Wald test intervals for small samples. However, the adjusted Wald test intervals seem to be as good as likelihood ratio intervals. When the coefficients of variation of the distributions are substantial, large samples are necessary even for the likelihood ratio intervals to be accurate.

Means		Standard Deviations		Sample Sizes		Coeff. of Var. of		Observed Coverage (Nominal Coverage=.95)		
μ_1	μ_2	σ_1	σ_2	n_1	n_2	$\bar{x}_1$	$\bar{x}_2$	S	W_z	W_a
6.38	0.76	4.170	0.257	10	10	20.67	10.69	0.946	0.931	0.954
6.38	0.76	4.170	0.257	20	20	14.61	7.56	0.949	0.941	0.954
6.38	0.76	4.170	0.257	30	30	11.93	6.17	0.949	0.944	0.953
6.38	0.76	4.170	0.257	40	40	10.33	5.35	0.950	0.944	0.953
10	20	7	5	10	10	22.13	7.90	0.938	0.925	0.952
10	20	7	5	20	20	15.65	5.59	0.947	0.937	0.953
10	20	7	5	30	30	12.78	4.56	0.949	0.942	0.953
10	20	7	5	40	40	11.06	3.95	0.949	0.945	0.952
10	20	7	15	10	10	22.13	23.71	0.935	0.919	0.954
10	20	7	15	20	20	15.65	16.77	0.948	0.936	0.957
10	20	7	15	30	30	12.78	13.69	0.950	0.943	0.958
10	20	7	15	40	40	11.06	11.86	0.948	0.942	0.953

Table 4.8: Observed coverage in 10,000 simulations for normal data. (Nominal coverage =95%)

Results are similar for more than one products. However, then the calculation of likelihood ratio intervals is much more time consuming. As an example, we generated 5,000 samples, and computed 95% confidence intervals for sums of three pairwise products of means. The means and standard deviations for this example were set equal to those in the fish problem. The results from this simulation are presented in Table 4.9.

Sample sizes						Nominal Coverage=95%		
n_{11}	n_{12}	n_{13}	n_{21}	n_{22}	n_{23}	S	W_z	W_a
20	20	20	20	20	20	0.945	0.937	0.950

Table 4.9: Observed coverage in 5,000 simulations for normal data. (Nominal coverage =95%)

4.5 Summary

In this chapter we discussed the problem of finding confidence intervals for sums of products of pairwise means. These confidence intervals were constructed by gathering all the parameter values that were not rejected by a formal hypothesis test. We looked at two approaches in detail; one based on the likelihood ratio test and the other on the Wald test.

In the case at hand, the direct inversion of the likelihood ratio test and the Wald test are not numerically attractive. Using Madansky's [25] approach, we suggested simplifications to the computational problem. Related theoretical details were provided for the normal data. Bartlett-type adjustment factors were computed to improve the accuracy of the coverage probabilities of the confidence intervals obtained using the Wald test approach. We also suggested a procedure (Method 2), particularly designed for a product of normal means, which is numerically more tractable than the approach described under Method 1. While this approach has the advantage that it can be easily extended to a product of any finite number of means, it has the limitation that it cannot be extended to a sum of products of means. Suggestions were made for other distributions when the sample sizes are large enough to justify an appeal to the central theorem.

The applicability of the methods was examined by a Monte Carlo study. For non normal data, a conservative approach with t quantiles was found to work well even with samples of moderate sizes. With very large samples, both likelihood ratio and Wald test approaches seem to work well. In the fish example, the sample sizes were daily aggregates. In order to produce more precise estimates accounting for daily cycles, it may be necessary to stratify the day into smaller time intervals. Then, the number of transects within a time interval may be as small as 30-40. Similarly, when fewer fish are in the river, the numbers from stationary soundings may be smaller. In this case it may be better to construct confidence intervals through W with t quantiles.

For normal data, in situations such as when the coefficients of variation of the distributions are large, confidence intervals based on Wald test were often found to have smaller than the nominal coverage. Monte Carlo study shows that the confidence intervals obtained using the adjusted Wald statistic are as appealing as those obtained from the likelihood ratio test.

Chapter 5

Stratified Two-Sample Tag Recovery Census of Closed Populations

In the simplest type of tag recovery experiments a simple random sample of m animals is taken from a population, tagged and released. After allowing enough time for the tagged animals to mix with others, a second sample is taken and the numbers of tagged animals (n) and untagged animals (v) are counted. Assuming every animal has the same probability p of being sampled, and the death and emigration rates are negligible, p is estimated by $\widehat{p} = n/m$. The number V of untagged animals in the population at the time of recovery is then estimated by $\widehat{V} = v/\widehat{p} = mv/n$. This leads to the well known Petersen estimator of the total population size $\widehat{T} = \widehat{V} + m = m(v+n)/n$.

Very often a population is stratified geographically, so that the total population may be regarded as consisting of separate strata living in different areas. Sometimes, the stratification may be with respect to time instead of place. In this case, the population size can be estimated by a stratified version of the Petersen estimate. In this method, a known number of tagged animals is released in each stratum, using a different tag for each stratum. After allowing for the tagged animals to disperse, a random sample is taken from each recovery stratum. The numbers of untagged animals and tagged animals of each type in each sample are then recorded. Several authors have investigated this method. For example, Schaefer [30] stratified both the tagging and recovery with respect to time to estimate a migrating salmon run. Chapman and Junge [7] showed that Schaefer's estimator is not in general consistent. They estimated the numbers of animals in recovery strata from sets of equations relating to the expected values of observed frequencies. Darroch [13] derived the maximum likelihood estimators for the case of equal numbers of tagging and recovery strata. For other cases, he provided alternative estimators. Seber [32](Ch.11), presented an overview of Darroch's paper. Plante [28] derived maximum likelihood estimators for cases where the number of tagging strata is not equal to the number of recovery strata. In a similar experiment, Cormack and Skalski [9] employed log-linear models to analyze coded-wire tag returns from commercial catches, when the sampling proportions from commercial catches were known.

In this chapter we closely follow Chapman and Junge, and Darroch, but derive least-squares estimators of stratum sizes. These estimators are shown to overcome some difficulties that arise in the existing methods. The estimators are shown to be consistent. As well, the formulae for calculating asymptotic variances of the estimators are derived.

5.1 Notation

Let

$$s = \text{number of tagging strata, and}$$

$$t = \text{number of recovery strata.}$$

The numbers of animals in tagging and recovery stages are denoted by symbols as in Table 5.1.

	Tagging stage	Recovery stage	
	Population	Population	Sample
Marked	m	N	n
Unmarked	U	V	v
Total		T	k

Table 5.1: Symbols for Numbers of Animals

Subscripts are attached to the above symbols to denote the corresponding tagging or recovery stratum as follows.

m_i = number of tagged animals released in stratum i,

U_i = number of untagged animals in stratum i at the time of tagging,

N_{ij} = number of tagged animals migrating from stratum i to stratum j,

V_j = number of untagged animals in stratum j at the time of recovery,

T_j = total number of animals in stratum j at the time of recovery,

n_{ij} = the number of marked animals released in stratum i and recovered in stratum j,

v_j = number of untagged animals in jth recovery sample, and

k_j = size of sample from stratum j.

The moving, survival, and sampling probabilities are denoted as follows.

θ_{ij} = probability that a marked animal moves from stratum i to stratum j,

ϕ_i = $\sum_{j=1}^{t} \theta_{ij}$ = survival probability of a marked animal released in stratum i,

p_j = sampling probability in stratum j: assumed non-zero,

ρ_j = $1/p_j$, and

ψ_{ij} = $\theta_{ij} p_j$ = probability that a marked animal released in stratum i will be caught in stratum j.

A sum over any subscript is denoted by replacing the subscript by '+'. For example,

m_+ = $\sum_{i=1}^{s} m_i$ = total number of tagged animals released,

U_+ = $\sum_{i=1}^{s} U_i$ = total number of untagged animals in tagging strata,

N_{i+} = $\sum_{j=1}^{t} N_{ij}$ = total number of tagged animals successfully migrated to recovery strata from the ith tagging stratum.

V_+ = $\sum_{j=1}^{t} V_j$ = total number of untagged animals at the time of recovery,

T_+ = $\sum_{j=1}^{t} T_j$ = total population size,

n_{++} = $\sum_{i=1}^{s} \sum_{j=1}^{t} n_{ij}$ = total number of tagged animals in the t recovery samples, and

v_+ = $\sum_{j=1}^{t} v_j$ = total number of untagged animals in t recovery samples.

Vectors and matrices are denoted by boldface letters. For example let

$$\begin{aligned}
\mathbf{m} &= \begin{pmatrix} m_1 \\ \vdots \\ m_s \end{pmatrix}, \ \mathbf{U} = \begin{pmatrix} U_1 \\ \vdots \\ U_s \end{pmatrix}, \ \mathbf{V} = \begin{pmatrix} V_1 \\ \vdots \\ V_t \end{pmatrix}, \ \mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_t \end{pmatrix}, \ \mathbf{k} = \begin{pmatrix} k_1 \\ \vdots \\ k_t \end{pmatrix}, \\
\mathbf{p} &= \begin{pmatrix} p_1 \\ \vdots \\ p_t \end{pmatrix}, \ \boldsymbol{\rho} = \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_t \end{pmatrix}, \ \boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_s \end{pmatrix}, \\
\mathbf{n} &= (n_{ij})_{(s\times t)}, \ \boldsymbol{\Theta} = (\theta_{ij})_{(s\times t)}, \text{ and } \boldsymbol{\psi} = (\psi_{ij})_{(s\times t)}.
\end{aligned}$$

Expected values are denoted by an overline. For instance,

$$\overline{\mathbf{n}} = E(\mathbf{n}), \quad \overline{\mathbf{V}} = E(\mathbf{V}), \text{ and } \overline{\mathbf{v}} = E(\mathbf{v}).$$

In addition, let

$$\begin{aligned}
\mathbf{D_x} &= \text{the diagonal matrix formed from the vector, } \mathbf{x}, \\
\mathbf{A}' &= \text{the transpose of matrix } \mathbf{A}, \\
\mathbf{A}^\dagger &= \text{the Moore-Penrose inverse of matrix } \mathbf{A}, \\
\delta_{ij} &= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \\
\mathbf{1} &= \text{vector of 1's: the dimension to be understood from the context}
\end{aligned}$$

5.2 Estimation

First, we provide a brief overview of the estimators derived by Schaefer [30], Chapman and Junge [7], Darroch [13], and Seber [32]. One purpose of this is to point out the difficulties that can arise in using these estimators. The other purpose is to show the close relationships between these estimators and the least-squares estimators that we derive in this section.

Schaefer's [30] estimator for the total is

$$\hat{T}_+ = \sum_j \hat{T}_j = \sum_{i=1}^{s} \sum_{j=1}^{t} \frac{k_j m_i n_{ij}}{n_{i+} n_{+j}}. \tag{5.1}$$

Chapman and Junge [7] showed that this estimator is not consistent in general, but consistent if the samplings in recovery strata are proportional to the population sizes in each strata. They estimated the unknown stratum sizes (T_j's) from sets of equations relating to the expected values of the observed frequencies. Schaefer as well as Chapman and Junge assumed that each animal has zero probability of dying or emigrating between tagging and recovering, or in other words that the 'survival probability' is one. As this assumption is not often justified, Darroch framed the likelihood theory in such a way that it may be avoided if desired. The price paid for dropping this assumption is that one has to be content with estimating the stratum sizes that would have prevailed had there been no deaths or migration. These authors discussed the estimation in three cases, according to the relative sizes of s and t.

Case 1: $s = t$

If the tagged and untagged animals have the same probability of being captured, then

$$E(n_{ij}|N_{ij}) = N_{ij}\frac{k_j}{T_j} \quad \text{for all } i \text{ and } j.$$

Observing that

$$\sum_{j=1}^{t} \frac{T_j}{k_j} E\left(n_{ij}|N_{ij}\right) = N_{i+} = m_i \quad \text{for all } i,$$

Chapman and Junge formed the set of equations,

$$\sum_{j=1}^{t} \frac{\widehat{T}_j}{k_j} n_{ij} = m_i \quad (i = 1, \ldots, s), \tag{5.2}$$

which led to the estimator,

$$\widehat{\mathbf{T}} = \mathbf{D_k n^{-1} m}. \tag{5.3}$$

Darroch [13] adopted a product-multinomial sampling model to develop the following probability distributions to derive the maximum likelihood estimators:

$$p[\{n_{ij}\}|\{m_i\}] = \frac{\prod_i m_i!}{\prod_i (m_i - n_{i+})! \prod_{ij} n_{ij}!} \prod_i (1 - \sum_j \theta_{ij} p_j)^{(m_i - n_{i+})} \prod_{ij} (\theta_{ij} p_j)^{n_{ij}}, \tag{5.4}$$

and

$$p[\{v_j|V_j\}] = \prod_j \frac{V_j!}{v_j!(V_j - v_j)!} p_j^{v_j} (1 - p_j)^{V_j - v_j}. \tag{5.5}$$

Under this model, the maximum likelihood estimators of $\mathbf{V}$ and $\boldsymbol{\psi} = \boldsymbol{\Theta}\mathbf{D_p}$ are

$$\widehat{\mathbf{V}} = \mathbf{D_v} \widehat{\boldsymbol{\rho}}, \text{ and } \widehat{\boldsymbol{\psi}} = \mathbf{D_m^{-1} n}$$

respectively. If $\boldsymbol{\rho}$ can be estimated, so can $\mathbf{V}$. However, as Darroch pointed out, $\{\theta_{ij}\}$ and $\{p_j\}$ are non-identifiable to the extent of a multiplicative constant, for the likelihood of $\{\theta_{ij}\}$, $\{p_j\}$ is the same as that of $\{\gamma\theta_{ij}\}, \{p_j/\gamma\}$. Hence, neither $\boldsymbol{\rho}$ nor $\mathbf{V}$ can be calculated. To tie down this non-identifiability, he assumed a common survival probability, $\phi_i \equiv \phi$ for all i, and worked with new parameters $\boldsymbol{\Theta}^\star = \boldsymbol{\Theta}/\phi$, $\mathbf{p}^\star = \phi\mathbf{p}$, and the corresponding $\boldsymbol{\rho}^\star$, and $\mathbf{V}^\star = \mathbf{V}/\phi$. Using the facts that $\boldsymbol{\psi} = \boldsymbol{\Theta}^\star \mathbf{D}_{\mathbf{p}^\star}$, and $\boldsymbol{\Theta}^\star \mathbf{1} = \mathbf{1}$, he derived an estimator for $\boldsymbol{\rho}^\star$ as

$$\widehat{\boldsymbol{\rho}}^\star = \mathbf{n}^{-1}\mathbf{m}. \tag{5.6}$$

Since

$$\widehat{\mathbf{V}}^\star = \mathbf{D}_{\bar{\mathbf{v}}} \widehat{\boldsymbol{\rho}^\star}, \tag{5.7}$$

substitution of (5.6) into (5.7) led to the estimator

$$\hat{\mathbf{V}}^\star = \mathbf{D}_\mathbf{V}\mathbf{n}^{-1}\mathbf{m}. \tag{5.8}$$

Consequently,

$$\hat{V}_+^\star = \mathbf{v}'\mathbf{n}^{-1}\mathbf{m}. \tag{5.9}$$

Darroch pointed out that if the marked and unmarked animals have the same probability of survival, then (5.9) is roughly the total unmarked size at the time of tagging. When $\phi \equiv 1$, the resulting estimator of T_+ is the same as Chapman and Junge's total estimator (5.3).

Case 2: $s > t$

In this case, Chapman and Junge's set of equations (5.2) contain more equations than the number of unknown T_j's. So, they suggested to estimate the unknown stratum sizes by pooling enough tagging strata to form a system that has a unique solution. Seber [32] suggested replacing a set of s equations of t unknown V_j's by a linear combination of them, giving the estimator,

$$\hat{\mathbf{V}} = \mathbf{D}_\mathbf{V}(\mathbf{G}\mathbf{n})^{-1}\mathbf{G}\mathbf{m}. \tag{5.10}$$

Here, $\mathbf{G}$ is a $t \times s$ matrix of rank t such that $\mathbf{Gn}$ is nonsingular.

Recall that in Case 1, Darroch assumed that $\phi_i \equiv \phi \ \ \forall i$, imposing $s - 1$ restrictions on the survival probabilities. When $s > t$, his approach requires to impose $t - 1$ restrictions instead. To see why, note that under assumption of common survival probabilities, there are $st + t - s$ independent $\{\theta_{ij}^\star\}, \{p_j^\star\}$ parameters while there are st $\{\psi_{ij}\}$ parameters. Hence, when $s > t$, there are more $\{\psi_{ij}\}$ parameters than $\{\theta_{ij}^\star\}, \{p_j^\star\}$ parameters. Therefore, apart from the relation $\boldsymbol{\psi} = \boldsymbol{\Theta}^\star\mathbf{D}_{\mathbf{p}^\star}$, there must be other dependencies between the two systems. Hence, the simple estimation procedure used in Case 1 is no longer valid. So, Darroch took an alternative approach by allowing ϕ_i to differ. He defined $\bar{\phi} = \sum_i \phi_i/s$ and worked with new parameters $\phi_i^{\star\star} = \phi_i/\bar{\phi}$, $\theta_{ij}^{\star\star} = \theta_{ij}/\bar{\phi}$, $p_j^{\star\star} = \bar{\phi}p_j$ and $\mathbf{V}^{\star\star} = \mathbf{V}/\bar{\phi}$. Then, he imposed

$t-1$ restrictions on $\phi_i^{\star\star}$s. These, including the inherent restriction $\sum_i \phi_i^{\star\star}/s = 1$, can be written in matrix form as

$$\mathbf{A}\boldsymbol{\phi}^{\star\star} = \mathbf{b}. \tag{5.11}$$

Here, $\mathbf{A}$ is a $t \times s$ matrix of rank t with each element in the last row equalling $1/s$, and $\mathbf{b}$ is the $1 \times t$ vector defined as $\mathbf{b}' = (0, \ldots, 0, 1)$. Now, the system $\{\theta_{ij}^{\star\star}\}, \{p_j^{\star\star}\}$ has the same number of independent parameters as the system $\{\psi_{ij}\}$. Hence, following the same steps as in Case 1, but using the facts that $\boldsymbol{\psi} = \boldsymbol{\Theta}^{\star\star}\mathbf{D}_{\mathbf{p}^{\star\star}}$ and $\mathbf{A}\boldsymbol{\Theta}^{\star\star}\mathbf{1} = \mathbf{b}$, estimators for $\boldsymbol{\rho}^{\star\star}$ and $\mathbf{V}^{\star\star}$ were respectively derived as

$$\widehat{\boldsymbol{\rho}^{\star\star}} = (\mathbf{A}\mathbf{D}_{\mathbf{m}}^{-1}\mathbf{n})^{-1}\mathbf{b}, \text{ and } \hat{\mathbf{V}}^{\star\star} = \mathbf{D}_{\mathbf{v}}(\mathbf{A}\mathbf{D}_{\mathbf{m}}^{-1}\mathbf{n})^{-1}\mathbf{b}. \tag{5.12}$$

Case 3: $s < t$

In this case, Chapman and Junge's system of equations cannot be solved uniquely. However, they commented that with additional assumptions, the total population size can be estimated. Darroch also used moment equations to derive estimators in this case. Assuming that the untagged and tagged animals have the same movement pattern, he formed the moment equations,

$$\begin{aligned} \sum_i U_i\psi_{ij} &= v_j, \text{ and} \\ m_i\psi_{ij} &= n_{ij}, \end{aligned}$$

which led to the following set of t equations of s unknown U_i's:

$$\sum_{i=1}^{s} U_i n_{ij}/m_i = v_j \quad (j = 1, \ldots, t). \tag{5.13}$$

Since there are more than enough equations to estimate $\mathbf{U}$, he suggested replacing the system of equations by s linear combinations of them or increasing the number of parameters by relaxing the equality of some of movement probabilities or by introducing some immigration parameters. Following the first suggestion, Seber proposed the estimator

$$\hat{\mathbf{U}} = \mathbf{D}_{\mathbf{m}}(\mathbf{H}\mathbf{n}')^{-1}\mathbf{H}\mathbf{v}, \tag{5.14}$$

where **H** is an $s \times t$ matrix of rank s such that $(\mathbf{H}\mathbf{n}')$ is nonsingular.

In practice, calculating the estimators (5.10), (5.12) and (5.14) might cause problems. For example, the estimators (5.10) and (5.14) can be influenced strongly by the choice of matrices **G** and **H**. Hence, these estimators can be unappealing unless the matrices **G** and **H** are chosen sensibly. Calculation of (5.12), may be difficult because it requires imposing restrictions (5.11) on unknown survival probabilities. Since ϕ_i's cannot be estimated from our data, extra information is needed to impose these restrictions.

In the next three subsections of this chapter, we closely follow Chapman and Junge, and Darroch, but derive least-squares estimators. We will show that these estimators are special cases of the estimators (5.10) and (5.14). They provide sensible choices for matrices **G** and **H**, and thereby avoid the subjectivity that can arise in using estimators (5.10) and (5.14). They also overcome the difficulties that arise by the requirement of imposing restrictions on unknown survival probabilities.

5.2.1 Assumptions

To derive the least-squares estimators we make the following assumptions.

1. Animals behave independently of one another in regard to moving between strata.
2. All tagged animals released in a given stratum have the same probability distribution of movement to recovery strata.
3. All animals in the jth recovery stratum behave independently in regard to being caught, and have the same (non-zero) probability of being caught in the sample.
4. The matrices $\overline{\mathbf{n}}$, and $\mathbf{n}$ are of full rank. This assumption is required for the consistency of the estimates. In Section 5.5, we show that this assumption can be relaxed under certain conditions.

5. There are negligible numbers of births, deaths, immigration, or emigration, between the two sampling times; i.e.,

   $$\sum_j \theta_{ij} = 1 \quad (i = 1, \ldots, s).$$

   (In situations like the samplings of dead fish at the spawning grounds, this assumption means that the numbers of fish that die before they reach the spawning grounds is negligible.)

   This assumption is required in order to estimate the numbers of animals at the recovery stage. For some animal populations, it may be valid if the time interval between tagging and recovery are short. But, for many animal populations, this assumption is not valid. For example, it is certainly not valid if fish are caught in the interim. In such cases, the investigator has several choices:

   (a) Be content with the estimators of scaled numbers of animals at the recovery stage; e.g. Darroch's estimators (5.8) or (5.12).

   (b) Use additional information to estimate the survival probabilities and using these estimators, unscale Darroch's estimators to get actual numbers at the recovery stage.

   (c) Estimate the numbers at the tagging stage based on Assumption 6 instead.

6. The movement pattern as well as the death and migration rates are the same for tagged animals and untagged animals in a given stratum.

   This assumption is useful in order to derive estimators of the numbers at the tagging stage. It is more reasonable than Assumption 5 for many animal populations. Therefore, we will often make this assumption and estimate the numbers of animals at the tagging stage.

5.2.2 Estimation of the Numbers of Animals in the Recovery Strata ($\overline{\mathbf{V}}$)

Suppose that Assumptions 1-5 hold. Assumptions 2 and 3 imply that

$$\begin{aligned} E(v_j) &= E\left[E(v_j|V_j)\right] = \overline{V}_j p_j, \quad (j = 1, \ldots, t), \quad \text{and} && (5.15) \\ E(n_{ij}) &= m_i \theta_{ij} p_j \;\; (i = 1, \ldots, s \;\; \text{and} \;\; j = 1, 2, \ldots, t). && (5.16) \end{aligned}$$

If an estimator of $\mathbf{p}$ were available, then (5.15) could be used to derive an estimator for $\overline{\mathbf{V}}$. We shall derive estimators for $\mathbf{p}$ from (5.16). First note that the system of equations (5.15) can be written in matrix form as

$$\mathbf{D}_{\overline{\mathbf{v}}} = \mathbf{D}_{\overline{\mathbf{V}}} \mathbf{D}_{\mathbf{p}}. \qquad (5.17)$$

The system (5.16) in matrix form is

$$\overline{\mathbf{n}} = \mathbf{D}_{\mathbf{m}} \boldsymbol{\Theta} \mathbf{D}_{\mathbf{p}}. \qquad (5.18)$$

Assumption 5 implies that

$$\boldsymbol{\Theta} \mathbf{1} = \mathbf{1}. \qquad (5.19)$$

Equations (5.17), (5.18), and (5.19) are the three key equations that we use to derive estimators in this case. Equations (5.18) and (5.19) lead to,

$$\overline{\mathbf{n}} \boldsymbol{\rho} = \overline{\mathbf{n}} \mathbf{D}_{\boldsymbol{\rho}} \mathbf{1} = \mathbf{D}_{\mathbf{m}} \boldsymbol{\Theta} \mathbf{1} = \mathbf{D}_{\mathbf{m}} \mathbf{1} = \mathbf{m}. \qquad (5.20)$$

So, it is reasonable to estimate $\boldsymbol{\rho}$ by a $\widehat{\boldsymbol{\rho}}$ that comes as close as possible to satisfying

$$\mathbf{n} \widehat{\boldsymbol{\rho}} = \mathbf{m}. \qquad (5.21)$$

When $s = t$, under Assumption 4, system (5.21) has a unique solution with high probability. This solution is given by

$$\widehat{\boldsymbol{\rho}} = \mathbf{n}^{-1} \mathbf{m} \;\; \text{when } s = t. \qquad (5.22)$$

When $s > t$, the system (5.21) does not have a solution since it contains s linear equations of t unknowns ρ_j. One can at best attain a good approximation. Under Assumption 4, a least-squares solution that minimizes the sum of squares of errors,

$$\epsilon_i = \sum_{j=1}^{t} n_{ij}\rho_j - m_i$$

is

$$\widehat{\boldsymbol{\rho}} = \left[(\mathbf{n}'\mathbf{n})^{-1}\mathbf{n}'\right]\mathbf{m} \text{ when } s \geq t. \tag{5.23}$$

Now, note that from (5.17),

$$\overline{\mathbf{V}} = \mathbf{D}_{\overline{\mathbf{V}}}\mathbf{1} = \mathbf{D}_{\overline{\mathbf{v}}}\mathbf{D}_{\mathbf{p}}^{-1}\mathbf{1} = \mathbf{D}_{\overline{\mathbf{v}}}\mathbf{D}_{\boldsymbol{\rho}}\mathbf{1} = \mathbf{D}_{\overline{\mathbf{v}}}\boldsymbol{\rho}. \tag{5.24}$$

By replacing $\overline{\mathbf{v}}$ and $\boldsymbol{\rho}$ by $\mathbf{v}$ and $\widehat{\boldsymbol{\rho}}$ respectively, an estimator of $\overline{\mathbf{V}}$ can be derived as

$$\widehat{\overline{\mathbf{V}}} = \begin{cases} \mathbf{D}_{\mathbf{v}}\mathbf{n}^{-1}\mathbf{m} & \text{when } s = t \\ \mathbf{D}_{\mathbf{v}}\left[(\mathbf{n}'\mathbf{n})^{-1}\mathbf{n}'\right]\mathbf{m} & \text{when } s > t. \end{cases} \tag{5.25}$$

Consequently, $\overline{V}_+$ can be estimated by

$$\widehat{\overline{V}}_+ = \mathbf{1}'\overline{\mathbf{V}} = \begin{cases} \mathbf{v}'\mathbf{n}^{-1}\mathbf{m} & \text{when } s = t \\ \mathbf{v}'\left[(\mathbf{n}'\mathbf{n})^{-1}\mathbf{n}'\right]\mathbf{m} & \text{when } s > t. \end{cases} \tag{5.26}$$

When $s < t$, $\boldsymbol{\rho}$ cannot be estimated from the system (5.21). Therefore, $\overline{\mathbf{V}}$ also cannot be estimated. However, as Chapman and Junge [7], and Darroch [13] pointed out, an estimator of U is available if Assumption 6 is satisfied. This estimator is derived in the next subsection.

5.2.3 Estimation of the Numbers of Animals in the Tagging Strata (U)

Suppose that Assumptions 1-4 and 6 hold. Under Assumption 6,

$$E(v_j) = \sum_i U_i \theta_{ij} p_j \;\; (j = 1, \ldots, t),$$

or in matrix notation,

$$\overline{\mathbf{v}}' = \mathbf{U}'\boldsymbol{\Theta}\mathbf{D}_{\mathbf{p}} = \mathbf{U}'\boldsymbol{\psi}. \tag{5.27}$$

This is the first key equation that we use in this case to derive estimators. The second key equation is (5.18). That is,

$$\overline{\mathbf{n}} \;=\; \mathbf{D}_{\mathbf{m}}\boldsymbol{\Theta}\mathbf{D}_{\mathbf{p}}. \tag{5.28}$$

Now, obtaining

$$\boldsymbol{\psi} = \boldsymbol{\Theta}\mathbf{D}_{\mathbf{p}} = \mathbf{D}_{\mathbf{m}}^{-1}\overline{\mathbf{n}} \tag{5.29}$$

from (5.28), substituting in (5.27), and taking the transpose, yields

$$\overline{\mathbf{v}} \;=\; \overline{\mathbf{n}}'\mathbf{D}_{\mathbf{m}}^{-1}\mathbf{U}. \tag{5.30}$$

Now, let

$$\boldsymbol{\zeta} = \mathbf{v} - \mathbf{n}'\mathbf{D}_{\mathbf{m}}^{-1}\mathbf{U}.$$

We propose to estimate $\mathbf{U}$ by $\hat{\mathbf{U}}$ that minimizes the sum of squares of the errors,

$$\zeta_j = v_j - \sum_{i=1}^{s} n_{ij}\hat{U}_i/m_i. \tag{5.31}$$

Under Assumption 4, this estimator is given by

$$\hat{\mathbf{U}} = \begin{cases} \mathbf{D}_{\mathbf{m}}[(\mathbf{n}\mathbf{n}')^{-1}\mathbf{n}]\mathbf{v} & \text{when } s < t \\ \mathbf{D}_{\mathbf{m}}(\mathbf{n}')^{-1}\mathbf{v} & \text{when } s = t. \end{cases} \tag{5.32}$$

The resulting estimators of the total untagged animals are,

$$\hat{U}_+ = \hat{\mathbf{U}}'\mathbf{1} = \begin{cases} \mathbf{v}'[\mathbf{n}'(\mathbf{n}\mathbf{n}')^{-1}]\mathbf{m} & \text{when } s < t \\ \mathbf{v}'(\mathbf{n})^{-1}\mathbf{m} & \text{when } s = t\ . \end{cases} \tag{5.33}$$

5.2.4 Some Remarks

Remark 1

Suppose $\mathbf{\Theta 1} \neq \mathbf{1}$, but $\phi_i \equiv \phi \ \forall\, i$. Then, following Darroch, we can define scaled parameters $\theta^\star_{ij} = \theta_{ij}/\phi$, $p^\star_j = \phi p_j$, and $\overline{V}^\star_j = \overline{V}_j/\phi$. Then, the new parameters satisfy key equations (5.17), (5.18) and (5.19). Therefore, (5.25) now estimates $\overline{\mathbf{V}}^\star$. When $s = t$, estimator (5.25) is essentially the same as Darroch's estimator (5.8). However, this is not particularly useful unless the survival probabilities can be estimated.

Remark 2

The estimators (5.25) and (5.32) are special cases of (5.10) and (5.14) respectively. These suggest that $\mathbf{G} = \mathbf{n}'$ and $\mathbf{H} = \mathbf{n}$ are sensible choices to be used in (5.10) and (5.14) respectively.

Remark 3

According to Searle [31], the Moore-Penrose inverse $\mathbf{n}^\dagger$ of $\mathbf{n}$ is a unique matrix such that

(i) $\mathbf{n}\mathbf{n}^\dagger\mathbf{n} = \mathbf{n}$,

(ii) $\mathbf{n}^\dagger\mathbf{n}\mathbf{n}^\dagger = \mathbf{n}^\dagger$,

(iii) $\mathbf{n}\mathbf{n}^\dagger$ is symmetric, and

(iv) $\mathbf{n}^\dagger\mathbf{n}$ is symmetric.

One can show (e.g., Albert [1], pp. 20-21) by direct verification of (i)-(iv) that if $\mathbf{n}$ is of full rank, then

$$\mathbf{n}^\dagger = \begin{cases} \mathbf{n}'(\mathbf{n}\mathbf{n}')^{-1} & \text{when } s < t \\ \mathbf{n}^{-1} & \text{when } s = t \\ (\mathbf{n}'\mathbf{n})^{-1}\mathbf{n}' & \text{when } s > t. \end{cases} \tag{5.34}$$

Therefore, the estimators derived above can be summarized as follows.

$$\left.\begin{array}{lcll}\hat{\rho} & = & \mathbf{n}^{\dagger}\mathbf{m} & \text{when } s \geq t, \\ \hat{\mathbf{V}} & = & \mathbf{D_V n^{\dagger} m} & \text{when } s \geq t, \\ \hat{V}_{+} & = & \mathbf{v}'\mathbf{n}^{\dagger}\mathbf{m} & \text{when } s \geq t, \\ \hat{\mathbf{U}} & = & \mathbf{D_m}(\mathbf{n}')^{\dagger}\mathbf{v} & \text{when } s \leq t, \text{ and} \\ \hat{U}_{+} & = & \mathbf{v}'\mathbf{n}^{\dagger}\mathbf{m} & \text{when } s \leq t.\end{array}\right\} \tag{5.35}$$

Remark 4

In deriving the least squares estimates $\hat{\mathbf{V}}$ and $\hat{\mathbf{U}}$, we have neglected the variance-covariance structures of ϵ_i's and ζ_js. Better estimates may be derived by taking these into account.

In order to calculate the variance-covariance matrix of ϵ, we should adopt a suitable probability model. Under Assumptions 1-3, probability distributions (5.4) and (5.5) provide such a model. For this model, the variance of $\epsilon_i = \sum_{j=1}^{t} n_{ij}\rho_j - m_i$ is

$$\sigma_i^2 = m_i \sum_{j=1}^{t} \rho_j^2 \psi_{ij}(1-\psi_{ij}) - 2m_i \sum_{j=1}^{t-1} \sum_{k=j+1}^{t} \rho_j \rho_k \psi_{ij} \psi_{ik}. \tag{5.36}$$

The covariance terms are zero; i.e, $Cov(\epsilon_i, \epsilon_l) = 0 \ \ \forall i \neq l$. Weighted least-squares (WLS) estimators may be derived by minimizing

$$S = \sum_{i=1}^{s} (\epsilon_i^2/\tilde{\sigma}_i^2), \tag{5.37}$$

where

$$\tilde{\sigma}_i^2 = m_i \sum_{j=1}^{t} \rho_j^2 \hat{\psi}_{ij}(1-\hat{\psi}_{ij}) - 2m_i \sum_{j=1}^{t-1} \sum_{k=j+1}^{t} \rho_j \rho_k \hat{\psi}_{ij} \hat{\psi}_{ik},$$

with $\hat{\psi}_{ij}$'s defined by

$$\hat{\psi} = \mathbf{D}_{\mathbf{m}}^{-1}\mathbf{n}. \tag{5.38}$$

These estimates can be obtained by solving the estimating equations,

$$\frac{\partial S}{\partial \rho} = 0.$$

Since the weights are functions of unknown ρ_j's, these estimating equations are biased. However, an estimate which is unbiased conditional on $\mathbf{n}$, can be derived using an iteratively reweighted least-squares method.

Let $\Sigma(\boldsymbol{\rho})$ be the diagonal matrix with diagonal elements $\tilde{\sigma}_i^2$'s. Suppose that $\boldsymbol{\rho}_0$ is an initial guess for $\boldsymbol{\rho}$. So, at this point, the initial guess of the variance-covariance matrix is given by $\Sigma(\boldsymbol{\rho}_0)$. Let

$$\begin{aligned} \mathbf{z}_0 &= \mathbf{m} - \mathbf{n}\boldsymbol{\rho}_0, \text{ and} \\ S_0 &= [\mathbf{m} - \mathbf{n}\boldsymbol{\rho}]' \, \Sigma^{-1}(\boldsymbol{\rho}_0) \, [\mathbf{m} - \mathbf{n}\boldsymbol{\rho}]. \end{aligned}$$

Now, we minimize S_0 with respect to $\boldsymbol{\rho}$. The requirement, $\frac{\partial S_0(\widehat{\boldsymbol{\rho}})}{\partial \boldsymbol{\rho}} = 0$, leads to

$$\begin{aligned} \mathbf{n}'\Sigma^{-1}(\boldsymbol{\rho}_0)\mathbf{n}\widehat{\boldsymbol{\rho}} &= \mathbf{n}'\Sigma^{-1}(\boldsymbol{\rho}_0)\mathbf{m} \qquad (5.39) \\ &= \mathbf{n}'\Sigma^{-1}(\boldsymbol{\rho}_0)\,[\mathbf{z}_0 + \mathbf{n}\boldsymbol{\rho}_0], \\ \mathbf{n}'\Sigma^{-1}(\boldsymbol{\rho}_0)\mathbf{n}\,[\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0] &= \mathbf{n}'\Sigma^{-1}(\boldsymbol{\rho}_0)\mathbf{z}_0, \text{ and hence} \\ \widehat{\boldsymbol{\rho}} &= \boldsymbol{\rho}_0 + \left[\mathbf{n}'\Sigma^{-1}(\boldsymbol{\rho}_0)\mathbf{n}\right]^{-1} \mathbf{n}'\Sigma^{-1}(\boldsymbol{\rho}_0)\mathbf{z}_0. \end{aligned}$$

This gives rise to the iterative equation,

$$\widehat{\boldsymbol{\rho}}_{k+1} = \widehat{\boldsymbol{\rho}}_k + \left[\mathbf{n}'\Sigma^{-1}(\widehat{\boldsymbol{\rho}}_k)\mathbf{n}\right]^{-1} \mathbf{n}'\Sigma^{-1}(\widehat{\boldsymbol{\rho}}_k)\mathbf{z}_k, \qquad (5.40)$$

with $\mathbf{z}_k$ defined as

$$\mathbf{z}_k = \mathbf{m} - \mathbf{n}\widehat{\boldsymbol{\rho}}_k.$$

I.e.,

$$\widehat{\boldsymbol{\rho}}_{k+1} = \widehat{\boldsymbol{\rho}}_k + \left[\mathbf{n}'\Sigma^{-1}(\widehat{\boldsymbol{\rho}}_k)\mathbf{n}\right]^{-1} \mathbf{n}'\Sigma^{-1}(\widehat{\boldsymbol{\rho}}_k)\,[\mathbf{m} - \mathbf{n}\widehat{\boldsymbol{\rho}}_k] \quad k = 0, 1, 2, \ldots. \qquad (5.41)$$

This procedure typically does not minimize S. But it converges, when it converges, to a root of the equation (5.39) with $\boldsymbol{\rho}_0$ replaced by $\boldsymbol{\rho}$; i.e.,

$$\mathbf{n}'\Sigma^{-1}(\boldsymbol{\rho})\mathbf{n}\widehat{\boldsymbol{\rho}} = \mathbf{n}'\Sigma^{-1}(\boldsymbol{\rho})\mathbf{m}.$$

This estimating equation is unbiased conditional on $\mathbf{n}$ under the model which justifies (5.36).

In order to derive a similar estimate for $\mathbf{U}$, we need the variance-covariance structure of $\boldsymbol{\zeta}$. Recall that the estimator $\hat{\mathbf{U}}$ was derived under Assumptions 1-4 and 6. Under these assumptions, it is reasonable to assume a probability model for the unmarked animals that is parallel to (5.4). I.e,

$$p[\{v_{ij}\}|\{U_i\}] = \frac{\prod_i U_i!}{\prod_i (U_i - v_{i+})! \prod_{ij} v_{ij}!} \prod_i (1 - \sum_j \theta_{ij} p_j)^{(U_i - v_{i+})} \prod_{ij} (\theta_{ij} p_j)^{v_{ij}}, \qquad (5.42)$$

where

v_{ij} = the number of unmarked animals which are in the ith stratum at the tagging stage and in the jth recovery sample.

Then, from (5.31)

$$\begin{aligned} \zeta_j &= v_j - \sum_{i=1}^{s} n_{ij} U_i / m_i \\ &= \sum_{i=1}^{s} (v_{ij} - n_{ij} U_i / m_i). \end{aligned}$$

Therefore, under probability models (5.5) and (5.42),

$$\begin{aligned} Var(\zeta_j) &= \gamma_{jj} = \sum_{i=1}^{s} \left(U_i + U_i^2 / m_i \right) \psi_{ij} (1 - \psi_{ij}), \text{ and} \\ Cov(\zeta_j, \zeta_k) &= \gamma_{jk} = - \sum_{i=1}^{s} \left(U_i + U_i^2 / m_i \right) \psi_{ij} \psi_{ik}, \text{ when } j \neq k. \end{aligned}$$

Let $\Sigma(\mathbf{U})$ be the matrix (γ_{ij}), but with the ψ_{ij}'s replaced by estimators given by (5.38). Now, suppose that $\mathbf{U}_0$ is a suitable initial guess for $\mathbf{U}$. Then, by similar arguments to those above, the iterative procedure,

$$\hat{\mathbf{U}}_{k+1} = \hat{\mathbf{U}}_k + \mathbf{D_m} \left[\mathbf{n} \Sigma^{-1} (\hat{\mathbf{U}}_k) \mathbf{n}' \right]^{-1} \mathbf{n} \Sigma^{-1} (\hat{\mathbf{U}}_k) \left[\mathbf{v} - \mathbf{n}' \mathbf{D_m^{-1}} \hat{\mathbf{U}}_k \right], \; k = 0, 1, 2, \ldots \qquad (5.43)$$

can be used to derive an estimate for $\mathbf{U}$.

In the subsequent sections, we refer to the estimators derived via iterative procedures (5.41) and (5.43) as iteratively reweighted least squares estimates (IRLSE).

Remark 5

When $s \geq t$, by premultiplying (5.20) by $\overline{\mathbf{n}}'$, and performing simple matrix manipulation under Assumption 4, $\boldsymbol{\rho}$ can be expressed as

$$\boldsymbol{\rho} \;=\; [(\overline{\mathbf{n}}'\,\overline{\mathbf{n}}\,)^{-1}\overline{\mathbf{n}}'\,]\mathbf{m}. \tag{5.44}$$

Substitution of (5.44) in (5.24) yields

$$\overline{\mathbf{V}} = \mathbf{D}_{\overline{\mathbf{v}}}[(\overline{\mathbf{n}}'\,\overline{\mathbf{n}}\,)^{-1}\overline{\mathbf{n}}'\,]\mathbf{m} \text{ when } s \geq t\ . \tag{5.45}$$

Consequently,

$$\overline{V}_{+} = \mathbf{1}'\overline{\mathbf{V}} = \overline{\mathbf{v}}'[(\overline{\mathbf{n}}'\,\overline{\mathbf{n}}\,)^{-1}\overline{\mathbf{n}}'\,]\mathbf{m} \text{ when } s \geq t\ . \tag{5.46}$$

Similarly, when $s \leq t$, by premultiplying (5.30) by $\overline{\mathbf{n}}$, and performing simple matrix manipulation under Assumption 4, $\mathbf{U}$ can be written as

$$\mathbf{U} \;=\; \mathbf{D}_{\mathbf{m}}[(\overline{\mathbf{n}}\,\overline{\mathbf{n}}'\,)^{-1}\overline{\mathbf{n}}]\overline{\mathbf{v}}. \tag{5.47}$$

Hence,

$$U_{+} = \mathbf{U}'\mathbf{1} \;=\; \overline{\mathbf{v}}'[\overline{\mathbf{n}}'(\overline{\mathbf{n}}\,\overline{\mathbf{n}}'\,)^{-1}]\mathbf{m}. \tag{5.48}$$

Now, note that under Assumption 4,

$$\overline{\mathbf{n}}^{\dagger} = \begin{cases} \overline{\mathbf{n}}'(\overline{\mathbf{n}}\,\overline{\mathbf{n}}')^{-1} & \text{when } s < t \\ \overline{\mathbf{n}}^{-1} & \text{when } s = t \\ (\overline{\mathbf{n}}'\overline{\mathbf{n}})^{-1}\overline{\mathbf{n}}' & \text{when } s > t. \end{cases} \tag{5.49}$$

Therefore, the above quantities can be written as

$$\left.\begin{array}{lcll} \boldsymbol{\rho} & = & \overline{\mathbf{n}}^{\dagger}\mathbf{m} & \text{when } s \geq t, \\ \overline{\mathbf{V}} & = & \mathbf{D}_{\overline{\mathbf{v}}}\overline{\mathbf{n}}^{\dagger}\mathbf{m} & \text{when } s \geq t, \\ \overline{V}_{+} & = & \overline{\mathbf{v}}'\overline{\mathbf{n}}^{\dagger}\mathbf{m} & \text{when } s \geq t, \\ \mathbf{U} & = & \mathbf{D}_{\mathbf{m}}(\overline{\mathbf{n}}')^{\dagger}\overline{\mathbf{v}} & \text{when } s \leq t, \text{ and} \\ U_{+} & = & \overline{\mathbf{v}}'\overline{\mathbf{n}}^{\dagger}\mathbf{m} & \text{when } s \leq t. \end{array}\right\} \tag{5.50}$$

These representations are useful in proving the consistency of the estimators in Section 5.3, and calculating the variances in Section 5.4.

5.3 Consistency

As Chapman and Junge pointed out, the consistency of estimators based on samples from finite populations has been variously defined. According to one such usage, an estimator $\hat{\lambda}$ of parameter λ would be called consistent if $\hat{\lambda} = \lambda$ whenever the sample taken without replacement exhausts the population. However, from a practical point of view, in a study of populations that number several hundred thousands or millions (e.g. salmon populations), it is unreasonable to think of a sample equaling or nearly equaling the population size. Yet, at the same time, it is possible that the samples are random and very large so that the weak law of large numbers should be applicable. Hence we consider the limit process: $\{\theta_{ij}\}$, $\{p_j\}$ constant, all $m_i \rightarrow \infty$ and $V_i \rightarrow \infty$ such that m_i/m_+, V_j/V_+ and m_+/V_+ approach fixed constants. We say that $\widehat{\overline{\mathbf{V}}}$ is a consistent estimator of $\overline{\mathbf{V}}$ if under these conditions, $\widehat{\overline{V}}_j/\overline{V}_j \rightarrow 1$ in probability for all j. Similarly, we say that $\hat{\mathbf{U}}$ is consistent for $\mathbf{U}$ if $\hat{U}_i/U_i \rightarrow 1$ in probability for all i.

First notice that under probability model (5.4), and in the above limit process, $n_{ij}/\bar{n}_{ij} \rightarrow 1$ in probability, for all i and j. In other words, $\mathbf{n}$ is consistent for $\bar{\mathbf{n}}$. The matrix product is a continuous function and the matrix inverse is locally continuous in a neighborhood of an invertible matrix. Therefore, under Assumption 4, the Slutsky-Frechet theorem (see Appendix) implies that the Moore-Penrose inverse $\mathbf{n}^{\dagger}$ of $\mathbf{n}$ is consistent for the Moore-Penrose inverse $\bar{\mathbf{n}}^{\dagger}$ of $\bar{\mathbf{n}}$. This, in turn implies that $\hat{\boldsymbol{\rho}} = \mathbf{n}^{\dagger}\mathbf{m}$ is consistent for $\bar{\mathbf{n}}^{\dagger}\mathbf{m}$, which by (5.50) is equal to $\boldsymbol{\rho}$.

Next, notice that under probability model (5.5) and the above limit process, $\mathbf{v}$ is consistent for $\bar{\mathbf{v}}$. Further notice that the only random variable involved in $\hat{\boldsymbol{\rho}}$ is $\mathbf{n}$. Since $\mathbf{v}$ and $\mathbf{n}$ are independent by Assumptions 1 and 3, so are $\mathbf{v}$ and $\hat{\boldsymbol{\rho}}$. Therefore, $\widehat{\overline{\mathbf{V}}} = \mathbf{D}_{\mathbf{v}}\mathbf{n}^{\dagger}\mathbf{m}$ is consistent for $\mathbf{D}_{\bar{\mathbf{v}}}\bar{\mathbf{n}}^{\dagger}\mathbf{m}$, which by (5.50) is equal to $\overline{\mathbf{V}}$. A similar argument will show that $\hat{\mathbf{U}}$ is consistent for $\mathbf{U}$.

5.4 Variance

Darroch used probability distributions (5.4) and (5.5), to derive the formulae for the approximate variances of his estimators. Accordingly, when $s = t$,

$$\begin{aligned} Var(\widehat{\boldsymbol{\rho}^{\star}}) = \boldsymbol{\Sigma} &\approx \overline{\mathbf{n}}^{-1}\mathbf{D}_{\mathbf{m}}\mathbf{D}_{\boldsymbol{\mu}}(\overline{\mathbf{n}}^{-1})', & (5.51)\\ Var(\widehat{\mathbf{V}^{\star}}) &\approx \mathbf{D}_{\overline{\mathbf{v}}}\boldsymbol{\Sigma}\mathbf{D}_{\overline{\mathbf{v}}} + \mathbf{D}_{\mathbf{V}}(\mathbf{D}_{\boldsymbol{\rho}} - \mathbf{I}), \quad \text{and} & (5.52)\\ Var(\widehat{V_{+}^{\star}}) &\approx \overline{\mathbf{v}}'\boldsymbol{\Sigma}\overline{\mathbf{v}} + \mathbf{V}'(\boldsymbol{\rho} - \mathbf{1}). & (5.53)\end{aligned}$$

Here, $\mu_i = \sum_j \theta_{ij}\rho_j - 1$.

In the next two subsections, we derive formulae for the variances of estimators (5.35).

5.4.1 Variance of $\widehat{\overline{\mathbf{V}}}$

Recall from (5.50), (5.34) and (5.35) that when $s \geq t$,

$$\begin{aligned} \overline{\mathbf{V}} &= \mathbf{D}_{\overline{\mathbf{v}}}\boldsymbol{\rho}, \text{ and} & (5.54)\\ \widehat{\overline{\mathbf{V}}} &= \mathbf{D}_{\mathbf{v}}\widehat{\boldsymbol{\rho}} = \mathbf{D}_{\mathbf{v}}\mathbf{n}^{\dagger}\mathbf{m} = \mathbf{D}_{\mathbf{v}}\left[(\mathbf{n}'\mathbf{n})^{-1}\mathbf{n}'\right]\mathbf{m}. \end{aligned}$$

Now let

$$\mathbf{y} = (\mathbf{v} - \overline{\mathbf{v}}). \tag{5.55}$$

Then,

$$\widehat{\overline{\mathbf{V}}} - \overline{\mathbf{V}} = \mathbf{D}_{\overline{\mathbf{v}}}(\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}) + \mathbf{D}_{\mathbf{y}}\widehat{\boldsymbol{\rho}}. \tag{5.56}$$

The only random variable involved in $\widehat{\boldsymbol{\rho}}$ is $\mathbf{n}$. Since $\mathbf{v}$ and $\mathbf{n}$ are independent by Assumptions 1 and 3, so are $\mathbf{y}$ and $\widehat{\boldsymbol{\rho}}$. Furthermore, $E(\mathbf{y}) = 0$. Therefore, the bias of $\widehat{\overline{\mathbf{V}}}$ is

$$E(\widehat{\overline{\mathbf{V}}} - \overline{\mathbf{V}}) = \mathbf{D}_{\overline{\mathbf{v}}}E(\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}). \tag{5.57}$$

The mean squared error of $\widehat{\overline{\mathbf{V}}}$ is

$$\begin{aligned} E[(\widehat{\overline{\mathbf{V}}} - \overline{\mathbf{V}})(\widehat{\overline{\mathbf{V}}} - \overline{\mathbf{V}})'] &= \mathbf{D}_{\overline{\mathbf{v}}}E\left[(\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho})(\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho})'\right]\mathbf{D}_{\overline{\mathbf{v}}} + E\left[\mathbf{D}_{\mathbf{y}}\widehat{\boldsymbol{\rho}}\widehat{\boldsymbol{\rho}}'\mathbf{D}_{\mathbf{y}}\right]\\ &\quad + 2\mathbf{D}_{\overline{\mathbf{v}}}E\left[(\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho})\widehat{\boldsymbol{\rho}}'\mathbf{D}_{\mathbf{y}}\right]. \end{aligned}$$

Again, because $\mathbf{y}$ and $\hat{\boldsymbol{\rho}}$ are independent and $E(\mathbf{y}) = 0$, the last term is zero. Hence,

$$E[(\widehat{\overline{\mathbf{V}}} - \overline{\mathbf{V}})(\widehat{\overline{\mathbf{V}}} - \overline{\mathbf{V}})'] = \mathbf{D}_{\overline{\mathbf{V}}} E\left[(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho})(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho})'\right] \mathbf{D}_{\overline{\mathbf{V}}} + E\left[\mathbf{D}_{\mathbf{y}} \hat{\boldsymbol{\rho}} \hat{\boldsymbol{\rho}}' \mathbf{D}_{\mathbf{y}}\right]. \quad (5.58)$$

We need the bias and the mean squared error of $\hat{\boldsymbol{\rho}}$ to calculate those of $\widehat{\overline{\mathbf{V}}}$. So, let us calculate those quantities first.

Calculation of bias and mean squared error of $\hat{\boldsymbol{\rho}}$

This calculation contains the following steps.

Step 1. Recall that when $s \geq t$,

$$\begin{aligned} \hat{\boldsymbol{\rho}} &= \mathbf{n}^{\dagger}\mathbf{m} = \left[(\mathbf{n}'\mathbf{n})^{-1}\mathbf{n}'\right] \mathbf{m}, \text{ and} \\ \boldsymbol{\rho} &= \overline{\mathbf{n}}^{\dagger}\mathbf{m} = \left[(\overline{\mathbf{n}}'\overline{\mathbf{n}})^{-1}\overline{\mathbf{n}}'\right] \mathbf{m}. \end{aligned} \quad (5.59)$$

Step 2. Let

$$\mathbf{n} = \overline{\mathbf{n}} + \mathbf{X}. \quad (5.60)$$

Show that

$$\hat{\boldsymbol{\rho}} = \boldsymbol{\rho} - \overline{\mathbf{n}}^{\dagger}\mathbf{X}\boldsymbol{\rho} + \text{ a random quantity with expected value } O(m_+^{-1}). \quad (5.61)$$

Thereby show that the bias of $\hat{\boldsymbol{\rho}}$ is $O(m_+^{-1})$. The expression (5.61) is obtained by expanding the inverse matrix in (5.59) using the following identity:

If $\mathbf{A} = \bar{\mathbf{A}} + \mathbf{Z}$ and both $\mathbf{A}$ and $\bar{\mathbf{A}}$ are nonsingular, then,

$$\mathbf{A}^{-1} = [\mathbf{I} - \bar{\mathbf{A}}^{-1}\mathbf{Z}]\bar{\mathbf{A}}^{-1} + (\bar{\mathbf{A}}^{-1}\mathbf{Z})^2\mathbf{A}^{-1}. \quad (5.62)$$

Step 3. Show that the leading term in the approximate mean squared error is

$$\boldsymbol{\Sigma} = \overline{\mathbf{n}}^{\dagger} E[\mathbf{X}\boldsymbol{\rho}\boldsymbol{\rho}'\mathbf{X}'](\overline{\mathbf{n}}^{\dagger})', \quad (5.63)$$

which is $O(m_+^{-1})$. Hence, the bias is asymptotically negligible compared to the root mean squared error, which is $O(m_+^{-1/2})$, and (5.63) can be considered as the approximate variance-covariance matrix of $\hat{\boldsymbol{\rho}}$.

Now, we present the detailed calculation in Step 2. Let $\mathbf{X}$ be defined as in (5.60). Then,

$$\begin{aligned}
\mathbf{n}'\mathbf{n} &= (\bar{\mathbf{n}}' + \mathbf{X}')(\bar{\mathbf{n}} + \mathbf{X}) \\
&= \bar{\mathbf{n}}'\bar{\mathbf{n}} + \bar{\mathbf{n}}'\mathbf{X} + \mathbf{X}'\bar{\mathbf{n}} + \mathbf{X}'\mathbf{X} \\
&= \boldsymbol{\Gamma} + \mathbf{Z},
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\Gamma} &= \bar{\mathbf{n}}'\bar{\mathbf{n}}, \text{and} && (5.64) \\
\mathbf{Z} &= \bar{\mathbf{n}}'\mathbf{X} + \mathbf{X}'\bar{\mathbf{n}} + \mathbf{X}'\mathbf{X}. && (5.65)
\end{aligned}$$

Hence, substitution of $\mathbf{A} = (\mathbf{n}'\mathbf{n})$ and $\bar{\mathbf{A}} = \boldsymbol{\Gamma}$ in (5.62) yields

$$(\mathbf{n}'\mathbf{n})^{-1} = [\mathbf{I} - \boldsymbol{\Gamma}^{-1}\mathbf{Z}]\boldsymbol{\Gamma}^{-1} + (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^2(\mathbf{n}'\mathbf{n})^{-1}. \tag{5.66}$$

Therefore,

$$\mathbf{n}^{\dagger} = (\mathbf{n}'\mathbf{n})^{-1}\mathbf{n}' = [\mathbf{I} - \boldsymbol{\Gamma}^{-1}\mathbf{Z}]\boldsymbol{\Gamma}^{-1}(\bar{\mathbf{n}}' + \mathbf{X}') + (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^2\mathbf{n}^{\dagger}. \tag{5.67}$$

Post-multiplying (5.67) by $\mathbf{m}$ and substituting $\bar{\mathbf{n}}'\mathbf{X} + \mathbf{X}'\bar{\mathbf{n}} + \mathbf{X}'\mathbf{X}$ for $\mathbf{Z}$ in the square bracket, leads to

$$\begin{aligned}
\hat{\boldsymbol{\rho}} &= \boldsymbol{\Gamma}^{-1}\bar{\mathbf{n}}'\mathbf{m} + \boldsymbol{\Gamma}^{-1}\mathbf{X}'\mathbf{m} - \boldsymbol{\Gamma}^{-1}(\bar{\mathbf{n}}'\mathbf{X} + \mathbf{X}'\bar{\mathbf{n}})\boldsymbol{\Gamma}^{-1}\bar{\mathbf{n}}'\mathbf{m} - \mathbf{b} + \mathbf{r} && (5.68) \\
&= \boldsymbol{\rho} + \boldsymbol{\Gamma}^{-1}\mathbf{X}'[\mathbf{m} - \bar{\mathbf{n}}\boldsymbol{\rho}] - \bar{\mathbf{n}}^{\dagger}\mathbf{X}\boldsymbol{\rho} - \mathbf{b} + \mathbf{r} \\
&= \boldsymbol{\rho} - \bar{\mathbf{n}}^{\dagger}\mathbf{X}\boldsymbol{\rho} - \mathbf{b} + \mathbf{r}, && (5.69)
\end{aligned}$$

where

$$\mathbf{b} = \boldsymbol{\Gamma}^{-1}(\bar{\mathbf{n}}'\mathbf{X} + \mathbf{X}'\bar{\mathbf{n}})\boldsymbol{\Gamma}^{-1}\mathbf{X}'\mathbf{m} + \boldsymbol{\Gamma}^{-1}(\mathbf{X}'\mathbf{X})\boldsymbol{\Gamma}^{-1}(\bar{\mathbf{n}}' + \mathbf{X}')\mathbf{m}, \tag{5.70}$$

and

$$\mathbf{r} = (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^2\hat{\boldsymbol{\rho}}. \tag{5.71}$$

Since $E(\mathbf{X}) = 0$, (5.69) implies that the bias, $E(\hat{\boldsymbol{\rho}}) - \boldsymbol{\rho} = E(-\mathbf{b} + \mathbf{r})$. Now, we show that this bias is $O(m_{+}^{-1})$. For this, consider the limit process: $\{\theta_{ij}\}$, $\{p_j\}$ constant and all $m_i \to \infty$ in such a way that m_i/m_+ approach fixed constants. First, note that $\overline{\mathbf{n}} = O(m_+), \mathbf{\Gamma}^{-1} = O(m_+^{-2})$ in this limit process, and $\mathbf{m} = O(m_+)$.

Next note that since $(n_{i1}, n_{i2}, \ldots, n_{it})$ are multinomial random variables, $E(X_{ij}X_{lk}) = \delta_{il}O(m_+)$ and $E(X_{ij}X_{lk}X_{ab}) = \delta_{il}\delta_{ia}O(m_+)$. Hence,

$$E(\mathbf{b}) = O(m_+^{-1}). \tag{5.72}$$

Now, in order to find out the order of $E(\mathbf{r})$, let

$$\mathbf{F} \ = \ (F_{ij}) = (\mathbf{\Gamma}^{-1}\mathbf{Z})^2.$$

Then,

$$r_j \ = \ \sum_k F_{jk}\widehat{\rho}_k.$$

Using Jensen's inequality and Cauchy-Schwarz inequality (see Appendix), we can see that

$$\begin{aligned} E(r_j) \le |E(r_j)| &= \left| E\left(\sum_k F_{jk}\widehat{\rho}_k\right)\right| \\ &\le \sum_k \left|E\left(F_{jk}\widehat{\rho}_k\right)\right| \\ &\le \sum_k \left|E\left(F_{jk}^2\widehat{\rho}_k^2\right)\right|^{\frac{1}{2}} \\ &\le \sum_k \left[E(F_{jk}^2)\, E(\widehat{\rho}_k^2)\right]^{\frac{1}{2}}. \end{aligned}$$

In F_{ij}, the contribution from $\mathbf{\Gamma}^{-1}$ is $O(m_+^{-4})$. So, its contribution to $E(F_{ij}^2)$ is $O(m_+^{-8})$. The terms in $\mathbf{Z}$ that contribute the largest order terms to $E(F_{jk}^2)$ are $\overline{\mathbf{n}}'\mathbf{X} + \mathbf{X}'\overline{\mathbf{n}}$. So, the highest order of moments of $\mathbf{X}$ generated in F_{ij}^2 by these terms is 4. Since these are multinomial moments, the maximum contribution is $O(m_+^2)$. Finally, the highest contribution of terms $\overline{\mathbf{n}}$ that come from $\mathbf{Z}$ is $O(m_+^4)$. Therefore, the terms $E[F_{jk}^2]$ are $O(m_+^{-8+2+4}) = O(m_+^{-2})$. Furthermore, the order of $E(\widehat{\rho}_k^2)$ is the same as that of ρ_k^2, which is $O(1)$. Therefore,

$$E(\mathbf{r}) = \left[O(m_+^{-2})\right]^{1/2} = O(m_+^{-1}). \tag{5.73}$$

Order expressions, (5.72) and (5.73) imply that the bias of $\widehat{\boldsymbol{\rho}}$ is $O(m_+^{-1})$.

Now, we proceed to Step 3. From (5.69), the leading term in $E\left[(\hat{\rho}-\rho)(\hat{\rho}-\rho)')\right]$ is

$$\mathbf{\Sigma} = \bar{\mathbf{n}}^{\dagger} E[\mathbf{X}\rho\rho'\mathbf{X}'](\bar{\mathbf{n}}^{\dagger})'. \tag{5.74}$$

Note that $\bar{\mathbf{n}}^{\dagger} = O(m_+^{-1})$, second order moments of $\mathbf{X}$ are $O(m_+)$ and $\rho = O(1)$. Hence, $\mathbf{\Sigma}$ is $O(m_+^{-1})$, and the bias of $\hat{\rho}$, which is $O(m_+^{-1})$, is negligible compared to the root mean squared error. Therefore (5.74) is the approximate variance of $\hat{\rho}$.

Calculation of the mean squared error of $\hat{\bar{\mathbf{V}}}$

First, in order to find the approximate bias of $\hat{\bar{\mathbf{V}}}$, augment the above limit process by supposing that all $V_j \to \infty$ in such a way that V_j/V_+ and m_+/V_+ approach fixed constants. Equation (5.57) gives the bias of $\hat{\bar{\mathbf{V}}}$ as

$$E(\hat{\bar{\mathbf{V}}} - \bar{\mathbf{V}}) = \mathbf{D}_{\bar{\mathbf{V}}} E[\hat{\rho} - \rho]. \tag{5.75}$$

Since $\bar{\mathbf{v}}$ is $O(V_+)$ and the bias of $\hat{\rho}$ is $O(m_+^{-1})$, the bias of $\hat{\bar{\mathbf{V}}}$ is $O(V_+/m_+) = O(1)$ in the above limit process.

As given by (5.58), the mean squared error of $\hat{\bar{\mathbf{V}}}$ is

$$E[(\hat{\bar{\mathbf{V}}} - \bar{\mathbf{V}})(\hat{\bar{\mathbf{V}}} - \bar{\mathbf{V}})'] \;=\; \mathbf{D}_{\bar{\mathbf{V}}} E\left[(\hat{\rho}-\rho)(\hat{\rho}-\rho)'\right] \mathbf{D}_{\bar{\mathbf{V}}} + E\left[\mathbf{D}_{\mathbf{y}}\hat{\rho}\hat{\rho}'\mathbf{D}_{\mathbf{y}}\right]. \tag{5.76}$$

The first term in (5.76) is $\mathbf{D}_{\bar{\mathbf{V}}}\mathbf{\Sigma}\mathbf{D}_{\bar{\mathbf{V}}}$, where $\mathbf{\Sigma}$ is as given by (5.74). This is $O(V_+^2)O(m_+^{-1}) = O(V_+) = O(m_+)$ in the augmented limit process.

Substituting $\rho - \bar{\mathbf{n}}^{\dagger}\mathbf{X}\rho - \mathbf{b} + \mathbf{r}$ for $\hat{\rho}$, the second term can be written as

$$E\left[\mathbf{D}_{\mathbf{y}}\hat{\rho}\hat{\rho}'\mathbf{D}_{\mathbf{y}}\right] \;=\; E\left[\mathbf{D}_{\mathbf{y}}\rho\rho'\mathbf{D}_{\mathbf{y}}\right] + O(V_+^{-1}).$$

This leading term is $O(V_+) = O(m_+)$ in the limit process. So, the mean squared error of $\hat{\bar{\mathbf{V}}}$ is $O(m_+)$. Since the bias of $\hat{\bar{\mathbf{V}}}$ is $O(1)$, it is negligible compared to the root mean squared error. So, the approximate variance of $\hat{\bar{\mathbf{V}}}$ is

$$Var(\hat{\bar{\mathbf{V}}}) \approx \mathbf{D}_{\bar{\mathbf{V}}}\mathbf{\Sigma}\mathbf{D}_{\bar{\mathbf{V}}} + E(\mathbf{D}_{\mathbf{y}}\rho\rho'\mathbf{D}_{\mathbf{y}}). \tag{5.77}$$

Estimation of variances

First, to calculate (5.74), note that the (i,l)th element of $E[\mathbf{X}\rho\rho'\mathbf{X}']$ is $\sum_{jk}\rho_j\rho_k E[X_{ij}X_{lk}]$. According to the probability models (5.4) and (5.5),

$$E(X_{ij}X_{lk}) = Cov(n_{ij}, n_{lk}) = \delta_{il}m_i[\delta_{jk}\theta_{ij}p_j - \theta_{ij}p_j\theta_{ik}p_k]. \tag{5.78}$$

Therefore,

$$\begin{aligned}
\sum_{jk}\rho_j\rho_k E[X_{ij}X_{lk}] &= \delta_{il}m_i\sum_{jk}(\delta_{jk}\theta_{ij}\rho_k - \theta_{ij}\theta_{ik}) \\
&= \delta_{il}m_i\left[\sum_j\theta_{ij}\rho_j - (\sum_j\theta_{ij})(\sum_k\theta_{ik})\right] \\
&= \delta_{il}m_i\left[\sum_j\theta_{ij}\rho_j - 1\right] \\
&= \delta_{il}m_i\mu_i, \qquad (5.79)
\end{aligned}$$

where $\mu_i = \sum_j\theta_{ij}\rho_j - 1 = \sum_j\psi_{ij}\rho_j^2 - 1$. Noting that this is the (i,l)th element of $\mathbf{D_m}\mathbf{D}_{\boldsymbol{\mu}}$, the approximate variance (5.74) of $\widehat{\rho}$ can be written as

$$\Sigma \approx \overline{\mathbf{n}}^{\dagger}\mathbf{D_m}\mathbf{D}_{\boldsymbol{\mu}}(\overline{\mathbf{n}}^{\dagger})'. \tag{5.80}$$

Next, in order to calculate the second term in (5.77) note that

$$E(y_jy_k) = Cov(v_j, v_k) = \delta_{jk}\overline{V}_jp_j(1-p_j). \tag{5.81}$$

Hence,

$$E(\mathbf{D_y}\rho\rho'\mathbf{D_y}) = \mathbf{D}_{\overline{\mathbf{V}}}(\mathbf{D}_\rho - \mathbf{I}). \tag{5.82}$$

Then, (5.80) and (5.82) lead (5.77) to

$$\begin{aligned}
Var(\widehat{\overline{\mathbf{V}}}) &\approx \mathbf{D}_{\overline{\mathbf{v}}}\Sigma\mathbf{D}_{\overline{\mathbf{v}}} + \mathbf{D}_{\overline{\mathbf{V}}}(\mathbf{D}_\rho - \mathbf{I}), \quad \text{and} \qquad (5.83) \\
Var(\widehat{V}_+) &\approx \overline{\mathbf{v}}'\Sigma\overline{\mathbf{v}} + \overline{\mathbf{V}}'(\rho - \mathbf{1}). \qquad (5.84)
\end{aligned}$$

The approximate variances may be estimated by replacing the unknown parameters by their estimates. An estimator for $\boldsymbol{\mu}$ is,

$$\widehat{\boldsymbol{\mu}} = \mathbf{D}_{\mathbf{m}}^{-1}\mathbf{n}\mathbf{D}_{\widehat{\rho}}\widehat{\rho} - \mathbf{1}. \tag{5.85}$$

Remark 6

When $s = t$, Darroch's formula (5.51) can be deduced from (5.80) by noting that $\overline{\mathbf{n}}^{\dagger} = \overline{\mathbf{n}}^{-1}$. Consequently, (5.52) and (5.53) are also equivalent to (5.83) and (5.84) respectively.

5.4.2 Variance of $\widehat{\mathbf{U}}$

Now, we derive a formula for the approximate variance of $\hat{\mathbf{U}}$ given by (5.35). Since the derivation is similar to that in Subsection 5.4.1, here we present only the main steps.

When $s \leq t$, let

$$\begin{aligned} \mathbf{B} &= \mathbf{D_m}(\overline{\mathbf{n}}\,\overline{\mathbf{n}}')^{-1}\overline{\mathbf{n}} = \mathbf{D_m}(\overline{\mathbf{n}}')^{\dagger}, \quad \text{and} && (5.86) \\ \hat{\mathbf{B}} &= \mathbf{D_m}(\mathbf{n}\,\mathbf{n}')^{-1}\mathbf{n} = \mathbf{D_m}(\mathbf{n}')^{\dagger}. && (5.87) \end{aligned}$$

Then, from (5.50) and (5.35),

$$\begin{aligned} \mathbf{U} &= \mathbf{B}\overline{\mathbf{v}}, \quad \text{and} && (5.88) \\ \hat{\mathbf{U}} &= \hat{\mathbf{B}}\mathbf{v} && (5.89) \end{aligned}$$

respectively. Let $\mathbf{y}$ be defined as in (5.55). Then,

$$\begin{aligned} \hat{\mathbf{U}} - \mathbf{U} &= (\hat{\mathbf{B}} - \mathbf{B})\overline{\mathbf{v}} + \hat{\mathbf{B}}\mathbf{y}, \text{ and} \\ Var(\hat{\mathbf{U}}) &\approx E\left[(\hat{\mathbf{B}} - \mathbf{B})\overline{\mathbf{v}}\,\overline{\mathbf{v}}'(\hat{\mathbf{B}} - \mathbf{B})'\right] + E\left[\mathbf{B}\mathbf{y}\mathbf{y}'\mathbf{B}'\right] \\ &= E\left[(\hat{\mathbf{B}} - \mathbf{B})\overline{\mathbf{v}}\,\overline{\mathbf{v}}'(\hat{\mathbf{B}} - \mathbf{B})'\right] + \mathbf{B}E\left[\mathbf{y}\mathbf{y}'\right]\mathbf{B}'. && (5.90) \end{aligned}$$

Letting $Var(\hat{\mathbf{U}}) = \mathbf{\Omega}$, $E\left[(\hat{\mathbf{B}} - \mathbf{B})\overline{\mathbf{v}}\,\overline{\mathbf{v}}'(\hat{\mathbf{B}} - \mathbf{B})'\right] = \mathbf{\Lambda}$, and $E\left[\mathbf{y}\mathbf{y}'\right] = \mathbf{\Upsilon}$, respectively, we have

$$\mathbf{\Omega} = \mathbf{\Lambda} + \mathbf{B}\mathbf{\Upsilon}\mathbf{B}'. \qquad (5.91)$$

Calculation of Λ

Let $\mathbf{X}$ be defined as in (5.60). Then,

$$\begin{aligned}\mathbf{n}\mathbf{n}' &= (\overline{\mathbf{n}} + \mathbf{X})(\overline{\mathbf{n}}' + \mathbf{X}') \\ &= \overline{\mathbf{n}}\,\overline{\mathbf{n}}' + \overline{\mathbf{n}}\mathbf{X}' + \mathbf{X}\overline{\mathbf{n}}' + \mathbf{X}\mathbf{X}' \\ &= \mathbf{\Gamma} + \mathbf{Z},\end{aligned}$$

where $\mathbf{\Gamma} = \overline{\mathbf{n}}\,\overline{\mathbf{n}}'$ and $\mathbf{Z} = \overline{\mathbf{n}}\mathbf{X}' + \mathbf{X}\overline{\mathbf{n}}' + \mathbf{X}\mathbf{X}'$ (Note the new definition for $\mathbf{\Gamma}$. It was previously defined as $\mathbf{\Gamma} = \overline{\mathbf{n}}'\,\overline{\mathbf{n}}$). Then, using the identity (5.62),

$$\begin{aligned}(\mathbf{n}')^{\dagger} &= (\mathbf{n}\mathbf{n}')^{-1}\mathbf{n} \\ &= [\mathbf{I} - \mathbf{\Gamma}^{-1}\mathbf{Z}]\mathbf{\Gamma}^{-1}(\overline{\mathbf{n}} + \mathbf{X}) + (\mathbf{\Gamma}^{-1}\mathbf{Z})^2(\mathbf{n}\mathbf{n}')^{-1}\mathbf{n}. \end{aligned} \tag{5.92}$$

Pre-multiplying (5.92) by $\mathbf{D_m}$ and substituting $\overline{\mathbf{n}}\mathbf{X}' + \mathbf{X}\overline{\mathbf{n}}' + \mathbf{X}\mathbf{X}'$ for $\mathbf{Z}$ in the square bracket,

$$\begin{aligned}\widehat{\mathbf{B}} &= \mathbf{D_m}\mathbf{\Gamma}^{-1}\overline{\mathbf{n}} + \mathbf{D_m}\mathbf{\Gamma}^{-1}\mathbf{X}(\mathbf{I} - \overline{\mathbf{n}}^{\dagger}\,\overline{\mathbf{n}}) - \mathbf{D_m}\mathbf{\Gamma}^{-1}\overline{\mathbf{n}}\mathbf{X}'\mathbf{\Gamma}^{-1}\overline{\mathbf{n}} + \mathbf{R} \\ &= \mathbf{B} + \mathbf{A}\mathbf{X}\mathbf{Q} - \mathbf{B}\mathbf{X}'(\overline{\mathbf{n}}')^{\dagger} + \mathbf{R}, \end{aligned} \tag{5.93}$$

where

$$\begin{aligned}\mathbf{Q} &= \mathbf{I} - \overline{\mathbf{n}}^{\dagger}\overline{\mathbf{n}}, \\ \mathbf{A} &= \mathbf{D_m}\mathbf{\Gamma}^{-1},\end{aligned}$$

and $\mathbf{R}$ is the remainder term with expected value of order $O(m_{+}^{-1})$. Post-multiplication of (5.93) by $\overline{\mathbf{v}}$ leads to

$$\begin{aligned}(\widehat{\mathbf{B}} - \mathbf{B})\overline{\mathbf{v}} &\approx \mathbf{A}\mathbf{X}\mathbf{Q}\overline{\mathbf{v}} - \mathbf{B}\mathbf{X}'(\overline{\mathbf{n}}')^{\dagger}\overline{\mathbf{v}}, \text{ and} \\ \Lambda = E\left[(\widehat{\mathbf{B}} - \mathbf{B})\overline{\mathbf{v}}\,\overline{\mathbf{v}}'(\widehat{\mathbf{B}} - \mathbf{B})'\right] &\approx \mathbf{A}E\left[\mathbf{X}\mathbf{Q}\overline{\mathbf{v}}\,\overline{\mathbf{v}}'\mathbf{Q}'\mathbf{X}'\right]\mathbf{A}' + \mathbf{B}E\left[\mathbf{X}'(\overline{\mathbf{n}}')^{\dagger}\overline{\mathbf{v}}\,\overline{\mathbf{v}}'(\overline{\mathbf{n}})^{\dagger}\mathbf{X}\right]\mathbf{B}' \\ &\quad -\mathbf{A}E\left[\mathbf{X}\mathbf{Q}\overline{\mathbf{v}}\,\overline{\mathbf{v}}'\overline{\mathbf{n}}^{\dagger}\mathbf{X}\right]\mathbf{B}' - \mathbf{B}E\left[\mathbf{X}'(\overline{\mathbf{n}}')^{\dagger}\overline{\mathbf{v}}\,\overline{\mathbf{v}}'\mathbf{Q}'\mathbf{X}'\right]\mathbf{A}' \\ &= \mathbf{A}E\left[\mathbf{X}\boldsymbol{\alpha}\boldsymbol{\alpha}'\mathbf{X}'\right]\mathbf{A}' + \mathbf{B}E\left[\mathbf{X}'\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}\right]\mathbf{B}' \\ &\quad -\mathbf{A}E\left[\mathbf{X}\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X}\right]\mathbf{B}' - \mathbf{B}E\left[\mathbf{X}'\boldsymbol{\beta}\boldsymbol{\alpha}'\mathbf{X}'\right]\mathbf{A}', \end{aligned} \tag{5.94}$$

where

$$\begin{aligned}\boldsymbol{\alpha} &= \mathbf{Q}\overline{\mathbf{v}}, \text{ and}\\ \boldsymbol{\beta} &= (\overline{\mathbf{n}}')^{\dagger}\overline{\mathbf{v}}.\end{aligned}$$

Now letting

$$\begin{aligned}\mathbf{D} &= E\left[\mathbf{X}\boldsymbol{\alpha}\boldsymbol{\alpha}'\mathbf{X}'\right],\\ \mathbf{F} &= E\left[\mathbf{X}'\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}\right], \text{ and}\\ \mathbf{G} &= E\left[\mathbf{X}\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X}\right],\end{aligned}$$

and noting that $\mathbf{Q} = \mathbf{0}$ when $s = t$, $\boldsymbol{\Lambda}$ can be written as

$$\boldsymbol{\Lambda} = \begin{cases}\mathbf{ADA}' + \mathbf{BFB}' - \mathbf{AGB}' - \mathbf{BG}'\mathbf{A}', & \text{if } s < t, \text{ and}\\ \mathbf{BFB}', & \text{if } s = t.\end{cases} \tag{5.95}$$

Calculation of $Var(\hat{U}_+)$

From (5.35),

$$\hat{U}_+ = \mathbf{v}'\mathbf{n}^{\dagger}\mathbf{m}.$$

But, since this is a scalar, it can also be written as

$$\hat{U}_+ = \mathbf{m}'(\mathbf{n}')^{\dagger}\mathbf{v}. \tag{5.96}$$

Similarly, from (5.50),

$$U_+ = \mathbf{m}'(\overline{\mathbf{n}}')^{\dagger}\overline{\mathbf{v}}. \tag{5.97}$$

Letting

$$\mathbf{b}' = \mathbf{m}'(\overline{\mathbf{n}}')^{\dagger}, \text{ and} \tag{5.98}$$

$$\hat{\mathbf{b}}' = \mathbf{m}'(\mathbf{n}')^{\dagger}, \tag{5.99}$$

(5.96) and (5.97) can be written as

$$\hat{U}_+ = \hat{\mathbf{b}}'\mathbf{v}, \text{ and} \tag{5.100}$$

$$U_+ = \mathbf{b}'\overline{\mathbf{v}}. \tag{5.101}$$

Comparing (5.89) and (5.88) with (5.100) and (5.101) respectively, we can see that the calculations in Subsection 5.4.2 can be repeated with **U** and **B** replaced by U_+ and **b**′ respectively, to calculate the variance of $\widehat{U}_+$. This also requires replacing **A** by

$$\mathbf{a}' \;=\; \mathbf{m}'\mathbf{T}^{-1}. \tag{5.102}$$

Consequently, a formulae for the variance of $\widehat{U}_+$ can be obtained by replacing **A** and **B** in the variance formula for $\widehat{\mathbf{U}}$ by **a**′ and **b**′ respectively. Letting

$$\lambda \;=\; \begin{cases} \mathbf{a}'\mathbf{D}\mathbf{a} + \mathbf{b}'\mathbf{F}\mathbf{b} - \mathbf{a}'\mathbf{G}\mathbf{b} - \mathbf{b}'\mathbf{G}'\mathbf{a}, & \text{if } s < t, \text{ and} \\ \mathbf{b}'\mathbf{F}\mathbf{b}, & \text{if } s = t, \end{cases} \tag{5.103}$$

the approximate variance of $\widehat{U}_+$ can be written as

$$Var(\widehat{U}_+) \approx \lambda + \mathbf{b}'\boldsymbol{\Upsilon}\mathbf{b}. \tag{5.104}$$

Estimation of variances

Since $E(X_{ij}X_{lk}) = \delta_{il}m_i[\delta_{jk}\psi_{ij} - \psi_{ij}\psi_{ik}]$, the matrices **D**, **F** and **G** can be evaluated as follows:

$$\begin{aligned} D_{il} &= \sum_{jk} \alpha_j\alpha_k E(X_{ij}X_{lk}) \\ &= \sum_{jk} \alpha_j\alpha_k\delta_{il}m_i\left[\delta_{jk}\psi_{ij} - \psi_{ij}\psi_{ik}\right], \end{aligned} \tag{5.105}$$

$$\begin{aligned} F_{jk} &= \sum_{il} \beta_i\beta_l E(X_{ij}X_{lk}) \\ &= \sum_{il} \beta_i\beta_l\delta_{il}m_i[\delta_{jk}\psi_{ij} - \psi_{ij}\psi_{ik}] \\ &= \sum_{i} \beta_i^2 m_i[\delta_{jk}\psi_{ij} - \psi_{ij}\psi_{ik}], \end{aligned} \tag{5.106}$$

and

$$\begin{aligned} G_{ik} &= \sum_{jl} \alpha_j\beta_l E(X_{ij}X_{lk}) \\ &= \sum_{jl} \alpha_j\beta_l\delta_{il}m_i\left[\delta_{jk}\psi_{ij} - \psi_{ij}\psi_{ik}\right]. \end{aligned} \tag{5.107}$$

Now we evaluate

$$\Upsilon = E\left[\mathbf{y}\mathbf{y}'\right]. \tag{5.108}$$

Since $\mathbf{p}$ cannot be estimated when $s < t$, the expression (5.81) cannot be used to estimate Υ. Instead, we should express it in terms of the estimable quantities, ψ_{ij}. According to (5.81), Υ is a diagonal matrix. Under probability models (5.5) and (5.42), the jth diagonal element of Υ is

$$\begin{aligned}
\Upsilon_{jj} &= E(y_j^2) \\
&= Var(v_j) \\
&= Var\left[\sum_{i=1}^{s} v_{ij}\right] \\
&= \sum_{i=1}^{s} Var(v_{ij}) \\
&= \sum_{i=1}^{s} U_i \psi_{ij}(1-\psi_{ij}).
\end{aligned} \tag{5.109}$$

Now, the approximate variances may be estimated by replacing all the parameters by their estimates.

5.5 Relaxing Assumption 4

So far, all the calculations were done under the assumption that $\overline{\mathbf{n}}$ and $\mathbf{n}$ were of full rank (Assumption 4). As long as Assumptions 1-3 and 5 or 6 are satisfied, this is sufficient for the estimators to be consistent. This assumption was also useful in deriving formulae for the approximate variances of the estimators.

According to Seber [33](pg. 327), and Albert [1](see Appendix), regardless of whether or not $\mathbf{X}$ is of full rank, $\hat{\beta} = \mathbf{X}^{\dagger}\mathbf{y}$ minimizes $\|\mathbf{y} - \mathbf{X}\beta\|$. So, regardless of whether or not $\mathbf{n}$ is of full rank, (5.35) provides least-squares estimates. But now, $\mathbf{n}^{\dagger}$ is not necessarily defined by (5.34). In general, $\mathbf{n}^{\dagger}$ can be calculated via a singular-value decomposition. Even though least-squares estimates are available in this way under any circumstances, they may not necessarily be consistent.

Recall from (5.18) that

$$\bar{\mathbf{n}} = \mathbf{D_m}\Theta\mathbf{D_p}.$$

The matrix $\bar{\mathbf{n}}$ has full rank if and only if Θ is of full rank. If $\bar{\mathbf{n}}$ has full rank and the samples are large, a non-full rank matrix $\mathbf{n}$ is extremely unlikely. So, as long as Θ has full rank, and samples are large, we can estimate the numbers of animals in each of the strata (tagging or recovery, depending on relative sizes of s and t, and on validity of Assumptions 5 or 6) consistently.

As Darroch noted, there are a few obvious instances where the Θ matrix is not of full rank. For example,

1. $\theta_{ij} = 0 \ \ \forall i$,
2. θ_{ij}/θ_{lj} is constant independent of j, and
3. θ_{ij}/θ_{ik} is constant independent of i.

In the first instance, the jth stratum is effectively non-existent, and can be ignored. In the other two instances, the relevant strata may be pooled to avoid the non-full rank $\bar{\mathbf{n}}$ matrices. In Chapter 6, we show that such poolings do not affect the consistency of the pooled estimators. So, even if Θ is of less than full rank, we are still able to consistently estimate numbers of animals in unpooled strata and totals in pooled strata. It is possible to test for the above two collinearities in the Θ matrix as follows.

Test for collinearity of two rows of Θ matrix

Tagging stratum	# marked released	# marked recovered in stratum j				# marked not recovered
		1	2	...	t	
i	m_i	n_{i1}	n_{i2}	...	n_{it}	$m_i - n_{i+}$
l	m_l	n_{l1}	n_{l2}	...	n_{lt}	$m_l - n_{l+}$
Total	$m_i + m_l$	$n_{i1} + n_{l1}$	$n_{i2} + n_{l2}$	...	$n_{it} + n_{lt}$	$m_i + m_l - (n_{i+} + n_{l+})$

We are interested in testing the hypothesis $H_0 : \theta_{ij}/\theta_{lj}$ =constant for all j. Equivalently, we can test

$$H_0 : \theta_{ij}p_j/\theta_{lj}p_j = \psi_{ij}/\psi_{lj} = \text{constant for all } j.$$

Under H_0, for $a = i$ or l,

$$\hat{\psi}_{aj} = \frac{n_{a+}(n_{ij} + n_{lj})}{(n_{i+} + n_{l+})m_a}.$$

Let e_{ij} be the expected number of marked animals which are released in stratum i and recovered in stratum j. Then,

$$e_{aj} = m_a\hat{\psi}_{aj} = \frac{n_{i+}(n_{ij} + n_{lj})}{n_{i+} + n_{l+}}$$

Under H_0, the test statistic

$$T_r = \sum_{a=i,l} \sum_j \frac{(n_{aj} - e_{aj})^2}{e_{aj}},$$

has an approximate χ^2 distribution with $t - 1$ degrees of freedom.

As Darroch pointed out, the goodness of fit of the special hypothesis $H_0 : \theta_{ij} = \theta_{lj} \;\; \forall j$ is equivalent to a test of homogeneity for the rows of $2 \times (t + 1)$ contingency table including the last column in the above table. Now, under the null hypothesis, T_r has an approximate χ^2 distribution with t degrees of freedom.

Test for collinearity of two columns of Θ matrix

Similarly, the hypothesis $H_0 : \theta_{ij}/\theta_{ik}$=constant for all i, can be tested using the test statistic

$$T_c = \sum_i \sum_{a=j,k} \frac{(n_{ia} - e_{ia})^2}{e_{ia}}.$$

Here,

$$e_{ia} = \frac{n_{+a}(n_{ij} + n_{ik})}{n_{+j} + n_{+k}}.$$

Under H_0, T_c has an approximate χ^2 distribution with $s - 1$ degrees of freedom.

5.6 Discussion

In this chapter we addressed the problem of estimating the population size using stratified two-sample tag recovery data. We derived least-squares estimators and provided algorithms for obtaining weighted least-squares estimates. These estimators have the advantage that they can be easily applied even when the numbers of tagging strata and recovery strata are unequal.

However, these estimators share some weaknesses that are inherent to the existing methodology. For example, the unrestricted nature of estimation procedure can lead to probability estimators that are out of range, and population size estimators that are negative. These problems are particularly likely to arise when the data are sparse or the model is inadequate. Although the deficiencies should be dealt with directly, it is possible to impose direct restrictions on estimates of population sizes and probabilities. For instance, the procedure for estimating $\overline{\mathbf{V}}$ neglected that $0 < p_j \leq 1 \ \forall j$. This restriction can be incorporated into the estimation procedure by minimizing the error sums of squares subject to constraints or by converting the problem into an unconstrained one. The latter may be accomplished by re-parameterizing the problem using $\alpha_j = \log(p_j/1 - p_j)$ and minimizing the sum of squares of the errors, $\epsilon_i = \sum_{j=1}^{t} n_{ij}(1 + e^{-\alpha_j}) - m_i$, with respect to the α_j's. Similarly, in estimating

U, we neglected that $U_i \geq 0 \ \forall \ i$. This restriction may be imposed by writing $U_i = e^{\alpha_i}$ and minimizing the sum of squares of errors with respect to these α_i's. This approach does not necessarily provide reasonable estimates. However, it avoids unacceptable estimates.

Example 1 in Section 6.6 shows that least-squares estimation leads to reasonable estimates which are close to those obtained by other methods with more assumptions.

Chapter 6

Pooling in a Stratified Two-Sample Tag Recovery Census

Investigators often pool strata for various reasons. The pooling can occur before or after the experiment. For instance, if the investigators do not know how the population is stratified, then they may conduct an unstratified experiment and use the Petersen estimator. Also, if they are unable to use different tags for some of the tagging strata, then those strata may be pooled together. Investigators often pool data after the experiment when the numbers of tagged animals recovered in some strata are small. One such pooling can be found in Darroch's [13] paper, where he used Schaefer's sockeye salmon data set to estimate a run of migrating salmon. This data set has eight tagging strata and nine recovery strata. Darroch pooled the first three and last three weeks of tagging into single strata and the first three and last four weeks of recovery into single strata because frequencies in these weeks were too small to be used in large sample theory. As pointed out in Section 5.5, experimenters may also pool strata to avoid non-full rank matrices.

In any case, when the strata are pooled, the estimators can be inconsistent. For example, the Petersen estimator,

$$\widehat{V}_{+} = \frac{m_{+}v_{+}}{n_{++}}, \tag{6.1}$$

is not in general consistent. Darroch showed that (6.1) is consistent under each of the following practically appealing conditions:

i. A constant proportion of each recovery stratum is sampled ($p_j = p$ for $j = 1, \ldots, t$).

ii. A constant proportion of each stratum is tagged ($m_i/V_i = m_+/V_+$ for $i = 1, \ldots, s$) and the movement pattern for tagged animals and untagged animals is the same.

iii. Complete mixing of the whole population ($\theta_{ij} = \theta_j$ for all i and j) and the movement pattern for tagged animals and untagged animals is the same.

In the next three sections, we extend Darroch's work by finding sufficient conditions for pooling strata partially, so that the estimates may be consistent. We first consider pooling tagging and recovery strata separately, and finally combine the results to facilitate both types of pooling simultaneously. Then in Section 6.5, we examine the biases that can occur in the estimators if these conditions are not satisfied. Throughout the rest of this chapter, the pooled quantities are denoted by attaching a superscript '*'. The numbers of tagging and recovery strata after pooling are denoted by s^* and t^* respectively ('*' used in the subsequent sections should not be confused with the '$\star$' used previously to describe Darroch's estimates.)

6.1 Partial Pooling of Tagging Strata

Suppose that two of the tagging strata are pooled. To simplify the notation, without loss of generality, renumber the strata such that the pooled strata are the last two. The pooled quantities are,

$$\mathbf{m}^* = \begin{pmatrix} m_1 \\ \vdots \\ m_{s-2} \\ m_{s-1} + m_s \end{pmatrix}, \ \mathbf{U}^* = \begin{pmatrix} U_1 \\ \vdots \\ U_{s-2} \\ U_{s-1} + U_s \end{pmatrix}, \ \mathbf{n}^* = \begin{pmatrix} n_{11} & \cdots & n_{1t} \\ \vdots & \vdots & \vdots \\ n_{(s-2)1} & \cdots & n_{(s-2)t} \\ n_{(s-1)1} + n_{s1} & \cdots & n_{(s-1)t} + n_{st} \end{pmatrix}.$$

Naturally, one would use

$$\widehat{\overline{\mathbf{V}}} = \mathbf{D}_{\mathbf{v}}(\mathbf{n}^*)^{\dagger}\mathbf{m}^* \ \text{ when } s^* \geq t. \tag{6.2}$$

and

$$\hat{\mathbf{U}}^* = \mathbf{D}_{\mathbf{m}^*}([\mathbf{n}^*]')^{\dagger}\mathbf{v} \ \text{ when } s^* \leq t, \tag{6.3}$$

to estimate $\overline{\mathbf{V}}$ and $\mathbf{U}^*$ respectively. Here, $\mathbf{n}^{*\dagger}$ is the Moore-Penrose inverse of $\mathbf{n}^*$ defined similarly to (5.34).

Arguing as in Section 5.3, it can be shown that (6.2) and (6.3) are consistent for $\mathbf{D}_{\overline{\mathbf{v}}}(\overline{\mathbf{n}^*})^{\dagger}\mathbf{m}^*$ and $\mathbf{D}_{\mathbf{m}^*}([\overline{\mathbf{n}^*}]')^{\dagger}\overline{\mathbf{v}}$ respectively. So to derive sufficient conditions for the consistency of (6.2) and (6.3), we find conditions under which $\mathbf{D}_{\overline{\mathbf{v}}}(\overline{\mathbf{n}^*})^{\dagger}\mathbf{m}^* = \overline{\mathbf{V}}$ and $\mathbf{D}_{\mathbf{m}^*}(\overline{\mathbf{n}^*})^{\dagger}\overline{\mathbf{v}} = \mathbf{U}^*$ respectively. For this, it is enough to find conditions under which the pooled quantities satisfy the key equations in Sections 5.2.2 and 5.2.3. Then, by the same line of calculations as in these Sections, and in Remark 5, the consistency of the pooled estimators will follow.

Result 6.1 *Suppose that two of the tagging strata are pooled. Also suppose that Assumptions 1-3 are satisfied and matrices* $\overline{\mathbf{n}}^*$ *and* $\mathbf{n}^*$ *are of full rank.*

R.6.1.1

If Assumption 5 is satisfied, and $s^* \geq t$*, then the estimator given by (6.2) is consistent for* $\overline{\mathbf{V}}$.

R.6.1.2

If Assumption 6 is satisfied, and $s^* \leq t$*, then the following Conditions 1 and 2 are individually sufficient for the estimator given by (6.3) to be consistent for* $\mathbf{U}^*$.

Condition 1

The ratios of the numbers of tagged to untagged animals are the same in the two pooled strata. I.e,

$$U_i = cm_i \text{ for the two pooled tagging strata.}$$

(This is the analogue of Darroch's sufficient condition (ii) for the Petersen estimator to be consistent.)

Condition 2

The patterns of moving from pooled tagging strata to any given recovery stratum are the same. I.e.,

$$\theta_{(s-1)j} = \theta_{sj} \text{ for all recovery strata } j.$$

(This is analogous to Darroch's condition (iii)).

Proof

To prove R.6.1.1, it is enough to show that the pooled data satisfy the three key equations in Subsection 5.2.2. Since we pool only the tagging strata, the first key equation (5.17) in Subsection 5.2.2 is unchanged. That is,

$$\mathbf{D}_{\overline{\mathbf{v}}} = \mathbf{D}_{\overline{\mathbf{V}}}\mathbf{D}_{\mathbf{p}}. \tag{6.4}$$

Next, we show that there exists a matrix $\tilde{\boldsymbol{\Theta}}$, such that

$$\overline{\mathbf{n}}^* = \mathbf{D}_{\mathbf{m}\bullet}\tilde{\boldsymbol{\Theta}}\mathbf{D}_{\mathbf{p}}, \tag{6.5}$$

and

$$\tilde{\boldsymbol{\Theta}}\mathbf{1} = \mathbf{1}. \tag{6.6}$$

For this, let

$$\tilde{\boldsymbol{\Theta}} = \begin{pmatrix} \theta_{11} & \dots & \theta_{1t} \\ \vdots & \vdots & \vdots \\ \theta_{(s-2)1} & \cdots & \theta_{(s-2)t} \\ \tilde{\theta}_{(s-1)1} & \cdots & \tilde{\theta}_{(s-1)t} \end{pmatrix},$$

where $\tilde{\theta}_{(s-1)j}$'s are probabilities such that $\sum_{j=1}^{t} \tilde{\theta}_{(s-1)j} \leq 1$. Then, for $i = 1, \dots, s-2$ and $j = 1, \dots, t$,

$$\begin{aligned} E(n^*_{ij}) &= E(n_{ij}) = m_i\theta_{ij}p_j \\ &= m^*_i\tilde{\theta}_{ij}p_j, \end{aligned}$$

and for $j = 1, \dots, t$,

$$\begin{aligned} E\left[n^*_{(s-1)j}\right] &= E\left[n_{(s-1)j} + n_{sj}\right] \\ &= m_{s-1}\theta_{(s-1)j}p_j + m_s\theta_{sj}p_j \\ &= (m_{s-1} + m_s)\left[\frac{m_{s-1}\theta_{(s-1)j} + m_s\theta_{sj}}{m_{s-1} + m_s}\right]p_j. \end{aligned} \tag{6.7}$$

Therefore, if

$$\tilde{\theta}_{(s-1)j} = \left[\frac{m_{s-1}\theta_{(s-1)j} + m_s\theta_{sj}}{m_{s-1} + m_s}\right] \text{ for } j = 1, \ldots, t, \tag{6.8}$$

then, $\tilde{\mathbf{\Theta}}$ satisfies (6.5). Further note that if $\mathbf{\Theta}$ satisfies (5.15), then so does $\tilde{\mathbf{\Theta}}$; i.e., it satisfies (6.6). Therefore, from (6.4), (6.5), and (6.6), by the same line of calculations as in Subsection 5.2.2 and Remark 5, it follows that $\overline{\mathbf{V}} = \mathbf{D}_{\overline{\mathbf{v}}}(\overline{\mathbf{n}^*})^{\dagger}\mathbf{m}^*$. This proves R.6.1.1.

To prove R.6.1.2, we find conditions under which the pooled data satisfy the two key equations (5.27), and (5.28) in Subsection 5.2.3. Note that for $j = 1, \ldots, t$,

$$\begin{aligned} E(v_j) &= \sum_{i=1}^{s} U_i\theta_{ij}p_j \\ &= \sum_{i=1}^{s-2} U_i^*\tilde{\theta}_{ij}p_j + (U_{s-1} + U_s)\left[\frac{U_{s-1}\theta_{(s-1)j} + U_s\theta_{sj}}{U_{s-1} + U_s}\right]p_j. \end{aligned} \tag{6.9}$$

This can be written as

$$\overline{\mathbf{v}}' = \mathbf{U}^{*\prime}\tilde{\mathbf{\Theta}}\mathbf{D}_{\mathbf{p}}, \tag{6.10}$$

if

$$\left[\frac{U_{s-1}\theta_{(s-1)j} + U_s\theta_{sj}}{U_{s-1} + U_s}\right] = \tilde{\theta}_{(s-1)j} \text{ for } j = 1, \ldots, t. \tag{6.11}$$

We have already shown that the second key equation (5.28) in Subsection 5.2.3, which is identical to (5.18) is satisfied by the pooled data if (6.8) is satisfied. So, both key equations in Subsection 5.2.3 are satisfied by the pooled data if

$$\left[\frac{U_{s-1}\theta_{(s-1)j} + U_s\theta_{sj}}{U_{s-1} + U_s}\right] = \left[\frac{m_{s-1}\theta_{(s-1)j} + m_s\theta_{sj}}{m_{s-1} + m_s}\right] \text{ for } j = 1, \ldots, t. \tag{6.12}$$

Conditions 1 and 2 individually satisfy (6.12). This completes the proof.

Remark

As pointed out in Section 5.5, investigators may have to pool some tagging strata if the Θ is non-full rank due to collinearity of the rows. As we showed in Result 6.1, the pooled estimates are consistent if the corresponding rows in the Θ matrix are equal. However, if the rows are not equal, but collinear, then such pooling does not necessarily lead to consistent estimates of $\mathbf{U}$ (although the estimate of $\overline{\mathbf{V}}$ is still consistent). Recall that $\hat{\mathbf{U}}^*$ is consistent if (6.12) is satisfied. From this we can see that the estimates $\hat{\mathbf{U}}^*$ obtained by pooling tagging strata corresponding to two collinear rows of the Θ matrix are consistent if Condition 1 is also satisfied. That is, if

$$\begin{aligned} \theta_{ij}/\theta_{lj} &= \text{a constant not equal to 1, and} && (6.13) \\ U_i &= cm_i, && (6.14) \end{aligned}$$

for some tagging strata, then the investigator can pool those tagging strata to get consistent estimates of $\mathbf{U}$.

6.2 Partial Pooling of Recovery Strata

Now, we consider the partial pooling of recovery strata. Suppose that two of the recovery strata are pooled. Without loss of generality, renumber the recovery strata such that pooled strata are the last two strata. Then, the pooled quantities are

$$\mathbf{v}^* = \begin{pmatrix} v_1 \\ \vdots \\ v_{t-2} \\ v_{t-1} + v_t \end{pmatrix}, \quad \overline{\mathbf{V}}^* = \begin{pmatrix} \overline{V}_1 \\ \vdots \\ \overline{V}_{t-2} \\ \overline{V}_{t-1} + \overline{V}_t \end{pmatrix}, \quad \mathbf{n}^* = \begin{pmatrix} n_{11} & \dots & n_{1(t-2)} & n_{1(t-1)} + n_{1t} \\ \vdots & \vdots & \vdots & \vdots \\ n_{s1} & \dots & n_{s(t-2)} & n_{s(t-1)} + n_{st} \end{pmatrix},$$

and

$$\Theta^* = \begin{pmatrix} \theta_{11} & \dots & \theta_{1(t-2)} & \theta_{1(t-1)} + \theta_{1t} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{s1} & \dots & \theta_{s(t-2)} & \theta_{s(t-1)} + \theta_{st} \end{pmatrix}. \tag{6.15}$$

Naturally, one would use the estimators,

$$\widehat{\overline{\mathbf{V}}}^{*} = \mathbf{D}_{\mathbf{v}^{*}}(\mathbf{n}^{*})^{\dagger}\mathbf{m} \text{ when } s \geq t^{*}, \tag{6.16}$$

and

$$\hat{\mathbf{U}} = \mathbf{D}_{\mathbf{m}}(\mathbf{n}^{*\prime})^{\dagger}\mathbf{v}^{*} \text{ when } s \leq t^{*}, \tag{6.17}$$

to estimate $\overline{\mathbf{V}}$ and $\mathbf{U}$ respectively.

Result 6.2 *Suppose that two of the recovery strata are pooled. Also suppose that Assumptions 1-3 are satisfied and matrices* $\overline{\mathbf{n}}^{*}$ *and* $\mathbf{n}^{*}$ *are of full rank.*

R.6.2.1

(a) If Assumption 5 is satisfied, and $s \geq t^{*}$*, then the following Condition 3 is sufficient for the estimator given by (6.16) to be consistent for* $\overline{\mathbf{V}}^{*}$.

(b) If Assumptions 5 and 6 are satisfied, and $s \geq t^{*}$*, then the following Condition 4 is sufficient for the estimator given by (6.16) to be consistent for* $\overline{\mathbf{V}}^{*}$.

R.6.2.2

If Assumption 6 is satisfied, and $s \leq t^{*}$*, then the following Conditions 3 and 4 are individually sufficient for the estimator given by (6.17) to be consistent for* $\mathbf{U}$.

Condition 3

The sampling proportion is the same for both pooled strata; i.e, $p_{t-1} = p_{t}$. *(This is the analogue of Darroch's sufficient condition (i) for the Petersen estimator to be consistent.)*

Condition 4

The probabilities of moving to pooled recovery strata from any given tagging stratum are constant multiples of each other; i.e., $\theta_{it}/\theta_{i(t-1)} = k$ *(a constant) for* $i = 1, \ldots, s$.

Proof

Let

$$\tilde{\mathbf{p}} = \begin{pmatrix} p_1 \\ \vdots \\ p_{t-2} \\ \tilde{p}_{t-1} \end{pmatrix} \quad \text{for some } 0 < \tilde{p}_{t-1} \leq 1. \tag{6.18}$$

Note that if

$$\frac{\overline{V}_{t-1} p_{t-1} + \overline{V}_t p_t}{\overline{V}_{t-1} + \overline{V}_t} = \tilde{p}_{t-1}, \tag{6.19}$$

then

$$\mathbf{D}_{\overline{\mathbf{v}}^*} = \mathbf{D}_{\overline{\mathbf{V}}^*} \mathbf{D}_{\tilde{\mathbf{p}}}. \tag{6.20}$$

Next, note that for $i = 1, \ldots, s$ and $j = 1, \ldots (t-2)$

$$E(n^*_{ij}) \;=\; E(n_{ij}) = m_i \theta_{ij} p_j = m_i \theta^*_{ij} \tilde{p}_j, \tag{6.21}$$

and for $i = 1, \ldots, s$

$$\begin{aligned} E\left[n^*_{i(t-1)}\right] &= E(n_{i(t-1)} + n_{it}) = m_i \theta_{i(t-1)} p_{t-1} + m_i \theta_{it} p_t \\ &= m_i \left[\theta_{i(t-1)} + \theta_{it}\right] \left[\frac{\theta_{i(t-1)} p_{t-1} + \theta_{it} p_t}{\theta_{i(t-1)} + \theta_{it}}\right] \\ &= m_i \theta^*_{i(t-1)} \left[\frac{\theta_{i(t-1)} p_{t-1} + \theta_{it} p_t}{\theta_{i(t-1)} + \theta_{it}}\right]. \end{aligned} \tag{6.22}$$

Hence, if

$$\tilde{p}_{t-1} \;=\; \left[\frac{\theta_{i(t-1)} p_{t-1} + \theta_{it} p_t}{\theta_{i(t-1)} + \theta_{it}}\right] \quad \text{for } i = 1, \ldots, s, \tag{6.23}$$

then

$$\overline{\mathbf{n}}^* = \mathbf{D}_{\mathbf{m}} \Theta^* \mathbf{D}_{\tilde{\mathbf{p}}}. \tag{6.24}$$

I.e, if

$$\frac{\overline{V}_{t-1}p_{t-1}+\overline{V}_t p_t}{\overline{V}_{t-1}+\overline{V}_t} = \left[\frac{\theta_{i(t-1)}p_{t-1}+\theta_{it}p_t}{\theta_{i(t-1)}+\theta_{it}}\right], \quad \text{for } i=1,\ldots,s \tag{6.25}$$

then pooled data satisfy key equations in Subsection 5.2.2. Since Condition 3 satisfies (6.25), this condition is sufficient for the the pooled data to satisfy the first two key equations in Subsection 5.2.2. If Assumption 5 is satisfied, the third key equation is satisfied by Θ^*. This proves part (a) of R.6.2.1.

To prove part (b) of R.6.2.1, suppose that both Assumptions 5 and 6 are satisfied. Then for $j = 1,\ldots,t-2$,

$$\begin{aligned} E(v_j^*) &= \sum_i U_i\theta^*_{ij}\tilde{p}_j & (6.26)\\ &= \overline{V}_j\tilde{p}_j, & (6.27)\end{aligned}$$

and

$$\begin{aligned} E(v_{t-1}^*) &= \sum_i^s U_i\theta_{i(t-1)}p_{t-1} + \sum_i^s U_i\theta_{it}p_t \\ &= \sum_i^s U_i[\theta_{i(t-1)}p_{t-1}+\theta_{it}p_t] \\ &= \sum_i^s U_i\theta^*_{i(t-1)}\left[\frac{\theta_{i(t-1)}p_{t-1}+\theta_{it}p_t}{\theta_{i(t-1)}+\theta_{it}}\right]. & (6.28)\end{aligned}$$

If (6.23) is satisfied, then (6.27) and (6.28) can be written as (6.20). We have already shown that (6.24) is satisfied if (6.23) is satisfied. So, now, it is sufficient that

$$\left[\frac{\theta_{i(t-1)}p_{t-1}+\theta_{it}p_t}{\theta_{i(t-1)}+\theta_{it}}\right] = \text{a constant} \quad \text{for } i=1,\ldots,s, \tag{6.29}$$

for the pooled data to satisfy the key equations in Subsection 5.2.2. Condition 4 satisfies (6.29). This proves part (b) of R.6.2.1.

Next, to prove R.6.2.2, recall that (6.24) is valid if (6.23) is satisfied. Also, note from (6.26) and (6.28) that if (6.23) is satisfied, then

$$\mathbf{D}_{\overline{\mathbf{V}}^*} = \mathbf{U}'\Theta^*\mathbf{D}_{\tilde{\mathbf{p}}}. \tag{6.30}$$

In other words, both two key equations in Subsection 5.2.3 are satisfied if (6.29) is satisfied. Conditions 3 and 4 both individually satisfy (6.29). This proves R.6.2.2.

6.3 Partial Pooling of Tagging and Recovery Strata

Usually, in practice, the investigator may need to pool tagging strata as well as recovery strata. Then $\mathbf{U}^*$ and $\overline{\mathbf{V}}^*$ will be estimated by

$$\widehat{\overline{\mathbf{V}}}^* = \mathbf{D}_{\mathbf{v}^*}(\mathbf{n}^*)^{\dagger}\mathbf{m}^* \quad \text{when } s^* \geq t^*, \tag{6.31}$$

and

$$\widehat{\mathbf{U}}^* = \mathbf{D}_{\mathbf{m}^*}(\mathbf{n}^*)^{\dagger}\mathbf{v}^* \quad \text{when } s^* \leq t^*. \tag{6.32}$$

to estimate $\overline{\mathbf{V}}$ and $\mathbf{U}$ respectively. We now combine Results 1 and 2 to derive sufficient conditions for these estimators to be consistent.

Result 6.3 *Suppose that two of the tagging strata and two of the recovery are pooled. Also suppose that Assumptions 1-3 are satisfied and matrices $\overline{\mathbf{n}}^*$ and $\mathbf{n}^*$ are of full rank.*

R.6.3.1

(a) *If Assumption 5 is satisfied, and $s^* \geq t^*$, then Condition 3 is sufficient for the estimator given by (6.31) to be consistent for $\overline{\mathbf{V}}^*$.*

(b) *If Assumptions 5 and 6 are satisfied, $s^* \geq t^*$, then Condition 4 is sufficient for the estimator given by (6.31) to be consistent for $\overline{\mathbf{V}}^*$.*

R.6.3.2

If Assumption 6 is satisfied, and $s^* \leq t^*$, then the following combinations of conditions are sufficient for the estimator given by (6.32) to be consistent for $\mathbf{U}^*$.

(a) *Conditions 1 and 3.*

(b) Conditions 1 and 4.

(c) Conditions 2 and 3.

(d) Conditions 2 and 4.

So far we have considered pooling two tagging strata and two recovery strata only. But, investigators often pool more than two tagging and recovery strata to form more than one group of pooled strata of each type. Sufficient conditions for such poolings can be derived from Results 6.1, 6.2 and 6.3. In such cases, the conditions should be true for each group of strata that is pooled.

We can use Results 6.1, 6.2 and 6.3 to verify Darroch's conditions (for the particular case where all the tagging and recovery strata are pooled). To see this, first suppose that Assumptions 1-3 and 5 are satisfied and Darroch's condition (i) is true. Then, R.6.3.1(a) implies that $\widehat{V}_+$ is consistent for $\overline{V}_+$. Secondly suppose that Assumptions 1-3 and 6 are satisfied and Darroch's condition (ii) or (iii) is true. Then, by R.6.1.2, all the tagging strata can be pooled without affecting consistency of $\widehat{U}^*_+$. At this point, the corresponding $\tilde{\Theta}$ is $1 \times t$, and automatically satisfies Condition 4. So, by R.6.2.2, all the recovery strata can be pooled to get a consistent estimator of U_+. If in addition Assumption 1 is satisfied, this estimates $\overline{V}_+$.

6.4 Variances of Pooled Estimates

Having listed various conditions under which the pooled estimates are valid, let us now suppose that the investigator correctly assumes that one of them is satisfied and conducts an appropriate pooling. Although the investigator can consistently estimate the number of animals by substituting the pooled data in an estimator derived in Section 5.2, the variance formulae derived in Section 5.4 may no longer be valid for the pooled data. We now examine sufficient conditions for those formulae to correctly estimate the variances. The following discussion considers the variance formula for $\widehat{\mathbf{V}}$. The complicated form of the variance formula for $\widehat{\mathbf{U}}$ causes difficulties in determining the effect of pooling on that formula, and is not considered here.

Complete Pooling

First, suppose that one of Darroch's conditions is valid, and an unstratified experiment has been conducted. According to Darroch, then the approximate mean squared error of the Petersen estimate is

$$E[(\hat{V}_+ - V_+)^2] \approx (\nu^2 m_+^2/\eta^4) Var(n_+) + (m_+/\eta^2) Var(v_+), \tag{6.33}$$

where $\nu = E(v_+) = \sum_j V_j p_j$ and $\eta = E(n_+) = \sum_{i,j} m_i \theta_{i,j} p_j$. In order to estimate the mean squared error, the investigator has no course but to substitute estimates:

$$\begin{aligned}
\hat{\nu} &= \hat{V}_+(n_{++}/m_+), \\
\hat{\eta} &= m_+(n_{++}/m_+) = n_{++}, \\
\widehat{Var}(n_+) &= m_+(n_{++}/m_+)(1 - n_{++}/m_+), \quad \text{and} \\
\widehat{Var}(v_+) &= v_+(n_{++}/m_+)(1 - n_{++}/m_+)
\end{aligned}$$

in (6.33). Darroch showed that this leads to overestimation, unless the sampling probabilities (the p_j's) are equal.

Partial Pooling

Now suppose that some of the tagging and recovery strata are partially pooled under sufficient conditions provided by Result 6.3. Then, the calculations (5.54)-(5.77) in Subsection 5.4.1 and (5.86)-(5.104) in Subsection 5.4.2 are valid for pooled quantities. However, the evaluation of second moments of $\mathbf{X}$ and $\mathbf{y}$, based on probability models (5.4), (5.5) and (5.42) may or may not be valid.

For example, we shall show that when the tagging strata are pooled, the variances of $\widehat{\widehat{\mathbf{V}}}$ are estimated correctly. By contrast, when the recovery strata are pooled, the variances are overestimated unless the sampling probabilities are equal. To see this, first suppose that two of the tagging strata are pooled under conditions stated in Result 6.1. Without loss of generality, change the labels of tagging strata so that the pooled strata will be the last two.

Now, to check the validity of variance formulae (5.83) for the pooled data, it is enough to examine the validity of equations (5.79) and (5.81) for the pooled data.

It is easy to see that equation (5.81) is valid since only the tagging strata are pooled. Following (5.79), the investigator is obliged to use

$$m_i^* \left[\sum_j \tilde{\theta}_{ij} \rho_j - 1 \right] = m_i^* \mu_i^*, \tag{6.34}$$

in place of $\sum_{jk} \rho_j \rho_k E[X_{ij}^* X_{ik}^*]$, the (i,i)th element of $E[\mathbf{X}^* \boldsymbol{\rho} \boldsymbol{\rho}' \mathbf{X}^{*\prime}]$. Clearly, this is correct for $i = 1, \ldots, s-2$, since those strata are not pooled. Let us check its validity for $i = s-1$.

Note that for $j = 1, \ldots, t$,

$$\begin{aligned} E\left[(X_{(s-1)j}^*)^2\right] &= Var\left[n_{(s-1)j}^*\right] \\ &= Var\left[n_{(s-1)j} + n_{sj}\right] \\ &= Var\left[n_{(s-1)j}\right] + Var\left[n_{sj}\right] \\ &= E\left[(X_{(s-1)j})^2\right] + E\left[(X_{sj})^2\right]. \end{aligned}$$

Next note that for $j \neq k$,

$$\begin{aligned} E\left[X_{(s-1)j}^* X_{(s-1)k}^*\right] &= Cov\left[n_{(s-1)j}^*, n_{(s-1)k}^*\right] \\ &= Cov\left[n_{(s-1)j}, n_{(s-1)k}\right] + Cov\left[n_{sj}, n_{sk}\right] \\ &= E\left[X_{(s-1)j} X_{(s-1)k}\right] + E\left[X_{sj} X_{sk}\right]. \end{aligned}$$

Therefore, the $(s-1, s-1)$th element of $E[\mathbf{X}^* \boldsymbol{\rho} \boldsymbol{\rho}' \mathbf{X}^{*\prime}]$ is

$$\begin{aligned} \sum_{jk} \rho_j \rho_k E[X_{(s-1)j}^* X_{(s-1)k}^*] &= \sum_{jk} \rho_j \rho_k E[X_{(s-1)j} X_{(s-1)k}] + \sum_{jk} \rho_j \rho_k E[X_{sj} X_{sk}] \\ &= m_{s-1} \left[\sum_j \theta_{(s-1)j} \rho_j - 1 \right] + m_s \left[\sum_j \theta_{(s)j} \rho_j - 1 \right] \\ &= \sum_j \left(m_{s-1} \theta_{(s-1)j} + m_s \theta_{(s)j} \right) \rho_j - (m_{s-1} + m_s). \end{aligned}$$

But, according to (6.8),

$$m_{s-1} \theta_{(s-1)j} + m_s \theta_{sj} = (m_{s-1} + m_s) \tilde{\theta}_{(s-1)j} = m_{s-1}^* \tilde{\theta}_{(s-1)j}.$$

Hence, $\sum_{jk}\rho_j\rho_k E[X^*_{(s-1)j}X^*_{(s-1)k}]$ is in fact equal to (6.34). This implies that when only tagging strata are pooled under sufficient conditions, then, the variance formulae given in Section 5.4.1 are valid for pooled data.

Now, suppose that two of the recovery strata are pooled under sufficient conditions given by Result 6.2. Without loss of generality, change the labels of recovery strata so that the pooled strata will be the last two. Now, to estimate $Var(\widehat{\mathbf{V}}^*)$, the investigator is obliged to use

$$m_i\left[\sum_{j=1}^{t-1}\theta^*_{ij}\tilde{\rho}_j - 1\right],$$

in place of $\sum_{j=1}^{t-1}\sum_{k=1}^{t-1}\tilde{\rho}_j\tilde{\rho}_k E[X^*_{ij}X^*_{ik}]$, the (i,i)th element of $E[\mathbf{X}^*\tilde{\rho}\tilde{\rho}'\mathbf{X}^{*\prime}]$. To check the validity of this, note that

$$\begin{aligned}
\sum_{j=1}^{t-1}\sum_{k=1}^{t-1}\tilde{\rho}_j\tilde{\rho}_k E[X^*_{ij}X^*_{ik}] &= \sum_{j=1}^{t-2}\sum_{k=1}^{t-2}\rho_j\rho_k E[X_{ij}X_{ik}] \\
&\quad +\sum_{j=1}^{t-2}2\rho_j\tilde{\rho}_{t-1}E[X_{ij}(X_{i(t-1)}+X_{it})] \\
&\quad +\tilde{\rho}^2_{t-1}E[(X_{i(t-1)}+X_{it})^2] \\
&= m_i\sum_{j=1}^{t-2}\rho_j^2(\theta_{ij}p_j-\theta^2_{ij}p_j^2) \\
&\quad -2m_i\sum_{j=1}^{t-3}\sum_{k=j+1}^{t-2}\rho_j\rho_k(\theta_{ij}p_j\theta_{ik}p_k) \\
&\quad -2m_i\sum_{j=1}^{t-2}\rho_j\tilde{\rho}_{t-1}\left(\theta_{ij}p_j\theta_{i(t-1)}p_{t-1}+\theta_{ij}p_j\theta_{it}p_t\right) \\
&\quad +m_i\tilde{\rho}^2_{t-1}\left[(\theta_{i(t-1)}p_{t-1}-\theta^2_{i(t-1)}p^2_{t-1})+(\theta_{it}p_t-\theta^2_{it}p_t^2)\right] \\
&\quad -2m_i\tilde{\rho}^2_{t-1}(\theta_{i(t-1)}p_{t-1}\theta_{it}p_t) \\
&= m_i\sum_{j=1}^{t-2}\rho_j^2(\theta_{ij}p_j-\theta^2_{ij}p_j^2) \\
&\quad +m_i\tilde{\rho}^2_{t-1}\left[(\theta_{i(t-1)}p_{t-1}+\theta_{it}p_t)-(\theta_{i(t-1)}p_{t-1}+\theta_{it}p_t)^2\right] \\
&\quad -2m_i\sum_{j=1}^{t-3}\sum_{k=j+1}^{t-2}\rho_j\rho_k(\theta_{ij}p_j\theta_{ik}p_k)
\end{aligned}$$

$$-2m_i \sum_{j=1}^{t-2} \rho_j \tilde{\rho}_{t-1} \left(\theta_{ij} p_j \theta_{i(t-1)} p_{t-1} + \theta_{ij} p_j \theta_{it} p_t\right).$$

Now, recall from (6.15) and (6.23) that under sufficient conditions,

$$\theta_{i(t-1)} p_{t-1} + \theta_{it} p_t = \theta^*_{t-1} \tilde{p}_{t-1}.$$

Hence, the above can be written as

$$\begin{aligned}
\sum_{j=1}^{t-1} \sum_{k=1}^{t-1} \tilde{\rho}_j \tilde{\rho}_k E[X^*_{ij} X^*_{ik}] &= m_i \sum_{j=1}^{t-1} (\theta^*_{ij} \tilde{\rho}_j - (\theta^*_{ij})^2) - 2m_i \sum_{j=1}^{t-2} \sum_{k=j+1}^{t-1} \theta^*_{ij} \theta^*_{ik} \\
&= m_i \sum_{j=1}^{t-1} \sum_{k=1}^{t-1} (\delta_{jk} \theta^*_{ij} \tilde{\rho}_j - \theta^*_{ij} \theta^*_{ik}) \\
&= m_i \sum_{j=1}^{t-1} (\theta^*_{ij} \tilde{\rho}_j - 1)
\end{aligned}$$

This implies that (5.79) is correctly estimated if the pooling is done under the sufficient conditions. However, we can show that (5.81) is overestimated unless the sampling probabilities are equal. To see this, note that the investigator is obliged to use

$$\overline{V^*_j} \tilde{p}_j (1 - \tilde{p}_j) \tag{6.35}$$

in place of $E\left[(y^*_j)^2\right]$. This is correct for $j = 1, \ldots, t-2$, since those recovery strata are not pooled. To check its validity for $t-1$, note that

$$\begin{aligned}
E\left[(y^*_{t-1})^2\right] &= E\left[(y_{t-1} + y_t)^2\right] \\
&= \overline{V}_{t-1} p_{t-1} (1 - p_{t-1}) + \overline{V}_t p_t (1 - p_t).
\end{aligned}$$

Recalling from (6.19) that under sufficient conditions, $\tilde{p}_{t-1} = \frac{\overline{V}_{t-1} p_{t-1} + \overline{V}_t p_t}{\overline{V}_{t-1} + \overline{V}_t}$, one can show that

$$\overline{V^*_{t-1}} \tilde{p}_{t-1} (1 - \tilde{p}_{t-1}) - E\left[(y^*_{t-1})^2\right] = \frac{\overline{V}_{t-1} \overline{V}_t (p_{t-1} - p_t)^2}{\overline{V}_{t-1} + \overline{V}_t}.$$

This shows that the term $E\left[(y^*_{t-1})^2\right]$ is overestimated unless $p_{t-1} = p_t$. So, the overall effect of pooling recovery strata is to overestimate the variances unless the sampling probabilities in the pooled recovery strata are equal. This result is consistent with Darroch's finding that the variance of the Petersen estimate will be overestimated unless the sampling probabilities are equal.

6.5 Biases of Pooled Estimates

In general, pooling can produce biased estimators. If the pooling is done under sufficient conditions stated above, then the estimators are consistent and asymptotically unbiased. To calculate the bias that can occur otherwise, first suppose that $s^* \geq t^*$, and $\overline{\mathbf{n}}^*$ and $\mathbf{n}^*$ are of full rank. Then,

$$\begin{aligned} \hat{\bar{\boldsymbol{\rho}}} &= (\mathbf{n}^*)^{\dagger}\mathbf{m}^* = \left[(\mathbf{n}^{*\prime}\mathbf{n}^*)^{-1}\mathbf{n}^{*\prime}\right]\mathbf{m}^*, \text{ and} \\ \widehat{\overline{\mathbf{V}}^*} &= \mathbf{D}_{\mathbf{V}^*}\hat{\bar{\boldsymbol{\rho}}} = \mathbf{D}_{\mathbf{V}^*}(\mathbf{n}^*)^{\dagger}\mathbf{m}^*. \end{aligned} \tag{6.36}$$

By repeating arguments in Subsection 5.4.1 with pooled quantities, we can show that

$$\begin{aligned} E\left(\widehat{\overline{\mathbf{V}}^*}\right) &= \mathbf{D}_{\overline{\mathbf{V}}^*}\left(\overline{\mathbf{n}}^{*\prime}\overline{\mathbf{n}}^*\right)^{-1}(\overline{\mathbf{n}}^*)'\mathbf{m}^* + O(1) \\ &= \mathbf{D}_{\overline{\mathbf{V}}^*}(\overline{\mathbf{n}}^*)^{\dagger}\mathbf{m}^* + O(1). \end{aligned} \tag{6.37}$$

To see this, first let

$$\mathbf{n}^* = \overline{\mathbf{n}}^* + \mathbf{X}^*. \tag{6.38}$$

Then,

$$\mathbf{n}^{*\prime}\mathbf{n}^* = \boldsymbol{\Gamma}^* + \mathbf{Z}^*,$$

where

$$\boldsymbol{\Gamma}^* = \overline{\mathbf{n}}^{*\prime}\overline{\mathbf{n}}^*, \text{ and} \tag{6.39}$$

$$\mathbf{Z}^* = \overline{\mathbf{n}}^{*\prime}\mathbf{X}^* + \mathbf{X}^{*\prime}\overline{\mathbf{n}}^* + \mathbf{X}^{*\prime}\mathbf{X}^*. \tag{6.40}$$

Now, using the identity (5.62), we can get

$$\begin{aligned} (\mathbf{n}^{*\prime}\mathbf{n}^*)^{-1} &= [\mathbf{I} - \boldsymbol{\Gamma}^{*-1}\mathbf{Z}^*]\boldsymbol{\Gamma}^{*-1} + (\boldsymbol{\Gamma}^{*-1}\mathbf{Z}^*)^2(\mathbf{n}^{*\prime}\mathbf{n}^*)^{-1}, \text{ and} \\ (\mathbf{n}^*)^{\dagger} = (\mathbf{n}^{*\prime}\mathbf{n}^*)^{-1}\mathbf{n}^{*\prime} &= [\mathbf{I} - \boldsymbol{\Gamma}^{*-1}\mathbf{Z}^*]\boldsymbol{\Gamma}^{*-1}(\overline{\mathbf{n}}^{*\prime} + \mathbf{X}^{*\prime}) + (\boldsymbol{\Gamma}^{*-1}\mathbf{Z}^*)^2(\mathbf{n}^*)^{\dagger}. \end{aligned} \tag{6.41}$$

Now, Post-multiplying (6.41) by $\mathbf{m}^*$, substituting $\overline{\mathbf{n}}^{*\prime}\mathbf{X}^* + \mathbf{X}^{*\prime}\overline{\mathbf{n}}^* + \mathbf{X}^{*\prime}\mathbf{X}^*$ for $\mathbf{Z}^*$ in the square bracket, and arguing as in Subsection 5.4.1, it can be shown that

$$\hat{\bar{\boldsymbol{\rho}}} = \boldsymbol{\Gamma}^{*-1}\overline{\mathbf{n}}^{*\prime}\mathbf{m}^* + \text{ a random quantity with expected value } O(m_+^{-1}).$$

In other words,

$$\begin{aligned} E\left(\widehat{\tilde{\boldsymbol{\rho}}}\right) &= \boldsymbol{\Gamma}^{*-1}\overline{\mathbf{n}}^{*\prime}\mathbf{m}^* + O(m_+^{-1}) \\ &= (\mathbf{n}^*)^{\dagger}\mathbf{m}^* + O(m_+^{-1}). \end{aligned}$$

Next, let

$$\mathbf{v}^* = \overline{\mathbf{v}}^* + \mathbf{y}^* \tag{6.42}$$

Then, from (6.36)

$$\widehat{\overline{\mathbf{V}}^*} = \mathbf{D}_{\overline{\mathbf{v}}^*}\widehat{\tilde{\boldsymbol{\rho}}} + \mathbf{D}_{\mathbf{y}^*}\widehat{\tilde{\boldsymbol{\rho}}}.$$

Since $\mathbf{y}$ and $\widehat{\tilde{\boldsymbol{\rho}}}$ are independent, and $E(\mathbf{y}^*) = 0$, it follows that

$$\begin{aligned} E\left(\widehat{\overline{\mathbf{V}}^*}\right) &= \mathbf{D}_{\overline{\mathbf{v}}^*}(\mathbf{n}^*)^{\dagger}\mathbf{m}^* + O(V_+)O(m_+^{-1}) \\ &= \mathbf{D}_{\overline{\mathbf{v}}^*}(\mathbf{n}^*)^{\dagger}\mathbf{m}^* + O(1). \end{aligned}$$

If the sufficient conditions stated in Result 6.3 are satisfied, then $\mathbf{D}_{\overline{\mathbf{v}}^*}(\mathbf{n}^*)^{\dagger}\mathbf{m}^* = \overline{\mathbf{V}}^*$, and hence $\widehat{\overline{\mathbf{V}}^*}$ is asymptotically unbiased. Otherwise, it can be biased.

Now recall from (5.50) that $\overline{\mathbf{V}} = \mathbf{D}_{\overline{\mathbf{v}}}(\overline{\mathbf{n}})^{\dagger}\mathbf{m}$. Therefore, actual expected numbers of untagged animals in pooled recovery strata can be obtained by adding the corresponding $\overline{V}_j$'s in $\overline{\mathbf{V}}$. Let this be denoted by $\left(\overline{\mathbf{V}}\right)^*$. Then, the asymptotic bias of $\widehat{\overline{\mathbf{V}}^*}$ is given by $\mathbf{D}_{\overline{\mathbf{v}}^*}(\overline{\mathbf{n}}^*)^{\dagger}\mathbf{m}^* - \left(\overline{\mathbf{V}}\right)^*$.

The bias can be positive or negative depending on the movement patterns and sampling probabilities. To see this consider the following hypothetical examples. Let

$$\mathbf{U} = \begin{pmatrix} 200000 \\ 100000 \\ 300000 \\ 50000 \end{pmatrix}, \quad \mathbf{m} = \begin{pmatrix} 2000 \\ 2000 \\ 9000 \\ 500 \end{pmatrix}, \quad \boldsymbol{\Theta} = \begin{pmatrix} 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \\ 0.4 & 0.5 & 0.1 \\ 0.6 & 0.1 & 0.3 \end{pmatrix}$$

First, let

$$\mathbf{p} = \begin{pmatrix} 0.02 \\ 0.03 \\ 0.02 \end{pmatrix}.$$

Then,

$$\overline{\mathbf{n}} = \begin{pmatrix} 12 & 30 & 8 \\ 4 & 24 & 20 \\ 72 & 135 & 18 \\ 6 & 1.5 & 3 \end{pmatrix}, \quad \overline{\mathbf{v}} = \begin{pmatrix} 4400 \\ 8850 \\ 2700 \end{pmatrix}, \text{ and } \overline{\mathbf{V}} = \begin{pmatrix} 220000 \\ 295000 \\ 135000 \end{pmatrix}.$$

Pooling the last two tagging strata, and the last two recovery strata, we get

$$\mathbf{m}^* = \begin{pmatrix} 2000 \\ 2000 \\ 9500 \end{pmatrix} \quad \overline{\mathbf{n}}^* = \begin{pmatrix} 12 & 38 \\ 4 & 44 \\ 78 & 157.5 \end{pmatrix}, \text{ and } \overline{\mathbf{v}}^* = \begin{pmatrix} 4400 \\ 11550 \end{pmatrix}.$$

Hence,

$$\overline{\mathbf{V}}^* = \begin{pmatrix} 220000 \\ 430000 \end{pmatrix} \quad \text{and } E\left(\widehat{\overline{\mathbf{V}}^*}\right) \approx \begin{pmatrix} 164781 \\ 482102 \end{pmatrix}.$$

So, the bias of the total estimate is

$$\text{Bias}(\widehat{\overline{V}_+}) \approx (164781 + 482102) - (220000 + 430000) = -3117$$

Next, suppose that the sampling proportions in the last two recovery strata were interchanged. I.e,

$$\mathbf{p} = \begin{pmatrix} 0.02 \\ 0.02 \\ 0.03 \end{pmatrix}.$$

Then,

$$\overline{\mathbf{n}} = \begin{pmatrix} 12 & 20 & 12 \\ 4 & 16 & 30 \\ 72 & 90 & 27 \\ 6 & 1 & 4.5 \end{pmatrix}, \quad \overline{\mathbf{v}} = \begin{pmatrix} 4400 \\ 5900 \\ 4050 \end{pmatrix}, \text{ and } \overline{\mathbf{V}} = \begin{pmatrix} 220000 \\ 295000 \\ 135000 \end{pmatrix}.$$

Pooling the last two tagging strata, and the last two recovery strata, we get

$$\mathbf{m}^* = \begin{pmatrix} 2000 \\ 2000 \\ 9500 \end{pmatrix} \quad \overline{\mathbf{n}}^* = \begin{pmatrix} 12 & 32 \\ 4 & 46 \\ 78 & 122.5 \end{pmatrix}, \text{ and } \overline{\mathbf{v}}^* = \begin{pmatrix} 4400 \\ 9950 \end{pmatrix}.$$

Hence,

$$\overline{\mathbf{V}}^* = \begin{pmatrix} 220000 \\ 430000 \end{pmatrix} \quad \text{and} \quad E\left(\widehat{\overline{\mathbf{V}}^*}\right) \approx \begin{pmatrix} 271077 \\ 381681 \end{pmatrix}.$$

So, the bias of the total estimate is

$$\text{Bias}(\widehat{\overline{V}_+}) \approx (271077 + 381681) - (220000 + 430000) = 2758$$

Similar results can be shown for the bias of $\widehat{\mathbf{U}}$, when $s^* \leq t^*$.

6.6 Worked Examples and Monte Carlo Studies

6.6.1 Example 1: Schaefer's sockeye salmon data

As the first example, we illustrate our methods using a data set provided by Schaefer(1951). Darroch also used this data set to demonstrate his methods. So, to compare results from both approaches and to point out the advantages of our method, we reproduce main points of Darroch's calculation as well.

In this experiment both stratifications are with respect to time instead of place. The population consisted of adult sockeye salmon passing a certain point of a river in British

Columbia. The experiment was conducted by Schaefer. A total of 2,351 fish were tagged on the way to their spawning grounds, over an 8-week period. Later, samples were drawn regularly over a 9-week period as the fish spawned and died on the spawning grounds further upstream: 10,472 fish, of which 520 had been tagged were recovered in these samples. The data from the experiment are set out in Table 6.1.

Week of tagging (i)	m_i	Week of recovery (j) 1	2	3	4	5	6	7	8	9	Total
1	15	1	0	2	0	0	0	0	0	0	3
2	59	1	3	7	0	0	0	0	0	0	11
3	410	1	11	33	24	5	1	0	1	0	76
4	695	0	5	29	79	52	3	2	7	3	180
5	773	0	0	11	67	77	2	16	7	3	183
6	335	0	0	0	14	25	3	10	6	2	60
7	59	0	0	0	0	0	0	1	5	0	6
8	5	0	0	0	0	0	0	1	0	0	1
Total	2351	3	19	82	184	159	9	30	26	8	520
v_j		16	113	718	2664	3317	635	1217	904	368	9952

Table 6.1: Schaefer's data on sockeye salmon: n_{ij}, m_i and v_j for $s = 8, t = 9$ (from Darroch [1961:Table 1])

Darroch's approach

1. He reduced s and t to four by grouping the first three and last three weeks of tagging into two single strata and the first three and last four recovery strata into two single strata, since the n_{ij} in some of the outer weeks were too small. The new values of n_{ij}^*, m_i^* and v_j^* are given in Table 6.2.

Week of tagging (i)	m_i^*	Week of recovery (j)					
		1-3	4	5	6-9	Total	n_{i+}^*/m_i^*
1-3	484	59	24	5	2	90	0.186
4	695	34	79	52	15	180	0.259
5	773	11	67	77	28	183	0.237
6-8	399	0	14	25	28	67	0.168
Total		104	184	159	73		
v_j^*		874	2664	3317	3124		

Table 6.2: Schaefer's data on sockeye salmon: n_{ij}^*, m_i^* and v_j^* for $s^* = 4, t^* = 4$ (from Darroch [1961:Table 2])

2. He assumed that the ϕ_i's are equal, and estimated p^* through (5.6) as

$$(\widehat{\mathbf{p}^*})' = (0.1318, 1.9461, 0.1947, 0.1063). \tag{6.43}$$

He pointed out that the unsatisfactory value of $\widehat{p}_2^*$ may be just a symptom of the general inadequacy of the estimators, or it may indicate that the model is incorrect in assuming the ϕ_i equal.

3. He argued that n_{i+}^*/m_i^* indicate where the possible differences in the ϕ_i lie. Since the middle two are appreciably larger than the outer two (cf. Table 6.2), he suggested

imposing constraints, $\phi_1 = \phi_4$ and $\phi_2 = \phi_3$, and using (5.12), and (5.12) to obtain estimates.

4. To evaluate (5.12), $\mathbf{AD_m^{-1}n}$ should be a square matrix. To accomplish this, two more recovery strata should be pooled. It is permissible to group two recovery strata if R.6.2.1 (b) is satisfied. In Table 6.2, the only two columns that might satisfy this are the third and fourth columns. He pooled those columns.

5. Equations (5.12) and (5.52) led to the following estimates.

Recovery Stratum	$\widehat{\mathbf{p}^{**}}$	$\widehat{\mathbf{V}^{**}}$	$\sqrt{MSE(\widehat{\mathbf{V}^{**}})}$
1-3	0.1661	5,099	2,677
4	0.6223	4,282	12,936
5-9	0.1563	41,204	14,096
Total		50,585	4,573

Table 6.3: Darroch's Estimates of $\mathbf{V}^*$

Our approach

Now, we present our approach on Schaefer's data. We think that it is more meaningful to estimate the numbers at the tagging stage rather than the numbers at the recovery stage. The reason is that the survival probabilities are not equal to one. So, the estimated numbers at the recovery stage would actually estimate the numbers that would have reached recovery strata if the survival probabilities were one. However, in order to compare with Darroch's approach, we also estimate the numbers at the recovery stage. The steps followed were:

1. Pool only the recovery strata. As Darroch did, we pool the first three and last four recovery strata. Then, $s = 8, t^* = 4$.
2. Derive the least-squares estimate (LSE) and an iteratively re-weighted least-squares estimate (IRLSE) of $\tilde{\mathbf{p}}$ using (5.35) and (5.41) respectively. The least-squares estimate is taken as the initial guess for deriving the weighted least-squares estimate. The mean squared errors are estimated using (5.77). These estimates are shown in Table 6.4

Recovery	LSE			IRLSE	
Stratum	$\hat{\mathbf{p}}$	$\widehat{\mathbf{V}^*}$	$\sqrt{MSE(\widehat{\mathbf{V}^*})}$	$\hat{\mathbf{p}}$	$\widehat{\mathbf{V}^*}$
1-3	0.1314	6,444	1,648	0.1562	5,422
4	1.7899	1,488	9,128	0.4344	6,133
5	0.2084	15,914	14,134	0.3054	10,860
6-9	0.1005	31,055	12,520	0.0988	31,625
Total		54,900	6,528		54,040

Table 6.4: Our Estimates of $\mathbf{V}^*$

Even though the least-squares method produced an unacceptable estimate of the sampling probability in recovery stratum 4, the iteratively re-weighted least-squares estimate is acceptable and may be reasonable too. High root mean squared errors indicate that the estimates of individual stratum sizes are not reliable. However, the root mean squared error of the total estimate implies that this estimate can be reliable.

Next, we present our estimation of the numbers at the tagging stage. For this, we pooled only the tagging strata. The steps followed were:

1. Pool the same tagging strata as Darroch did; i.e, first three and last three tagging strata. Then, $s^* = 4$ and $t = 9$.

2. Derive the least-squares estimate (LSE) and an iteratively re-weighted least-squares estimate (IRLSE) of $\mathbf{U}^*$ using (5.35) and (5.43), respectively. The least-squares estimate is taken as the initial guess for deriving the weighted least-squares estimate. The mean squared errors are estimated using (5.91). These estimates are shown in Table 6.5.

Tagging	LSE		IRLSE
Stratum	$\widehat{\mathbf{U}^*}$	$\sqrt{MSE(\widehat{\mathbf{U}^*})}$	$\widehat{\mathbf{U}^*}$
1-3	4,343	4,049	2,301
4	4,842	13,751	10,755
5	8,693	17,806	1,790
6-8	32,755	11,555	37,614
Total	50,633	5,072	52,460

Table 6.5: Our Estimates of $\mathbf{U}^*$

Again, the high root mean squared errors indicate that the estimates of individual stratum sizes are not reliable. However, the estimates of total seem to be more reliable. This phenomena is common when the samples are not large enough.

6.6.2 Monte Carlo Studies

As an investigation of the accuracy of the above estimation for the total, we performed three Monte Carlo studies. Since the estimated U_+ in this problem is close to 50,000 fish and the total number of tagged animals is 2351, the proportion of untagged fish to tagged fish is close to 20:1. So, in our first simulation we matched this proportion with a little variation. In the other two simulations we gradually decreased this proportion allowing a large variation. But, the released numbers of tagged animals were doubled to produce reasonably large numbers of recovered tagged animals. The steps taken in the simulations are as follows:

1. Let $\mathbf{m}$ be the numbers of tagged fish released in Schaefer's data.
2. $\widehat{\psi} = \mathbf{D}_{\mathbf{m}}^{-1}\mathbf{n}$ was calculated from Schaefer's data.
3. Matrix $\widehat{\psi}$ was augmented by including the $(t+1)$th column of probabilities $(\mathbf{1} - \widehat{\psi}\mathbf{1})$. Let the augmented matrix be denoted by $\tilde{\psi}$.
4. Let $Y(a, b)$ denote a rounded uniform random number in the interval (a, b). In the three Monte Carlo studies, the numbers of untagged animals at the tagging stage ($\mathbf{U}$) were simulated as,
   (a)
   $$\begin{aligned} \mathbf{U}' &= 20\ \mathbf{m}' + [Y(0,\ 50),\ Y(0,\ 50),\ Y(0,\ 100),\ Y(0,\ 2{,}000), \\ &\quad Y(0,\ 2{,}000),\ Y(0,\ 5{,}000),\ Y(0,\ 1{,}000),\ Y(0,\ 500),\ Y(0,\ 400)], \end{aligned}$$
   (b) $U_i = 2\, m_i \times Y(100, 300)$, and
   (c) $U_i = 2\, m_i \times Y(250, 400)$
5. The recovered numbers of tagged and untagged numbers ($\mathbf{n}$ and $\mathbf{v}$) were simulated according to a multinomial model with probabilities given by $\tilde{\psi}$ and numbers at the tagging stage as follows.
   (a) the numbers of tagged animals = $\mathbf{m}$, and the numbers of untagged animals = $\mathbf{U}$ as in (a) above.
   (b) the numbers of tagged animals = $2\mathbf{m}$, and the numbers of untagged animals = $\mathbf{U}$ as in (b) above.
   (c) the numbers of tagged animals = $2\mathbf{m}$, and the numbers of untagged animals = $\mathbf{U}$ as in (c) above.

6. The least squares estimates of the total numbers of untagged fish at the tagging stage (U_+) were calculated based on the generated data, using the same steps as in estimation of Example 1 (pooling first three and last three tagging strata, and so on).

7. Percent relative error of the total was calculated as $\frac{(\hat{U}_+ - U_+)}{U_+} \times 100$.

8. Repeated steps 4-7 1000 times.

The histograms and box-plots of the observed percent relative errors of overall totals are shown in Figure 6.1. The figures corresponding to three studies are labeled as (a), (b) and (c) respectively.

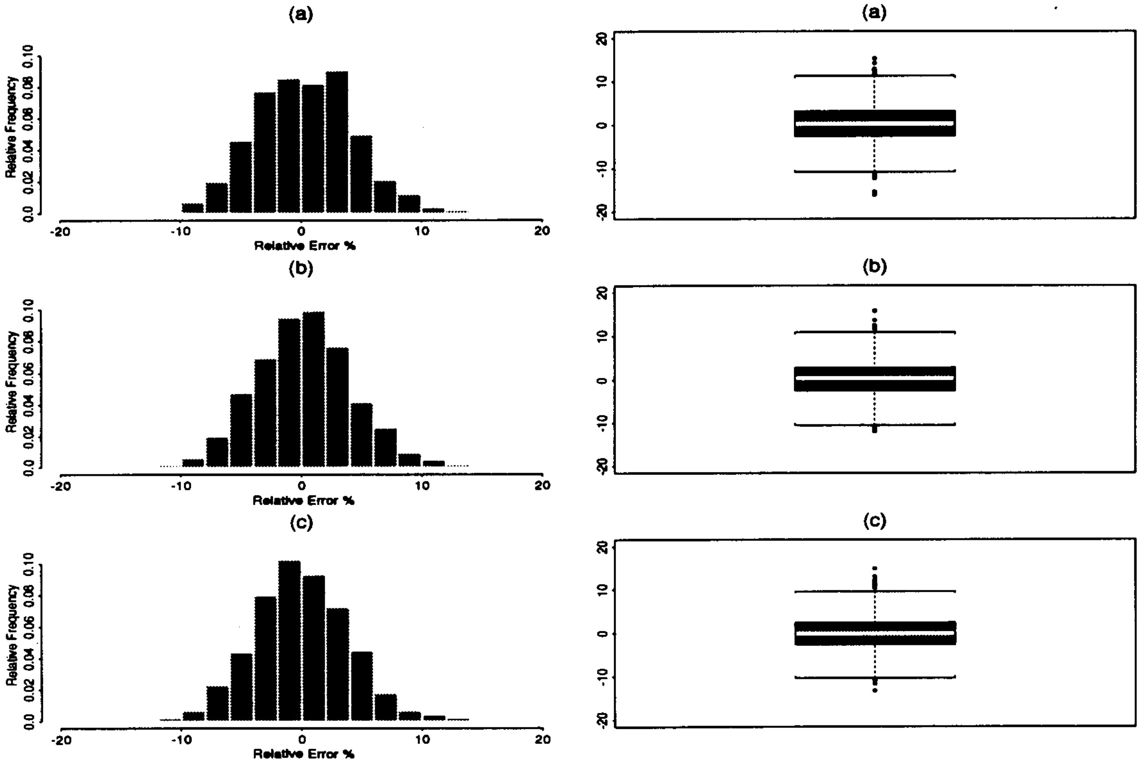


Figure 6.1: Percent relative error of $\hat{U}_+$ in three Monte Carlo studies (each study is based on 1000 simulations).

In the first study, the tagging proportions were close to 1/20. But the released numbers of tagged animals were rather small (as in Schaefer's data). Because of this, the n_{ij}'s were of the same sizes as in Schaefer's data set. In a number of simulations, the estimates of individual stratum sizes were negative. However, on average, the total estimates seem to be very reliable. It can be seen from the histogram (a) and the box-plot (a), that the percent absolute relative errors of the totals are reasonably low. In the second Monte Carlo study, the numbers of marked animals were twice the amount in Schaefer's data set. The tagging proportions were smaller; between 1/100 and 1/300. In the third study, the tagging proportions were further reduced to proportions between 1/250 and 1/400. In both studies, a number of negative estimates of stratum sizes were observed. However, the total number of fish at the tagging stage appears to have been estimated accurately with high probability.

6.6.3 Example 2: Schwarz's pink salmon data

Now we derive estimates using a data set provided by Dr. Carl Schwarz of the Department of Mathematics and Statistics, Simon Fraser University. These data were collected in the Fraser River, British Columbia. In this experiment, tagged male pink salmon were released over 6 weeks. As the fish reached each of 6 spawning areas, they were again sampled over several weeks. This produced 31 recovery strata, considering each recovery week at each spawning ground as a stratum. The numbers of marked and unmarked animals released and recovered are presented in Tables 6.6 and 6.7.

Area of recovery	Week of recovery	Week of tagging						Total number of unmarked fish recovered
		1	2	3	4	5	6	
		Number of marked fish released						
		3479	5354	4219	4227	1588	327	
		Number of marked fish recovered						
1	2	14	15	7	3	0	0	9,004
	3	5	12	7	3	0	0	9,279
	4	0	4	4	6	1	0	5,984
	5	0	5	1	4	0	0	3,287
	6	0	0	2	0	1	1	745
2	3	0	1	1	6	1	1	1,299
	4	0	2	6	22	15	1	8,374
	5	0	1	7	29	38	2	18,843
	6	1	2	5	23	42	10	19,401
	7	0	1	1	8	17	11	8,975
3	3	0	3	3	0	0	0	916
	4	0	5	7	7	7	1	5,837
	5	0	0	1	1	5	2	3,960
	6	0	0	1	3	2	0	1,710
	7	0	0	0	0	0	0	29

Continued on next page

Table 6.6: Fraser River pink salmon data: Areas 1-3

Area of recovery	Week of recovery	Week of tagging 1	2	3	4	5	6	Total number of unmarked fish recovered
		Number of marked fish recovered						
4	3	4	7	9	2	0	0	11,857
	4	5	11	15	9	1	0	22,540
	5	1	1	2	3	0	0	3,372
5	1	0	2	0	0	0	0	230
	2	9	6	2	0	0	0	2,636
	3	0	16	14	8	1	0	6,643
	4	0	11	17	35	11	1	22,295
	5	0	2	3	4	4	1	8,992
	6	0	0	0	0	0	0	1,508
6	1	4	3	2	0	0	0	226
	2	6	26	24	4	0	0	6,201
	3	14	48	55	45	3	0	10,583
	4	3	9	20	17	10	0	26,053
	5	0	0	0	6	1	0	11,545
	6	0	0	0	1	4	1	2,776
	7	0	0	0	0	0	0	394
Total in 6 Areas		66	193	216	249	164	32	235,494

Table 6.7: Fraser River pink salmon data: Areas 4-6

From these data, it is more meaningful to estimate the numbers of fish at the tagging strata than to estimate the numbers of fish at the recovery strata. There are two reasons for this: The first is that Assumption 6 may be more reasonable than Assumption 5. The second is that the number of tagging strata is smaller than the number of recovery strata ($s < t$). Following is a list of the more noteworthy features.

1. In this data set, there are several recovery strata in which the numbers of tagged fish are zero or very small. It may be reasonable to assume that within each recovery area, the sampling proportions are kept roughly equal. Then, it is permissible to pool strata within areas. So, we pool recovery strata as follows:

   Area 1: weeks 4-6, Area 2: weeks 3-4,
   Area 3: weeks 3-4, and 5-7, Area 4: weeks 4-5,
   Area 5: weeks 1-2, and 5-6, Area 6: weeks 1-2, and 4-7.

   The pooled data set is shown in Table 6.8.

Week of tagging (i)	Area of Recovery																	
	1			2				3		4		5				6		
1	14	5	0	0	0	1	0	0	0	4	6	9	0	0	0	10	14	3
2	15	12	9	3	1	2	1	8	0	7	12	8	16	11	2	29	48	9
3	7	7	7	7	7	5	1	10	2	9	17	2	14	17	3	26	55	20
4	3	3	10	28	29	23	8	7	4	2	12	0	8	35	4	4	45	24
5	0	0	2	16	38	42	17	7	7	0	1	0	1	11	4	0	3	15
6	0	0	1	2	2	10	11	1	2	0	0	0	0	1	1	0	0	1

Table 6.8: Fraser River pink salmon data: Recovered numbers of marked fish (after pooling).

2. The least squares estimates obtained using the pooled data is shown in Table 6.9. One of the estimates is unacceptable since it is negative. Such an unacceptable estimate can be a result of incorrect modeling or near singularities of the matrix $\mathbf{n}$. Table 6.8 suggests that there may be a collinearity between the second and third rows of corresponding $\tilde{\Theta}$ matrix (p-value =.12). So we pool those two tagging strata. In addition, the counts in the last row are very small except in Area 2. So, we pool the last two tagging strata as well. The least-squares and weighted least-squares estimates obtained using pooled data are presented in Table 6.10

	Week of tagging (i)						Total
	1	2	3	4	5	6	
$\hat{U}$	3,995,624	-7,360,795	5,652,406	305,512	534,366	64,331	3,191,444

Table 6.9: Estimated numbers of unmarked fish at the tagging stage

Week of tagging	LSE		IRLSE
	$\widehat{U^*}$	$\sqrt{MSE(\widehat{U^*})}$	$\widehat{U^*}$
1	1,696,810	2,000,490	1,476,959
2-3	82,683	1,718,784	1,124,423
4	1,738,166	1,183,717	1,910,050
5-6	493,535	293,300	460,718
Total	4,011,194	1,144,455	4,972,150

Table 6.10: Estimated numbers of unmarked fish at the tagging stage

6.7 Discussion

In stratified tag recovery experiments it is common to pool of some or all of the strata. The estimate of the population size in such a case may be inconsistent. Darroch [13] has described conditions for complete pooling which are sufficient for the resulting pooled estimate to be consistent.

We discussed the case of partial pooling and suggested conditions which are sufficient to preserve the consistency of the estimates. Conditions 1 and 3, proportional tagging and proportional sampling are potentially at the investigator's control, although they are difficult to implement without prior knowledge of the population sizes. However, preliminary estimates of proportional sizes of strata may be obtained from other sampling methods such as the 'CPUE' (catch per unit effort) method. Then it may be possible to maintain the conditions of proportional tagging and proportional sampling.

Naturally, the Conditions 2 and 4, which are on the movement probabilities, are not at the investigator's control. However, when they are satisfied, it is possible to detect the presence of collinearity caused by these conditions. Then the corresponding strata may be pooled to obtain consistent estimates.

If these conditions are met, the investigators can use the pooled data to derive estimates of the stratum sizes. However, the variance estimators provided for unpooled estimates are not necessarily be valid for the pooled data. For example, as Darroch showed that when all the tagging and recovery strata are pooled, the variance can be overestimated unless the sampling probabilities (the p_j's) are equal. We showed that under partial pooling of tagging strata, the variance of $\widehat{\mathbf{V}}^*$ is correctly estimated using formula (5.83) with pooled data. We also showed that effect of pooling recovery strata on this formula is to overestimate the variances unless the sampling probabilities are equal in the pooled strata. We were unable to determine the effect of pooling on the variance formula for $\widehat{\mathbf{U}}$, because of its complicated form. This problem is open for further research.

Appendix: Some Useful Results

In this appendix, we present some of the important results that were used throughout the thesis.

According to Sen and Singer [34](pg. 43),

Theorem 1 (Jensen's inequality) *Let X be a random variable, and $g(x)$, $x \in \mathbf{R}$, be a convex function such that $E[g(X)]$ exists. Then,*

$$E[g(X)] \geq g(E[X]),$$

with the equality sign holding only when g is linear almost everywhere.

According to Marsden and Tromba [26](pg. 274),

Theorem 2 *Let $f : U \subset \mathbf{R}^2 \to \mathbf{R}$ and $g : U \subset \mathbf{R}^2 \to \mathbf{R}$ be smooth functions. Let $\mathbf{v}_0 \in U$, $g(\mathbf{v}_0) = c$, and S be the level curve for g with value c. Assume that $\nabla g(\mathbf{v}_0) \neq \mathbf{0}$ and that there is a real number λ such that $\nabla f(\mathbf{v}_0) = \lambda \nabla g(\mathbf{v}_0)$. Form the auxiliary function $h = f - \lambda g$ and the bordered Hessian determinant*

$$|\overline{H}| = \begin{vmatrix} 0 & -\frac{\partial g}{\partial x} & -\frac{\partial g}{\partial y} \\ -\frac{\partial g}{\partial x} & \frac{\partial^2 h}{\partial x^2} & \frac{\partial h}{\partial x \partial y} \\ -\frac{\partial g}{\partial y} & \frac{\partial^2 h}{\partial x \partial y} & \frac{\partial^2 h}{\partial y^2} \end{vmatrix} \text{ evaluated at } \mathbf{v}_0.$$

1. *If $|\overline{H}| > 0$, then $\mathbf{v}_0$ is a local maximum point of f restricted to S.*

2. *If* $|\overline{H}| < 0$, *then* v_0 *is a local minimum point of* f *restricted to* S.

3. *If* $|\overline{H}| = 0$, *the test is inconclusive and* v_0 *may be a minimum, a maximum, or neither.*

According to Kotz and Johnson [21](pp. 386-387),

Theorem 3 (Cauchy-Schwarz inequality) *Let* (X, Y) *be a bivariate random variable. Then,*

$$E(X^2Y^2) \leq E(X^2)E(Y^2),$$

provided the expectation on the left hand side exists.

According to Kotz and Johnson [22](pg. 515),

Theorem 4 (Slutsky-Frechet) *If the sequence of random variables* $\{X_n\}$ *converges in probability to a random variable* X, *then so does* $f(X_n)$ *to* $f(X)$ *for any continuous function* f.

According to Albert [1](pg. 30),

Theorem 5 x_0 *minimizes*

$$||z - Hx||^2$$

if and only if x_0 *is of the form*

$$x_0 = H^\dagger z + (I - H^\dagger H)y$$

for some y. *The value of* x *which minimizes* $||z - Hx||^2$ *is unique if and only if* $H^\dagger H = I$.

Bibliography

[1] Albert, A. (1972). *Regression and the Moore-Penrose Pseudoinverse.* Academic press: New York.

[2] Banneheka, S.G., Routledge, R.D., Guthrie, I.C., and Woodey, J. C. (1995). Estimation of In-River Fish Passage Using a Combination of Transect and Stationary Hydroacoustic Sampling. To appear in *Can. J. Fish. Aquat. Sci.*

[3] Barndorff-Nielsen, O.E., and Cox, D.R. (1984). Bartlett Adjustments to the Likelihood Ratio Statistic and Distribution of the maximum Likelihood Estimator. *J.R. Statistics Soc. B, 46, No.3*, 483-495.

[4] Bartlett, M.S. (1937). Properties of sufficiency of and statistical tests. *Proc. R. Soc. A, 160*, 268-282.

[5] Berger, J. O., and Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *Journal of American Statistical Association. Vol. 84, No. 405*, pp 200-207.

[6] Burczynski, J. (1979). *Introduction to the use of sonar systems for estimating fish biomass. FAO Fisheries Technical Paper No.191.*

[7] Chapman, R. J. H., and Junge, C. O. (1956). The estimation of the size of a stratified animal population. *Ann. Math. Statist. 27*, 375-389.

[8] Clay, C. S., and Medwin, H. (1977). *Acoustical oceanography: principles and applications.* Wiley-Interscience, New York, NY. 544 p.

[9] Cormack, R.M., and Skalski, J.R. (1992). Analysis of coded wire tag returns from commercial catches.*Can. J. Fish. Aquat. Sci. 49*, 1816-1825.

[10] Cox, D.R., and Hinkley, D.V. (1974). *Theoretical statistics*, London: Chapman and Hall

[11] Crittenden, R.N. (1989). *Abundance estimation based on Echo counts.* Ph.D thesis, University of Washington.

[12] Crittenden, R.N., Thomas, G.L., Marino, D.A., and Thorne, R.E. (1988). A weighted duration-in-beam estimator for the volume sampled by a quantitative echo sounder. *Can. J. Fish. Aquat. Sci. 45*, 1249-1256.

[13] Darroch, J. N. (1961). The two sample capture-recapture census when tagging and sampling are stratified. *Biometrika, 48*, 241-260.

[14] Draper, N. R., and Smith, H. (1986). *Applied regression analysis.* John Wiley and Sons, New York, NY. 709 p.

[15] Foote, K. G. (1990). Correcting acoustic measurements of scatterer density for extinction. *J. Acoust. Soc. Am. 88*, 1543-1546.

[16] Johannesson, K.A., and Mitson, R.B. *Fisheries acoustics. A practical manual for aquatic biomass estimation. FAO Fish. Tech. Pap.*

[17] Kendall, M., and Stuart, A. (1977). *The advanced theory of statistics. Distribution theory.* p. 261-262. Macmillan, New York. NY.

[18] Kieser, R., and Ehrenberg, J. E. (1990). An unbiased, stochastic echo counting model. *Rapp. P.-v. Réun. Cons. int. Explor. Mer 189*, 65-72.

[19] Kimura, K. (1929). On the detection of fish-groups by an acoustical method. *J. Imp. Fish. Inst., Tokyo, 24*, 41-45.

[20] Kotz, S., and Johnson, N.L. (1983). *Encyclopedia of Statistical Sciences*, Vol. 4. John Wiley and Sons.

[21] Kotz, S., and Johnson, N.L. (1982). *Encyclopedia of Statistical Sciences*, Vol. 1. John Wiley and Sons.

[22] Kotz, S., and Johnson, N.L. (1988). *Encyclopedia of Statistical Sciences*, Vol. 8. John Wiley and Sons.

[23] Levy, D. A., Nealson, P. A., and Cheng, P. (1991). Fixed-aspect hydroacoustic estimation of Fraser River sockeye salmon abundance and distribution at Mission, B.C., in 1986. *Pacific Salmon Comm. Tech. Rep. No. 4*, 30 p.

[24] MacLennan, D. N., and Simmonds, E. J. (1992). Fisheries acoustics. Chapman and Hall, London. 325 p.

[25] Madansky, A. (1965). Approximate confidence limits for the reliability of series and parallel systems. *Technometrics 7*, 495-503.

[26] Marsden, J.E., and Tromba, A.J. (1988). *Vector Calculus*, W.H Freeman and Company, San Francisco.

[27] Nunnallee, E.P, JR., and Mathisen, O. A. (1983). An acoustic survey of Lake Wanatchee, Washington. Univ Wash. *Fish. Res. Inst. Circ.* 72-13:14p.

[28] Plante, N. (1990). ***Estimation de la taille d'une population animale à l'aide d'un modèle de capture-recapture avec stratification.*** M.Sc. thesis, Université Laval.

[29] Quinn, T.P., and terHart, B.A. (1987). Movements of adult sockeye salmon (*Oncorhynchus nerka*) in British Columbia coastal waters in relation to temperature and salinity stratification: ultrasonic telemetry results, p. 61-67. *In* H.D. Smith, L. Margolis, and C.C. Wood [ed.] Sockeye salmon (*Oncorhynchus nerka*) population biology and future management. Can. Spec. Publ. *Fish. Aquat. Sci.* 96.

[30] Schaefer, M. B. (1951). Estimation of size of animal populations by marking experiments. U.S. *Fish and Wildlife Service Fisheries Bulletin, 69*, 191-203.

[31] Searle, S. R. (1928). *Matrix algebra useful for statistics.* John Wiley and Sons.

[32] Seber, G. A. F. (1982). *Estimation of animal abundance and related parameters.* 2nd ed. Chapter 11. Charles Griffin and Company Ltd.

[33] Seber, G. A. F. (1977). *Linear Regression Analysis.* John Wiley and Sons.

[34] Sen, P. K., and Singer J. M. (1993). *Large Sample Methods in Statistics: An introduction with applications.* Chapman and Hall.

[35] Silvey, S.D. (1975). *Statistical Inference*, Chapman and hall.

[36] Skalski, J.R., Hoffmann, A., Ransom, B.H., and Steig, T.W. (1993). Fixed-location hydroacoustic monitoring designs for estimating fish passage using stratified random and systematic sampling. *Can. J. Fish. Aquat. Sci. 50*, 1208-1221.

[37] Southwood, T. (1978), *Ecological Methods with Parameter Reference to the Study of Insect Populations*, London: Chapman and Hall.

[38] Thorne, R.E., and Dawson, J.J. (1974). An acoustic estimate of the escapement of sockeye salmon (*Oncorhynchus nerka*) into Lake Washington in 1971. *J. Fish. Res. Board Can. 31*, 222-225.

[39] Thorne, R.E. (1988). An empirical evaluation of the duration-in-beam technique for hydroacoustic estimation. *Can. J. Fish. Aquat. Sci. 45*, 1244-1248.

[40] Thorne, R.E. (1983). *Hydroacoustics.* Ch. 12 *In* L.A. Nielson, and D.L. Johnson [ed.] *Fisheries Techniques.* American Fisheries Society.

[41] Wilks, S.S. (1938). The Large-sample Distribution of the Likelihood Ratio for Testing Composite Hypothesis. *Ann. Math. Stat., 9*, 60-62.

[42] Woodey, J.C. (1987). In-season management of Fraser River sockeye salmon (*Oncorhynchus nerka*): meeting multiple objectives, p.367-374. *In* H.D. Smith, L Margolis and C.C. Wood [ed.] Sockeye salmon (*Oncorhynchus nerka*) population biology and future management. *Can. Spec. Publ. Fish. Aquat. Sci.* 96.