Estimating the Abundance of Non-solitary Animals - the Effects of Contagion on Mark-recapture and Removal Methods

by

Peter K. Ott
B.Sc., University of Guelph, 1992.

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Mathematics and Statistics

We accept this thesis as conforming
to the required standard

[redacted]

Dr. W. J. Reed, Supervisor (Department of Mathematics & Statistics)

[redacted]

Dr. R. R. Davidson, Department Member (Department of Mathematics & Statistics)

[redacted]

Dr. M. Tsao, Department Member (Department of Mathematics & Statistics)

[redacted]

Dr. D. S. Eastman, Outside Member (Department of Biology)

[redacted]

Dr. C. J. Schwarz, External Examiner (Simon Fraser University)

University of Victoria

Supervisor: Dr. W. J. Reed (Department of Mathematics and Statistics)

Abstract

The removal method and mark-recapture experiment are two common techniques for estimating animal abundance based on capture data. In this thesis, the closed population estimation procedures for both models (under the assumption of constant catchability for all animals on all sampling occasions) are extended to allow for the possibility that the animals within the population reside in non-solitary groups. This is accomplished two different ways. The first is to construct a log likelihood based on a contagious distribution such as the Polya-Eggenberger or Dirichlet. Each model contains an unknown parameter meant to reflect the degree of contagion caused by animal clustering. The second is to derive quasi-likelihood estimating equations for an overdispersed model. The unknown overdispersion parameter is estimated using two different techniques, and these are used to adjust the confidence intervals of the other parameters. The new models are tested using both real data and the results of simulation (artificial populations of various degrees of clustering). The contagious distribution approach provided conservative estimates of population size, and likelihood profile-based confidence intervals were not always possible due to the shape of the likelihood surface. The quasi-likelihood technique yielded estimates very close to those from the classical model but had slightly wider confidence intervals, which were further broadened when the population was exhibiting significant overdispersion (attributable to clustering). These preliminary results reinforce the general belief that the effect of animal clustering has little effect on the classical estimate of population size but the standard confidence intervals are too narrow and should be wider.

Examiners:

[redacted]

Dr. W. Reed, Supervisor (Department of Mathematics & Statistics)

[redacted]

Dr. R. Davidson, Department Member (Department of Mathematics & Statistics)

[redacted]

Dr. M. Tsao, Department Member (Department of Mathematics & Statistics)

[redacted]

Dr. D. Eastman, Outside Member (Department of Biology)

[redacted]

Dr. C. Schwarz, External Examiner (Simon Fraser University)

Contents

List of Tables

List of Figures

Acknowledgement

Firstly, I would like to thank Dr. Bill Reed for his patience, guidance, encouragement and advice.

Further thanks go to Drs. Davidson, Eastman, Schwarz and Tsao for examining my thesis and providing suggestions and comments.

I would also like to thank Doug Steventon and Todd Mahon for letting me use their vole data.

Finally, special thanks go to my colleagues and friends from room D-252 who knew the right times to go for coffee.

Peter K. Ott

Chapter 1

Introduction

During the last 30 years, the use of statistical models in the estimation of animal abundance has received much attention among biologists, wildlife managers, and statisticians. In addition, many of these same models have been successfully applied in non-animal related disciplines. Estimation procedures generally fall into three broad classes: (1) those based on capture data which include methods such as mark-recapture, change-in-ratio, and removal, (2) models based on sampling units such as plots, strips or quadrats, and (3) distance methods. There is some overlap between these three classes but the main distinction is that the first class calls for the unique identification of certain individuals in the population, whereas the other classes require the researcher to *census* (i.e. count everything) within a certain area or note the relative location of a set of objects from a reference line or point.

This thesis will only be concerned with the first class of models and in particular, the mark-recapture and removal methods for closed populations. In fact, the removal method is actually a special case of the mark-recapture experiment if marked animals are considered as "removed". Rather than provide a comprehensive literature review of this very broad subject, readers completely unfamiliar with it are directed to works such as Seber (1982, 1986, 1992), Pollock (1991), Otis et. al. (1978), and White et. al. (1982).

Models based on capture data have undergone tremendous modifications in attempts to relax the assumptions that were required in their early development. For example, the mark-recapture experiment alone has scores of available estimators depending on which variables are considered known in advance, the manner in which marking/sampling is performed, whether the population being studied is closed or open, and so on. However, one particular problem that has received relatively little attention is how to incorporate the phenomenon of animal clustering into the estimation technique. This is because many of the above estimators are based on urn models where captured animals are considered to be equivalent to balls randomly drawn from an urn. This simple urn analogy obviously fails when the balls are 'sticky' and are no longer randomly distributed in the urn. This is especially disturbing since, quoting from Legendre and Fortin (1989, p.137): "In nature, living beings are distributed neither uniformly nor at random. Rather, they are aggregated on patches, or they form gradients or other kinds of spatial structures."

In the following chapters, we examine the influence of animal clumpiness, grouping, or aggregation on the classical closed population mark-recapture and removal models. To date, it is widely accepted that the presence of animal clustering has little effect on the value of these estimators but the usual standard errors will be underestimated (Arnason et. al. 1991). In other words, the estimate of population size will be misleadingly precise. We will explore this notion and also try to estimate the degree of clustering or grouping. Terms such as *overdispersion* and *heterogeneity* will be defined when appropriate to describe the effect of the clustering phenomenon on the estimators.

The main purpose of this thesis will be to examine the possibility of applying various methods for accommodating clustering or clumpiness in two different animal abundance estimation methods, and to derive the appropriate likelihoods, quasi-likelihoods, estimating equations, etc. The performance of the new methods will be based in analyzing single datasets, albeit ones in which the true parameter values are known (at least for the simulated data). Unfortunately, it is impossible to say anything definitive about accuracy or coverage probabilities of confidence intervals on the basis of a single dataset. To do this, one would need to artificially generate multiple samples from the same model with known parameter values and perform a full-scale Monte Carlo study. def'n of Monte Carlo

Chapter 2 introduces the removal model, and the mark-recapture method is dealt with in Chapter 3. It is not crucial for these chapters to be read in sequence (or together for that matter) although it is recommended. However, both chapters do rely heavily on material from the appendices so it will often be necessary for the reader to refer back to them, especially where indicated.

Chapter 2

The Removal Experiment

The removal (or depletion method) is a technique for estimating the size of a wild animal population. The idea is to capture a proportion of the population of interest on successive occasions and use the trend in declining catches to estimate the initial population size before any of the samples had been taken. The animals captured can be "removed" by either marking (i.e. marked animals are ignored for the remainder of the study), being put aside until the study is complete, or in some cases by killing.

2.1 Classical Approach

The common way to model the removal experiment, when effort is equal for all samples is to consider the successive catches as either multinomial (Moran 1951, Carle and Strub 1978) or binomial (Schnute 1983, Seber 1982) random variables. Both methods yield the same maximum likelihood estimates when the probability of capture is equal for all animals on each sampling occasion. The binomial likelihood of N and $\{p_i\}$ from the joint distribution of the $\{C_i\}$ is given by

$$\mathcal{L} = f(N, \{p_i\} \mid \{C_i\}) = \prod_{i=1}^{k} \binom{N - T_{i-1}}{C_i} p_i^{C_i} (1 - p_i)^{(N-Ti)} \tag{2.1}$$

where N = unknown population size

p_i = unknown probability of capture on ith sample for every animal remaining in the population

C_i = observed number of animals captured on ith sample, $C_0 = 0$

T_i = ith cumulative catch $= \sum_{j=1}^{i} C_j$, $T_0 = 0$

k = number of samples.

When the $\{p_i\}$ are assumed to be equal for all sampling periods, we can write the log likelihood (up to an additive constant) as

$$\begin{aligned} l &= \log f(N, p \mid \{C_i\}) + \mathit{constant} \\ &= \log N! + kN \log(1-p) - \log(N - T_k)! + T_k \log p - \log(1-p) \sum_{i=1}^{k} T_i \\ &- \sum_{i=1}^{k} C_i \log C_i + T_k \log T_k - \log T_k! \end{aligned} \tag{2.2}$$

The incorporation of the last three terms will obviously have no effect on the location of extrema but it does ensure that the function reaches a maximum value no greater than zero (Schnute 1983). The utility of this feature will become apparent in Section 2.3.

This model assumes that effort is fixed and equal for all samples, the population is closed (except for known removals), and animals are caught independently of one another during each sampling period. This last assumption will often be violated when the animal population being studied is not randomly distributed but divided into clusters or groups such as schools, herds, pairs, etc. We will take two approaches to address this grouping/clustering phenomenon. The first will be to specify a new likelihood using a contagious distribution such as the Pólya-Eggenberger (beta-binomial) or Dirichlet distribution (Appendix A). The second will be to use quasi-likelihood.

2.2 Contagious Distribution Approach

To reflect the lack of independence in the capture of animals one can employ a contagious distribution as a model for captures. Two such distributions are discussed in Appendix A. In both the Pólya-Eggenberger and Dirichlet models, the parameter ρ represents the degree of *contagion* present in the probability scheme. We shall use this feature along with the paradigm of maximum likelihood to fit the above models to the removal experiment. In other words, the degree of contagion will not be specified in advance but rather estimated from experimental data. We assume that the contagion effect exists in the capture of animals because they reside in clusters or groups. From this we can specify new likelihoods for the removal experiment. For example, the likelihood of obtaining the observed catches can be written as a product of Pólya-Eggenberger distributions. Namely,

$$\begin{aligned}\mathcal{L} &= f(N,\{p_i\},\rho \mid \{C_i\}) \\ &= \prod_{i=1}^{k} \binom{N-T_{i-1}}{C_i} \frac{\Gamma\left(\frac{1}{\rho}\right)}{\Gamma\left(N-T_{i-1}+\frac{1}{\rho}\right)} \frac{\Gamma\left(C_i+\frac{p_i}{\rho}\right)}{\Gamma\left(\frac{p_i}{\rho}\right)} \frac{\Gamma\left(N-T_i+\frac{1-p_i}{\rho}\right)}{\Gamma\left(\frac{1-p_i}{\rho}\right)}\end{aligned} \tag{2.3}$$

If the p_i are the same for each sampling occasion, this becomes

$$\begin{aligned}\mathcal{L} &= f(N,p,\rho \mid \{C_i\}) \\ &= \frac{N!}{(N-T_k)!} \frac{\Gamma^k\left(\frac{1}{\rho}\right)}{\Gamma^k\left(\frac{p}{\rho}\right)\Gamma^k\left(\frac{1-p}{\rho}\right)} \prod_{i=1}^{k} \frac{1}{C_i!} \frac{\Gamma\left(C_i+\frac{p}{\rho}\right)\Gamma\left(N-T_i+\frac{1-p}{\rho}\right)}{\Gamma\left(N-T_{i-1}+\frac{1}{\rho}\right)}\end{aligned} \tag{2.4}$$

with log likelihood,

$$\begin{aligned} l &= \log \mathcal{L} + \mathit{constant} \\ &= \log N! - \log(N-T_k)! + k\log\Gamma\left(\frac{1}{\rho}\right) - k\log\Gamma\left(\frac{p}{\rho}\right) - k\log\Gamma\left(\frac{1-p}{\rho}\right) \\ &+ \sum_{i=1}^{k}\log\Gamma\left(C_i+\frac{p}{\rho}\right) + \sum_{i=1}^{k}\log\Gamma\left(N-T_i+\frac{1-p}{\rho}\right) \\ &- \sum_{i=1}^{k}\log\Gamma\left(N-T_{i-1}+\frac{1}{\rho}\right) \end{aligned} \tag{2.5}$$

The limiting version (as the population size becomes very large) when the p_i are equal is Dirichlet. Namely,

$$\begin{aligned}\mathcal{L} &= f(N,p,\rho \mid \{C_i\}) \\ &= \frac{\Gamma^k\left(\frac{1}{\rho}\right)}{\Gamma^k\left(\frac{p}{\rho}\right)\Gamma^k\left(\frac{1-p}{\rho}\right)} \prod_{i=1}^{k} \left(\frac{C_i}{N-T_{i-1}}\right)^{\left(\frac{p}{\rho}-1\right)} \left(\frac{N-T_i}{N-T_{i-1}}\right)^{\left(\frac{1-p}{\rho}-1\right)}\end{aligned} \tag{2.6}$$

with log likelihood,

$$\begin{aligned} l &= \log \mathcal{L} + \mathit{constant} \\ &= k \log \Gamma\left(\frac{1}{\rho}\right) - k \log \Gamma\left(\frac{p}{\rho}\right) - k \log \Gamma\left(\frac{1-p}{\rho}\right) \\ &+ \frac{p}{\rho}\sum_{i=1}^{k} \log C_i - \left(\frac{1}{\rho} - 2\right) \sum_{i=1}^{k} \log\left(N - T_{i-1}\right) \\ &+ \left(\frac{1-p}{\rho} - 1\right) \sum_{i=1}^{k} \log\left(N - T_i\right) \end{aligned} \tag{2.7}$$

The log likelihoods (2.5) and (2.7) are quite complicated and obtaining closed form solutions for the maximum likelihood estimates $\hat{N}$, $\hat{p}$ and $\hat{\rho}$ is not possible because of the presence of the gamma functions. However, optimizing these functions numerically (Appendix D) over $N \in [T_k, \infty)$, $p \in [0, 1]$, and $\rho \in (0, \infty)$ is a straightforward procedure. Before examining the performance of these models, we first describe another method.

2.3 Using Quasi-likelihood

Rather than specifying an explicit model for contagion, another, usually simpler, approach is to use a quasi-likelihood function for an *overdispersed* model (Appendix B). To do this, only parametric specification of the first two moments of the process being modelled is required. This is an advantage since the contagion models used above were adopted more for mathematical convenience than for a literal belief in their correctness. We shall assume an overdispersed form of the above binomial model, namely one in which the variances of the catches $\{C_i\}$ are inflated by a fixed factor σ^2, but in which the expected catches are the same as the binomial model (see McCullagh and Nelder 1989, pp.124-126).

In cases where the underlying probability distribution is a member of the univariate *exponential family* (McCullagh and Nelder 1989) (i.e. binomial, hypergeometric, Poisson, etc.) **and** the observations are independent, the quasi-likelihood approach yields estimators identical to maximum likelihood (McCullagh and Nelder 1989, p.325, Seber and Wild 1989, p.46, Charnes et al. 1976, Wedderburn 1974). However, this result does **not** apply to our model because the underlying probability distribution (2.1) is no longer a simple exponential family member once we treat N as a parameter. Proofs which demonstrate the equivalence of maximum likelihood and quasi-likelihood estimates are based on the ability to ignore all constant functions of the data that do not involve the *probabilities* (the p_i from (2.2), for example) but which may include N; (see for example McCullagh and Nelder 1989, p.114). Because all terms involving N must be retained for our likelihood (2.2), the result does not apply to our model. Furthermore, since quasi-likelihood does not incorporate the effect of the higher order cumulants in the estimation process, the resulting estimates may differ from those obtained when the complete likelihood is known.

For most distributions, a quasi-likelihood function can be explicitly formulated by constructing a suitable line integral (McCullagh and Nelder 1989, p.333) and then integrating the quasi-score (Appendix B) function. The fact that N is a parameter in our model prevents us from calculating a quasi-likelihood function and we are forced to rely solely on the quasi-score function and the resulting estimating equations (Appendix B). However, little is lost since knowledge of the quasi-likelihood function usually only becomes important in certain hypothesis testing situations and is not required for parameter estimation (Seber and Wild 1989).

The removal experiment lends itself nicely to this approach because we can consider the experiment as k independent (quasi) binomial trials.

$$\text{Let} \quad \vec{y} = [C_1, C_2, \cdots, C_k] .$$

$$\text{The two dimensional unknown parameter vector is} \quad \vec{\theta} = [N, p] .$$

Therefore,

$$E(\vec{y}) = \vec{\mu} = p\,[(N - T_0), (N - T_1), \cdots, (N - T_{k-1})] .$$

In the absence of overdispersion, the variance-covariance matrix would be

$$Var\,(\vec{y}) = \mathbf{V}(\vec{\mu}) = p\,(1-p) \begin{bmatrix} (N - T_0) & & & 0 \\ & (N - T_1) & & \\ & & \ddots & \\ 0 & & & (N - T_{k-1}) \end{bmatrix}$$

and thus,

$$\mathbf{V}(\vec{\mu})^{-1} = \frac{1}{p\,(1-p)} \begin{bmatrix} \frac{1}{(N - T_0)} & & & 0 \\ & \frac{1}{(N - T_1)} & & \\ & & \ddots & \\ 0 & & & \frac{1}{(N - T_{k-1})} \end{bmatrix} .$$

The matrix $\mathbf{D}$ is constructed by taking the derivative of the mean vector $\vec{\mu}$ with respect to each element in the unknown parameter vector $\vec{\theta} = [N, p]$. Therefore,

$$\mathbf{D} = \left[\frac{\partial \mu_i}{\partial \theta_j}\right] = \begin{bmatrix} p & (N - T_0) \\ p & (N - T_1) \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ p & (N - T_{k-1}) \end{bmatrix}.$$

Assuming overdispersion, we replace the covariance matrix by $\sigma^2 \mathbf{V}(\vec{\mu})$ where σ^2 is an overdispersion parameter. Note how both Pólya-Eggenberger and Dirichlet distributions are of this form, with the overdispersion parameter σ^2 depending on the contagion parameter ρ (see Appendix A.1 and Appendix A.2). Therefore, the quasi-score function (see Appendix B) is

$$\vec{U}\left(\vec{\theta}\right) = \frac{\mathbf{D}'\,\mathbf{V}(\vec{\mu})^{-1}\,(\vec{y} - \vec{\mu})}{\sigma^2} = \frac{1}{\sigma^2\,(1-p)} \begin{bmatrix} \displaystyle\sum_{i=1}^{k} \frac{C_i}{N - T_{i-1}} - kp \\ \displaystyle\frac{1}{p}\sum_{i=1}^{k} C_i + \sum_{i=1}^{k} T_{i-1} - kN \end{bmatrix} \tag{2.8}$$

The quasi-likelihood estimates of N and p, denoted as $\breve{N}$ and $\breve{p}$, are obtained by numerically solving $\vec{U}\left(\vec{\theta}\right) = \vec{0}$. It can also be shown that

$$\left(\mathbf{D}'\,\mathbf{V}(\vec{\mu})^{-1}\,\mathbf{D}\right)^{-1} = \frac{(1-p)}{\left|\mathbf{D}'\,\mathbf{V}(\vec{\mu})^{-1}\,\mathbf{D}\right|} \begin{bmatrix} \left(\dfrac{N}{p} - \dfrac{1}{k}\displaystyle\sum_{i=1}^{k} T_{i-1}\right) & -1 \\ -1 & \dfrac{p}{k}\displaystyle\sum_{i=1}^{k} \frac{1}{N - T_{i-1}} \end{bmatrix}$$

where

$$\left|\mathbf{D}'\,\mathbf{V}(\vec{\mu})^{-1}\,\mathbf{D}\right| = \left(N - \frac{1}{k}\sum_{i=1}^{k} T_{i-1}\right)\left(\sum_{i=1}^{k} \frac{1}{N - T_{i-1}}\right) - k.$$

Furthermore, since

$$Var\left(\hat{\vec{\theta}}\right) \approx \sigma^2 \left(\mathbf{D}'\,\mathbf{V}(\vec{\mu})^{-1}\,\mathbf{D}\right)^{-1},$$

given σ^2, we could estimate the approximate standard errors of our quasi-likelihood estimates. However, σ^2 is unknown and has to be estimated. We have two options to estimate it. The first way is the conventional Pearson estimate based on the residual vector $\vec{y} - \vec{\hat{\mu}}$. It is given (Appendix B) by

$$\breve{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{k} \frac{(y_i - \hat{\mu}_i)^2}{\hat{V}_i(\hat{\mu}_i)} = \frac{1}{k-2}\sum_{i=1}^{k} \frac{\left(C_i - \hat{p}(\hat{N} - T_{i-1})\right)^2}{\hat{p}(1-\hat{p})(\hat{N} - T_{i-1})} \tag{2.9}$$

The other approach is based on the normalized residual deviance $D^*\left(\vec{y} \mid \vec{\hat{\mu}}\right)$ which is usually obtained from the quasi-likelihood function. Since we do not have a quasi-likelihood function to use, we shall utilize the classical log likelihood (2.2) as a crude approximation. It is given by

$$\hat{\sigma}^2 = \frac{D^*\left(\vec{y} \mid \vec{\hat{\mu}}\right)}{n-p} \approx \frac{-2\{l(\vec{\hat{\theta}} \mid \vec{y}) - l(\vec{y} \mid \vec{y})\}}{n-p} \tag{2.10}$$

The vector $\vec{\hat{\theta}}$ contains the maximum likelihood estimates of N and p. Recall that $l(\vec{y} \mid \vec{y})$ represents the log likelihood for a *saturated* model; essentially it is the largest value that a log likelihood can attain. For our model (2.2), this value has been shown to be *zero* (Schnute 1983). Therefore, (2.10) becomes

$$\hat{\sigma}^2 \approx \frac{2(-l(\vec{\hat{\theta}} \mid \vec{y}))}{k-2} \tag{2.11}$$

To evaluate and compare the performance of the quasi-likelihood technique with the contagious distribution models (2.5) and (2.7), and the classical model (2.2), we turn to some simulated data.

2.4 Simulation Results

The process of sampling from various clustered populations was simulated (Appendix C) by assuming that a population consists of a groups of size b. Animal clusters or groups (not individual animals) were assigned a particular capture probability $P(S)$, and members from these located clusters were caught with a high probability $P(C|S)$. A very low probability of capture $P(C|S')$ was given to animals whose associated group had not yet been found.

Different simulations for a six sample ($k = 6$) experiment were run on a population of consisting of 100 animals with capture probabilities pre-set at $P(S) = 0.2$, $P(C|S) = 0.8$, and $P(C|S') = 0.05$ (i.e. overall catchability $= p = 0.20$). In order to simulate various degrees of clustering, the number a and size b of the clusters were varied six different ways and the resultant data were then analyzed by the Pólya-Eggenberger model (2.5), the Dirichlet model (2.7), the quasi-score (2.8), and the classical product-binomial likelihood (2.2).

Two methods are customarily used for inference on an unknown parameter. The first relies on the observed (quasi) Fisher information (B.4), (B.10) and the second is based on likelihood profiles constructed from either the likelihood (B.5) or quasi-likelihood (B.11) ratio statistic. The computations required for both methods are discussed in Appendix D.

Because likelihood profile methods are invariant with respect to parameter transformation and are accepted as being more reliable than those based on observed

a	b	C_1	C_2	C_3	C_4	C_5	C_6	Total Catch
2	50	6	38	7	1	38	2	92
5	20	19	17	3	5	24	4	72
10	10	6	17	7	8	10	13	61
20	5	13	15	10	11	7	7	63
50	2	14	17	10	9	8	3	61
100	1	22	12	11	9	9	5	68

Table 2.1: Data generated by simulation for a population consisting of a groups of size b. The true values of N and p are 100 and 0.20, respectively.

information, especially for small samples (Schnute 1983), approximate confidence intervals for N from the classical model (2.2) and the contagious models (2.5) and (2.7) are based on the likelihood ratio statistic (B.5). Confidence Intervals for N resulting from the quasi-likelihood estimating equations (2.8) rely solely on the Wald statistic (B.10) because there is no explicit quasi-likelihood function.

Approximate standard errors based on (B.4) are also shown for the classical and contagious models. Approximate standard errors based on (B.10) (after being scaled by the estimate of overdispersion σ^2) are also presented for the quasi-likelihood model.

The data generated from the simulations is given in Table 2.1 and the results of the models are shown in Tables 2.2, 2.3, 2.4 and 2.5. The value of the negative log likelihood at its minimum $-l$ is provided as a coarse indicator of model fit.

a	b	$\hat{N}$	$S.\hat{E}.(\hat{N})$	$\hat{p}$	$-l$
2	50	304.80 $(136.02, \infty)$	26.61	0.06	53.76
5	20	126.32 $(85.98, 3668.25)$	9.83	0.13	19.40
10	10	+	+	+	+
20	5	100.39 $(71.91, 661.35)$	7.79	0.15	3.08
50	2	76.66 $(64.18, 129.97)$	4.52	0.22	3.67
100	1	84.47 $(71.70, 132.04)$	4.61	0.24	3.04

Table 2.2: Maximum Likelihood Estimates with approximate 95% Confidence Intervals underneath for classical model (2.2) using simulated data. The value of the negative log likelihood at its minimum is represented by $-l$. Confidence Intervals are based on the likelihood ratio statistic (B.5) and Standard Errors result from observed information (B.4).

Note:
* indicates estimator $\hat{N}$ is at the boundary of admissible values (i.e. $\hat{N}$ is equal to the total catch T_k)
# indicates result not applicable because of *
+ indicates parameter estimate $\hat{N}$ is entirely outside admissible range
– indicates result not available due to algorithm failure (i.e. no well defined extrema within parameter space)

a	b	$\hat{N}$	$S.\hat{E}.(\hat{N})$	$\hat{p}$	$\hat{\rho}$	$-l$
2	50	92.00* $(92, \infty)$	#	0.33	0.735	–201.13
5	20	72.00* $(72, \infty)$	#	0.32	0.350	–115.01
10	10	61.00* –	#	0.22	0.073	–57.33
20	5	63.72 –	1.16	0.29	0.011	–76.08
50	2	65.57 –	2.45	0.29	0.007	–81.86
100	1	68.00* –	#	0.33	0.000	–96.52

Table 2.3: Maximum Likelihood Estimates with approximate 95% Confidence Intervals underneath for Pólya-Eggenberger model (2.5) using simulated data. The value of the negative log likelihood at its minimum is represented by $-l$. Confidence Intervals are based on the likelihood ratio statistic (B.5) and Standard Errors result from observed information (B.4).

Note:	*	indicates estimator $\hat{N}$ is at the boundary of admissible values (i.e. $\hat{N}$ is equal to the total catch T_k)
	#	indicates result not applicable because of *
	+	indicates parameter estimate $\hat{N}$ is entirely outside admissible range
	–	indicates result not available due to algorithm failure (i.e. no well defined extrema within parameter space)

a	b	$\hat{N}$	$S.\hat{E}.\left(\hat{N}\right)$	$\hat{p}$	$\hat{\rho}$	$-l$
2	50	–	–	–	–	–
5	20	–	–	–	–	–
10	10	–	–	–	–	–
20	5	–	–	–	–	–
50	2	–	–	–	–	–
100	1	95.23 $(83.37, \infty)$	3.06	0.19	0.005	–27.12

Table 2.4: Maximum Likelihood Estimates with approximate 95% Confidence Intervals underneath for Dirichlet model (2.7) using simulated data. The value of the negative log likelihood at its minimum is represented by $-l$. Confidence Intervals are based on the likelihood ratio statistic (B.5) and Standard Errors result from observed information (B.4).

Note:	*	indicates estimator $\hat{N}$ is at the boundary of admissible values (i.e. $\hat{N}$ is equal to the total catch T_k)
	#	indicates result not applicable because of *
	+	indicates parameter estimate $\hat{N}$ is entirely outside admissible range
	–	indicates result not available due to algorithm failure (i.e. no well defined extrema within parameter space)

a	b	$\hat{N}$	$S.\hat{E}.\left(\hat{N}\right)$	$\hat{p}$	$\breve{\sigma}^2$	$\hat{\sigma}^2$
2	50	104.75		0.24		
		$(92, 191.50)$	240.56		29.553	
		$(92, 187.48)$	218.79			26.878
5	20	93.54		0.20		
		$(72, 182.34)$	138.20		9.300	
		$(72, 184.21)$	144.14			9.700
10	10	+	+	+	+	+
20	5	106.57		0.14		
		$(71.68, 141.46)$	9.41		0.279	
		$(63, 188.53)$	51.90			1.539
50	2	78.18		0.22		
		$(61, 96.00)$	7.73		0.722	
		$(61, 106.61)$	19.66			1.837
100	1	86.68		0.23		
		$(71.76, 101.60)$	4.80		0.398	
		$(68, 115.84)$	18.35			1.520

Table 2.5: Quasi-Likelihood Estimates with approximate 95% Confidence Intervals based on (2.8) using simulated data. Confidence Intervals are based on the value of σ^2 in that row. Both Standard Errors and Confidence Intervals result from observed quasi-information (B.10).

Note:	*	indicates estimator $\hat{N}$ is at the boundary of admissible values (i.e. $\hat{N}$ is equal to the total catch T_k)
	#	indicates result not applicable because of *
	+	indicates parameter estimate $\hat{N}$ is entirely outside admissible range
	–	indicates result not available due to algorithm failure (i.e. no well defined extrema within parameter space)

The results from the simulations show that the classical binomial approach performed poorly in the first three instances when the degree of clustering is high (Table 2.2). The first estimate of N is far too high and the third simply failed to converge because of the ascending catch structure that occurred for that simulation. Notice further that the approximate confidence intervals using likelihood profiles for instances with high clustering are too large to give the estimate much practical use. The small estimated standard errors using the observed Fisher information in these cases are misleading because of the asymmetry of the log likelihood. Not surprisingly however, the classical model (2.2) does provide reasonable parameter estimates in the situations of less extreme clustering. It is no accident that the value of the negative log likelihood at its minimum $-l$ generally declines with cluster size. This is because $-l$ has a direct connection to the overdispersion parameter given by (2.11).

The Pólya-Eggenberger likelihood appears to pick up the presence of clustering correctly as indicated by the corresponding values for $\hat{\rho}$ (Table 2.3). Notice that the estimates of N are very conservative and 4 out of the 6 are actually the total catch T_k. Estimates of p are generally higher than those obtained from the classical likelihood because of the estimator's apparent affinity for values near to the total catch. This model's main drawback is its inability to provide stable confidence intervals (using likelihood profiles), especially when moderate to low clustering is being exhibited (i.e. ρ is very small). Estimated standard errors could only be calculated for the two cases when N was not lying directly on the boundary of admissible values.

Unfortunately, the Dirichlet likelihood performed very poorly (Table 2.4) and results from other simulations (not shown here) appear to indicate that its likelihood surface has well defined extrema only in cases of low contagion. Notice however, that for the sole case where a result was obtained, the estimate of N is very close to the classical model's result (which is most appropriate). It is also closer to the true value of N than the classical model although quite different from the exact Pólya-Eggenberger approach.

The quasi-likelihood approach seems the most promising of the methods, especially in situations of high clustering (Table 2.5). All estimates of N are very close to the true value of 100 and all but one of the confidence intervals covers 100. The estimates of the overdispersion parameter σ^2 decrease as clustering is reduced. However, $\breve{\sigma}^2$ drops a long way below 1, while $\hat{\sigma}^2$ stays above 1; this results in narrower confidence intervals when $\breve{\sigma}^2$ is employed. This suggests that $\hat{\sigma}^2$ may be a more appropriate estimator. However, further (Monte Carlo) study is required to resolve this issue.

The preliminary results from the simulations indicate that the quasi-likelihood model is the most promising and differs from the classical model in two ways. Firstly, it yielded an estimate very close to the true value of N for the case of extreme grouping. Secondly, for cases of less extreme clustering, parameter estimates were comparable but the estimated standard errors in the classical model based on the observed Fisher information were apparently underestimated (i.e. the precision of the classical model is less than that claimed). This is obvious because the quasi-likelihood standard errors are larger even before being multiplied by the estimate of overdispersion. It is unfortunate that we do not have an explicit quasi-likelihood function to construct profiles comparable to those shown for the classical model (Table 2.2). Since confidence intervals based on estimated standard errors are likely not very realistic (because of the asymmetry of the objective function), this is an area worth exploring in future work.

2.5 An Application to Real Data: Brook Char

The above models were also tested by using data previously published by Schnute (1983). Between 1974 to 1976, various sections of three New Brunswick brooks within the Nashwaak Experimental Watershed Project were electroshocked. Each study consisted of four sampling occasions and the data for brook char *Salvelinus fontinalis* was examined. For our purposes, the two studies with the highest cumulative catch were chosen to examine the new models and the data are presented in Table 2.6.

To results of the classical model (2.2), the two contagious models (2.5) and (2.7), and the quasi-likelihood technique, are shown in Tables 2.7, 2.8, 2.9, and 2.10 respectively. The value of the negative log likelihood at its minimum $-l$ is again provided as a coarse indicator of model fit.

Sample	C_1	C_2	C_3	C_4	Total Catch
1	52	31	10	12	105
2	88	22	11	2	123

Table 2.6: Catches of brook char from the Upper Narrows Mountain Brook in the Nashwaak Experimental Watersheds Project, 1974-76. These examples correspond to samples 2 and 4 investigated by Schnute (1983), which are the two studies with the highest cumulative catch.

Sample	$\hat{N}$	$S.\hat{E}.\left(\hat{N}\right)$	$\hat{p}$	$-l$
1	114.59 $(106.99, 131.52)$	3.33	0.46	3.92
2	123.41 $(123, 126.69)$	1.00	0.70	1.81

Table 2.7: Maximum Likelihood Estimates with approximate 95% Confidence Intervals underneath for classical model (2.2) using brook char data. The value of the negative log likelihood at its minimum is represented by $-l$. Confidence Intervals are based on the likelihood ratio statistic (B.5) and Standard Errors result from observed information (B.4).

Sample	$\hat{N}$	$\hat{S.E.}(\hat{N})$	$\hat{p}$	$\hat{\rho}$	$-l$
1	105.00* –	#	0.38	0.245	–235.82
2	123.00* $(123, 131.48)$	#	0.73	0.000	–365.45

Table 2.8: Maximum Likelihood Estimates with approximate 95% Confidence Intervals underneath for Pólya-Eggenberger model (2.5) using brook char data. The value of the negative log likelihood at its minimum is represented by $-l$. Confidence Intervals are based on the likelihood ratio statistic (B.5) and Standard Errors result from observed information (B.4).

Note: * indicates estimator $\hat{N}$ is at the boundary of admissible values (i.e. $\hat{N}$ is equal to the total catch T_k)

# indicates result not applicable because of *

– indicates result not available due to algorithm failure (i.e. no well defined extrema within parameter space)

Sample	$\hat{N}$	$S.\hat{E}.\left(\hat{N}\right)$	$\hat{p}$	$\hat{\rho}$	$-l$
1	123.50	4.96	0.37	0.026	–16.784
	$(109.71, \infty)$				
2	123.90	0.28	0.70	0.020	–16.016
	$(123.21, 125.41)$				

Table 2.9: Maximum Likelihood Estimates with approximate 95% Confidence Intervals underneath for Dirichlet model (2.7) using brook char data. The value of the negative log likelihood at its minimum is represented by $-l$. Confidence Intervals are based on the likelihood ratio statistic (B.5) and Standard Errors result from observed information (B.4).

Sample	$\hat{N}$	$S.\hat{E}.\left(\hat{N}\right)$	$\hat{p}$	$\breve{\sigma}^2$	$\hat{\sigma}^2$
1	114.48		0.46		
	$(94.64, 134.32)$	13.87		1.874	
	$(94.19, 134.76)$	14.50			1.959
2	123.97		0.70		
	$(123, 126.61)$	1.31		0.939	
	$(123, 126.56)$	1.26			0.907

Table 2.10: Quasi-Likelihood Estimates with approximate 95% Confidence Intervals based on (2.8) using brook char data. Confidence Intervals are based on the value of σ^2 in that row. Both Standard Errors and Confidence Intervals result from observed quasi-information (B.10).

The second sample of Table 2.6 shows a constant decline in catches which lends itself nicely to the classical model. Notice that the classical model (Table 2.7) gives very tight confidence intervals for sample 2 and this is mostly because $\hat{p}$ is so high. This model also performs well in sample 1 although the intervals are wider because the last catch is larger than the third.

The Pólya-Eggenberger results (Table 2.8) yield very conservative estimates of N as in the simulation study. The outcome from sample 2 is very close to the classical model because of the very small $\hat{\rho}$. Sample 1 conveys high contagion ($\hat{\rho} = 0.245$) and because $\hat{N}$ is relatively close to the classical model's result, $\hat{p}$ is lower (i.e. much of the catchability is being attributed to clustering). Unfortunately, no confidence intervals could be obtained from the first sample due to numerical complications (see Appendix D for details).

The Dirichlet model (Table 2.9) yielded estimates similar to the classical model ,because the degree of contagion was relatively low. Besides having a very wide confidence interval for sample 1, the results are quite consistent with the classical model.

The quasi-likelihood results again appear to be the most satisfactory (Table 2.10) because overdispersion is picked-up as expected and point estimates are very close to those provided by the classical model. Notice however, that if an approximate confidence interval is constructed for N using the estimated standard error for the case when overdispersion was identified (i.e. sample 1), then the interval using quasi-likelihood is considerably wider than that computed for the classical model. This confirms that in such cases, the precision of the classical method may be lower than that claimed.

To summarize, data collected from a highly clustered population are characterized by the *absence* of purely diminishing catches (as in the first three simulations of Table 2.1). Instead, large catches may be collected even during later samples (e.g. first row of Table 2.1) and it is for such cases that the classical model is suspect, with quasi-likelihood providing the alternative. Seber and Whale (1970) provided a crude failure criterion at which the classical model will no longer yield point estimates and standard errors with good properties. The criterion is meant to indicate to reseachers when it is necessary to continue sampling, based on previous catches. However, only the simulation of 10 groups of size 10 exceeded the criterion, indicating that it may be too conservative. This may be another area worth exploring with repeated simulations.

Chapter 3

The Mark-Recapture Experiment

The standard version of the mark-recapture or capture-recapture experiment consists of collecting a series of at least two samples from the population of interest and tagging or marking those animals captured on each sampling occasion with a distinct tag or mark. A record is kept of the animals caught with and without tags for each occasion so that the complete capture history for each captured animal is known by the end of the study. Although each of the many possible samples has an equal chance of being selected, the examination of captured animals (for marks) is performed without replacement. The basic assumptions are that the population is closed, no marks are lost, and all marked animals are correctly identified. Further restrictions on the catchability structure will be discussed as the upcoming models are formulated.

3.1 Classical Methods

There have been several standard probability models used to describe the mark-recapture experiment when a series of more than two samples are taken. The most prevalent form regards the number of animals having each of the various capture histories as resulting from N independent trials from a multinomial experiment. The joint probability function (Darroch 1958, Seber 1982) is then

$$f\left(\{x_\omega\} \mid N, \{P_\omega\}\right) = \frac{N!}{(N-r)! \prod\limits_{\omega} x_\omega!} \left(1 - \sum_{\omega} P_\omega\right)^{(N-r)} \prod_{\omega} P_\omega^{x_\omega} \tag{3.1}$$

where N = unknown population size

P_ω = unknown probability of having capture history ω

ω = capture history (a series of 1's and 0's)

x_ω = number of animals having capture history ω

r = total number of different animals caught $= \sum\limits_{\omega} x_\omega$.

When the probability of being caught is the same for all animals on a particular sampling occasion, then (3.1) (Darroch 1958) becomes

$$f\left(\{x_\omega\} \mid N, \{p_i\}\right) = \frac{N!}{(N-r)! \prod\limits_{\omega} x_\omega!} \prod_{i=1}^{k} p_i^{n_i} \left(1 - p_i\right)^{N-n_i} \tag{3.2}$$

where p_i = unknown probability of being caught on the ith sample

n_i = number of animals captured on the ith occasion

If the probability of capture is assumed to be constant for all animals on all sampling occasions, (3.2) simplifies further to

$$f\left(\{x_\omega\} \mid N, p\right) = \frac{N!}{(N-r)! \prod_\omega x_\omega!} p^{\left(\sum_{i=1}^{k} n_i\right)} (1-p)^{\left(kN - \sum_{i=1}^{k} n_i\right)} \tag{3.3}$$

It is important to note here that the Poisson analogue (Cormack 1989) to (3.1) is

$$f\left(\{x_\omega\} \mid N, \{P_\omega\}\right) = \prod_\omega \frac{e^{-NP_\omega}(NP_\omega)^{x_\omega}}{x_\omega!} \tag{3.4}$$

This approximation should be good for large N and small P_ω. The utility of the Poisson approximation as an alternative to the multinomial approach will become apparent later. It should also be noted that both the multinomial and the Poisson likelihoods can be maximized either *unconditionally* or *conditionally* (on the P_ω) (Bishop et. al. 1975). This is due to the fact that estimation of the nuisance parameters (the P_ω) can be performed separately from N. For all our cases however, we shall proceed with the standard unconditional approach.

A slightly different approach from (3.1) is to first concentrate on the number of animals captured with marks on each sample $\{m_i\}$. In this case the joint probability mass function is a product of hypergeometric probability mass functions, each conditioned on the number of animals in the population bearing marks, when each sample was taken (Chapman 1952). The sample sizes $\{n_i\}$ may be treated as independent binomial random variables, in which case the two distributions can be multiplied together to form the joint probability density function for both the $\{m_i\}$ and the $\{n_i\}$

(Darroch 1958, Seber 1982), which yields the same maximum likelihood estimates as the multinomial approach shown by (3.2).

Our goal is to find a model that can be modified to properly mimic the situation when the animals are spatially clustered. The hypergeometric-binomial model illustrates why it is more difficult to incorporate contagion in the mark-recapture experiment than the removal situation. Contagion can enter into two different probability processes: the probability that an animal is captured, **and** the probability that an animal already has marks, given that it has been captured. In general, these two contagions in probability are not the same. The most straightforward approach, albeit rather crude, will be to work with the multinomial based model (3.3) rather than the hypergeometric-binomial because of its simplicity. We shall reconstruct it as a contagious distribution (i.e. the Pólya-Eggenberger and/or the Dirichlet) and also utilize quasi-likelihood.

3.2 Contagious Distribution Approach

One way to address the phenomenon of animal clustering in the mark-recapture experiment is to generalize the multinomial probability density for the $\{x_\omega\}$ described by (3.3) by employing the contagion models of Appendix A. These models are applicable because they incorporate a contagion parameter ρ to address the effect of overdispersion. In the multinomial probability density for the $\{x_\omega\}$, overdispersion results when one animal having a particular capture history ω increases the probability of another having that same history. Since it seems natural for animals to end up with the same (or at least similar) capture histories if they reside in groups, converting (3.3) into a multivariate contagious distribution seems appropriate.

This is quite different from the approach taken by Otis et. al. (1978), Burnham (1972), and Burnham and Overton (1978, 1979) where it is assumed that the capture *frequencies* (i.e. the number of animals captured i times) are multinomial and the animals' catchabilities (i.e. p_l, $l = 1, 2, \ldots, N$) result from an unknown distribution. Burnham (1972) assumed a parametric form for the distribution of these p (the class of beta distributions) and found the approach to be unacceptable using maximum likelihood.

A more robust nonparametric technique than the above is proposed by Otis et. al. (1978) and Burnham and Overton (1978, 1979) which uses a jackknife estimator. Chao (1989) expanded this approach further in order to deal with situations involving sparse data and then developed a more complicated approach (Chao et. al. 1992) which allows capture probabilities to vary by time **and** to be heterogeneous using the idea of *sample coverage.* Yip (1991) worked with a function of the capture frequencies (the number of animals captured *at least* i times) and then employed zero mean martingales to address the problem of heterogeneity among the capture probabilities.

Our immediate approach is different from all of the above because:

(1) We are not modelling capture frequencies, but the x_ω

(2) We are imposing a multivariate (Dirichlet) parametric form on the P_ω (see Appendix A.2) and not on the p as in Burnham (1972).

(3) We are trying to *estimate* the degree of heterogeneity or clustering along with N.

(4) Our model is **not** nonparametric.

Using the Pólya-Eggenberger distribution (Appendix A.2), we can write down the likelihood of obtaining the various $\{x_\omega\}$ as

$$\begin{aligned}\mathcal{L} &= f\left(N,\rho,\{P_\omega\} \mid \{x_\omega\}\right)\\ &= \frac{N!}{(N-r)!\prod_\omega x_\omega!}\frac{\Gamma\left(\frac{1}{\rho}\right)}{\Gamma\left(N+\frac{1}{\rho}\right)}\frac{\Gamma\left(N-r+\frac{1-\sum_\omega P_\omega}{\rho}\right)}{\Gamma\left(\frac{1-\sum_\omega P_\omega}{\rho}\right)}\\ &\quad\cdot\prod_\omega\frac{\Gamma\left(x_\omega+\frac{P_\omega}{\rho}\right)}{\Gamma\left(\frac{P_\omega}{\rho}\right)} \qquad (3.5)\end{aligned}$$

where as before N = unknown population size

P_ω = unknown probability of having capture history ω

ρ = unknown contagion parameter

ω = capture history (a series of 1's and 0's)

x_ω = number of animals having capture history ω

r = total number of different animals caught $= \sum_\omega x_\omega$.

Since P_ω is a function of animal catchability p, which we are assuming constant for all animals and all sampling periods, we can rewrite $\mathcal{L}$ in terms of p.

$$\begin{aligned}
\mathcal{L} &= f\left(N,\rho,p \mid \{x_\omega\}\right) \\
&= \frac{N!}{(N-r)!\prod_{\omega} x_\omega!} \frac{\Gamma\left(\frac{1}{\rho}\right)}{\Gamma\left(N+\frac{1}{\rho}\right)} \frac{\Gamma\left(N-r+\frac{(1-p)^k}{\rho}\right)}{\Gamma\left(\frac{(1-p)^k}{\rho}\right)} \\
&\quad \cdot \prod_{i=1}^{k} \left(\frac{1}{\Gamma^{\binom{k}{i}}\left(\frac{p^i(1-p)^{k-i}}{\rho}\right)} \prod_{\substack{\omega \supset \text{caught} \\ i \text{ times}}} \Gamma\left(x_\omega + \frac{p^i(1-p)^{k-i}}{\rho}\right) \right) \qquad (3.6)
\end{aligned}$$

Taking logarithms, we obtain

$$\begin{aligned}
l &= \log \mathcal{L} + \mathit{constant} \\
&= \log N! - \log(N-r)! + \log\Gamma\left(\frac{1}{\rho}\right) - \log\Gamma\left(N+\frac{1}{\rho}\right) \\
&+ \log\Gamma\left(N-r+\frac{(1-p)^k}{\rho}\right) - \log\Gamma\left(\frac{(1-p)^k}{\rho}\right) \\
&- \sum_{i=1}^{k}\binom{k}{i}\log\Gamma\left(\frac{p^i\,(1-p)^{k-i}}{\rho}\right) + \sum_{i=1}^{k}\sum_{\substack{\omega \supset \text{caught} \\ i \text{ times}}} \log\Gamma\left(x_\omega + \frac{p^i\,(1-p)^{k-i}}{\rho}\right) \qquad (3.7)
\end{aligned}$$

Because the Pólya urn scheme also has a limiting form, it is possible to create a likelihood knowing that the asymptotic (as N becomes large) distribution of the proportions $\{x_\omega/N\}$ of animals with the particular capture histories is Dirichlet (Appendix A.2). However, the Dirichlet likelihood turns out to have a major shortcoming: when any of the x_ω have entries of zero (a fairly likely case if the animals are spatially clustered), the log likelihood contains term involving the log of zero multiplied by unknown parameters. This difficulty could be side-stepped by artificially assigning a very small number in lieu of zero but investigations revealed that the resulting maximum likelihood estimates (especially $\hat{\rho}$) were found to be *very* sensitive to the value of this number. Also, the objective function experienced convergence problems, and for these reasons, the Dirichlet model was not considered to be useful. We discuss it no further and concentrate on other approaches.

Since the sufficient statistics for the log likelihood (3.7) are the x_ω, the experimenter must obtain explicit information from the study concerning capture histories rather than mere summary statistics such as capture frequencies, number of animals with marks per sample, etc. Although this likelihood may seem awkward, it is important to realize that it is easily computed numerically, which is all that is required for calculating maximum likelihood estimates of N, ρ, and p (Appendix D). Before discussing results based on log likelihood (3.7), we present another approach, quasi-likelihood.

3.3 Quasi-likelihood

So far, we've seen how a contagious distribution can be used to extend the classical model to include heterogeneity of the capture probabilities. An even simpler idea is to model the overdispersion caused by the heterogeneity using quasi-likelihood. This involves determining means and variances for the classical model, the inclusion of an overdispersion parameter reflecting the contagion effects, and the determination of a set of estimating equations.

When deciding to use a quasi-likelihood function as a parameter estimation tool, it is sometimes difficult to choose which format (or perhaps portion) of the collected data are most suitable. This difficulty becomes apparent when the multinomial approach to the mark-recapture experiment is taken because there are observations on $2^k - 1$ different capture histories ω plus a non-observable class representing the animals that are never captured. Because we are assuming constant catchability for all animals and all sampling occasions, we have the luxury of arranging the observed data in the following way: the various x_ω are grouped so that the first $\binom{k}{1}$ observations are the animals that are captured once, the next $\binom{k}{2}$ observations are the animals that are captured twice, and so on. Another arrangement could be for example, with the capture histories in the order that would result from a binary tree. Furthermore, the non-observable group has been entirely ignored giving our response vector only $2^k - 1$ elements. This is *not* for convenience but rather a necessity since the number in this group is unknown. We shall still exploit the fact that the response vector is multinomial but now the probabilities do not add to unity. The fact that N is *not* a fixed variable but an unknown parameter automatically excludes our model from any results pertaining to the equivalence of maximum likelihood and quasi-likelihood estimation procedures since we are dealing with a distribution no longer in the exponential family. Ignoring the non-observables gives us the ability to invert the quasi-multinomial variance-covariance matrix rather than resorting to a generalized inverse which would most certainly be required if the responses were truly multinomial (McCullagh and Nelder 1989).

$$\text{Let} \quad \vec{y} = \left[\{x_\omega\}_{\substack{\omega \supset \text{caught} \\ \text{once}}}, \{x_\omega\}_{\substack{\omega \supset \text{caught} \\ \text{twice}}}, \cdots, \{x_\omega\}_{\substack{\omega \supset \text{caught} \\ k \text{ times}}} \right]$$

and denote the unknown parameter vector by $\vec{\theta} = [N, p]$.

$$\begin{aligned} E(\vec{y}) &= \vec{\mu} = f\left(\vec{\theta}\right) \\ &= N \left[\{P_\omega\}_{\substack{\omega \supset \text{caught} \\ \text{once}}}, \{P_\omega\}_{\substack{\omega \supset \text{caught} \\ \text{twice}}}, \cdots, \{P_\omega\}_{\substack{\omega \supset \text{caught} \\ k \text{ times}}} \right] \\ &= N \left[\vec{P}_1, \vec{P}_2, \cdots, \vec{P}_k \right] \end{aligned} \tag{3.8}$$

where each $\vec{P}_i$ consists of $\{\, p^i(1-p)^{k-i} \,\}$ repeated $\binom{k}{i}$ times.

The $(2k-1) \times (2k-1)$ variance-covariance matrix in the absence of overdispersion can be partitioned into constituent block matrices of differing size. The matrices identified by the letter **A** represent the variance-covariance structure among groups of individuals captured *the same number of times* whereas the matrices labeled by **B** reflect the covariance between groups with *different capture frequencies.* The subscripts are used to denote the respective capture frequencies.

In other words,

$$\mathbf{A}_i \text{ is of order } \binom{k}{i} \text{ with elements } \begin{cases} p^i(1-p)^{k-i}\left(1 - p^i(1-p)^{k-i}\right), & i = j \\ -\left(p^i(1-p)^{k-i}\right)^2, & i \neq j \end{cases}$$

and

$\mathbf{B_{ij}}$ is $\binom{k}{i} \times \binom{k}{j}$ with elements $-\left(p^i(1-p)^{k-i}\right)\left(p^j(1-p)^{k-j}\right)$.

Therefore,

$$Var(\vec{y}) \doteq \mathbf{V}(\vec{\mu}) = N \left[\begin{array}{c|c|c|c} \mathbf{A_1} & \mathbf{B_{12}} & \dots & \mathbf{B_{1k}} \\ \hline \mathbf{B'_{12}} & \mathbf{A_2} & \dots & \mathbf{B_{2k}} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{B'_{1k}} & \mathbf{B'_{2k}} & \dots & \mathbf{A_k} \end{array}\right].$$

This has inverse

$$\mathbf{V}(\vec{\mu})^{-1} = \frac{1}{N(1-p)^k}\left[\,1\,\right] + \frac{1}{N}\left[\begin{array}{c|c|c|c} \mathbf{C_1} & \mathbf{0} & \dots & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{C_2} & \dots & \mathbf{0} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{0} & \mathbf{0} & \dots & \mathbf{C_k} \end{array}\right]$$

where

$\left[\,1\,\right]$ is a matrix of order $2^k - 1$ consisting entirely of **1**'s

and

$\mathbf{C_i}$ is a diagonal matrix of order $\binom{k}{i}$ with elements $\dfrac{1}{p^i(1-p)^{k-i}}$.

Assuming overdispersion, the covariance matrix is replaced by $\sigma^2\mathbf{V}(\vec{\mu})$ where σ^2 is an overdispersion parameter. This is analogous to the covariance matrices for the Pólya-Eggenberger and Dirichlet distributions (Appendix A.2) where overdispersion is provided by a function of the contagion parameter ρ.

Also, the matrix with elements consisting of the derivative of the mean vector $\vec{\mu}$ taken with respect to each element of the unknown parameter vector $\vec{\theta}$ is denoted by $\mathbf{D}$. It can be shown to be

$$\mathbf{D} = \left[\frac{\partial \mu_i}{\partial \theta_j}\right] = \left[\begin{array}{c|c} \vec{P}_1' & \vec{\Theta}_1' \\ \hline \vec{P}_2' & \vec{\Theta}_2' \\ \hline \vdots & \vdots \\ \hline \vec{P}_k' & \vec{\Theta}_k' \end{array}\right]$$

where

$\vec{P}_i'$ is a column vector of length $\binom{k}{i}$ with elements $p^i(1-p)^{k-i}$

and

$\vec{\Theta}_i'$ is a column vector of length $\binom{k}{i}$ with elements

$$\frac{\partial}{\partial p} N p^i (1-p)^{k-i} = N\left(i p^{i-1}(1-p)^{k-i} - (k-i) p^i (1-p)^{k-i-1}\right).$$

After some tedious algebra, the quasi-score function can be shown to be

$$\begin{aligned} \vec{U}\left(\vec{\theta}\right) &= \frac{\mathbf{D}'\,\mathbf{V}(\vec{\mu})^{-1}\,(\vec{y} - \vec{\mu})}{\sigma^2} \\ &= \frac{1}{\sigma^2}\left[\begin{array}{c} \dfrac{1}{N(1-p)^k}\left[r - N\left(1 - (1-p)^k\right)\right] \\ \\ \dfrac{1}{p(1-p)} \displaystyle\sum_{i=1}^{k}\left[i \sum_{\omega \supset i} x_\omega - iN\binom{k}{i} p^i (1-p)^{k-i}\right] \end{array}\right] \end{aligned}$$

$$= \frac{1}{\sigma^2}\begin{bmatrix} 1 - \dfrac{N-r}{N(1-p)^k} \\ \\ \dfrac{1}{p(1-p)}\left[\sum_{i=1}^{k} n_i - kNp\right] \end{bmatrix} \tag{3.9}$$

In other words, the sufficient statistics required for parameter estimation are now merely the n_i and r rather than the x_ω. We will still need the x_ω however, for estimating σ^2.

The quasi-likelihood estimates of N and p, denoted as $\check{N}$ and $\check{p}$, are obtained by numerically solving $\vec{U}\left(\vec{\theta}\right) = \vec{0}$. In this case, they are exactly equivalent to the maximum likelihood estimates derived from the Poisson likelihood (3.4) with catchability assumed constant for all animals and all sampling periods. It can also be shown that

$$\left(\mathbf{D}'\mathbf{V}(\vec{\mu})^{-1}\mathbf{D}\right)^{-1} = \frac{(1-p)}{\left|\mathbf{D}'\mathbf{V}(\vec{\mu})^{-1}\mathbf{D}\right|}\begin{bmatrix} \dfrac{N}{p} & -1 \\ \\ -1 & \dfrac{1-(1-p)^k}{kN(1-p)^{k-1}} \end{bmatrix} \tag{3.10}$$

where,

$$\left|\mathbf{D}'\mathbf{V}(\vec{\mu})^{-1}\mathbf{D}\right| = \frac{1-(1-p)^k}{p(1-p)^{k-1}} - k\,.$$

Furthermore, since

$$Var\left(\vec{\theta}\right) \approx \sigma^2\left(\mathbf{D}'\mathbf{V}(\vec{\mu})^{-1}\mathbf{D}\right)^{-1}$$

we can estimate the approximate standard errors of our quasi-likelihood estimates once we find an estimate for σ^2.

For *polytomous* data (i.e. more than two categories), both the conventional moment estimator $\breve{\sigma}^2$ and the residual deviance based estimator $\hat{\sigma}^2$ are appropriate, even in the presence of *sparse* data (McCullagh and Nelder 1989, p.129). The moment estimator of overdispersion $\breve{\sigma}^2$ is given by

$$\begin{aligned}
\breve{\sigma}^2 &= \frac{\left(\vec{y}-\vec{\hat{\mu}}\right)' \mathbf{V}(\vec{\hat{\mu}})^{-1}\left(\vec{y}-\vec{\hat{\mu}}\right)}{n-p} \\
&= \frac{1}{2^k-3}\left\{\frac{\left[r+(1-p)^k-1\right]\left[r-N\left(1-(1-p)^k\right)\right]}{N(1-p)^k}\right. \\
&+ \left.\sum_{i=1}^{k}\sum_{\substack{\omega \supset \text{ caught} \\ i \text{ times}}} \frac{\left(x_\omega - Np^i(1-p)^{k-i}\right)^2}{Np^i(1-p)^{k-i}}\right\} \\
&= \frac{1}{2^k-3}\left\{\sum_{\substack{\omega \supset \text{ caught} \\ i \text{ times}}} \frac{\left(x_\omega - Np^i(1-p)^{k-i}\right)^2}{Np^i(1-p)^{k-i}}\right\}. \qquad (3.11)
\end{aligned}$$

Notice that although $\breve{\sigma}^2$ incorporated the covariance structure of the observation vector, the resultant estimate is very simple. This is because the piece involving r in the above formula is equal to zero because of the quasi-score function (3.9).

Since the estimating equations resulting from the quasi-multinomial approach are identical to those obtained from the Poisson likelihood, it seems prudent to also pursue an overdispersed Poisson model. A Poisson model such as (3.4) is also characterized by independent observations so the covariance matrix will be diagonal and therefore easy to manipulate. This is beneficial since even poorly misspecified covariance matrices will yield consistent quasi-likelihood parameter estimates (Liang and Zeger 1986, Zeger and Liang 1986).

For example, as before

$$E(\vec{y}) = \vec{\mu} = N\left[\vec{P}_1,\, \vec{P}_2,\, \cdots,\, \vec{P}_k\right] \tag{3.12}$$

and in the absence of overdispersion

$$Var\,(\vec{y}) = \mathbf{V}(\vec{\mu}) = N\mathbf{diag}\{\,\vec{P}_1,\, \vec{P}_2,\, \cdots,\, \vec{P}_k\,\} \tag{3.13}$$

which is obviously different from the multinomial variance-covariance matrix.

This leads to the same estimating equations as before, namely (3.9). It also yields the same moment estimator of overdispersion σ^2. However, the quasi-information matrix is different and its inverse (which can be used to obtain approximate standard errors) is the following:

$$\left(\mathbf{D}'\,\mathbf{V}(\vec{\mu})^{-1}\,\mathbf{D}\right)^{-1} = \frac{1}{\left|\mathbf{D}'\,\mathbf{V}(\vec{\mu})^{-1}\,\mathbf{D}\right|}\begin{bmatrix} \dfrac{kN\left(1-p-kp(1-p)^k\right)}{p(1-p)^2} & -k(1-p)^{k-1} \\ & \\ -k(1-p)^{k-1} & \dfrac{\left(1-(1-p)^k\right)}{N} \end{bmatrix} \tag{3.14}$$

where

$$\left|\mathbf{D}'\,\mathbf{V}(\vec{\mu})^{-1}\,\mathbf{D}\right| = \left[\left(\frac{k\left(1-(1-p)^k\right)\left(1-p-kp(1-p)^k\right)}{p(1-p)^2}\right) - \left(k(1-p)^{k-1}\right)^2\right].$$

The major benefit of the quasi-Poisson model to the quasi-multinomial is that we are able to calculate an explicit quasi-likelihood function. Note from Appendix B that the log quasi-likelihood is defined as

$$Q(\vec{\mu} \mid \vec{y}) = \sum_{i=1}^{n} Q_i(\mu_i \mid y_i)$$

where

$$\mathcal{Q}_i(\mu_i \mid y_i) = \int_{y_i}^{\mu_i} \frac{y_i - t_i}{\sigma^2 V_i(t_i)} dt_i.$$

For the quasi-Poisson,

$$\begin{aligned}
\mathcal{Q}_i(\mu_i \mid y_i) = \mathcal{Q}_\omega(\mu_\omega \mid y_\omega) &= \int_{x_\omega}^{\mu_\omega} \frac{x_\omega - t_\omega}{\sigma^2 V_\omega(t_\omega)} dt_\omega \\
&= \frac{1}{\sigma^2}\left[x_\omega \log(\frac{\mu_\omega}{x_\omega}) - (\mu_\omega - x_\omega)\right]. \qquad (3.15)
\end{aligned}$$

Therefore,

$$\begin{aligned}
Q(\vec{\mu} \mid \vec{y}) &= \frac{1}{\sigma^2}\left[\sum_\omega x_\omega \log(\mu_\omega) - \sum_\omega x_\omega \log(x_\omega) - \sum_\omega \mu_\omega + \sum_\omega x_\omega\right] \\
&= \frac{1}{\sigma^2}\left[\sum_\omega x_\omega \log(NP_\omega) - \sum_\omega x_\omega \log(x_\omega) - \sum_\omega NP_\omega + \sum_\omega x_\omega\right] \qquad (3.16)
\end{aligned}$$

Recall that P_ω is a function of p representing the probability that an animal has capture history ω. Therefore, (3.16) can be simplified (with constant functions of the data removed) to

$$Q(\vec{\mu} \mid \vec{y}) = \frac{1}{\sigma^2}\left[r \log(N) - N(1-(1-p)^k) + \sum_{i=1}^{k} f_i \log\left(p^i(1-p)^{k-i}\right)\right] \qquad (3.17)$$

where f_i is the number of animals captured exactly i times.

This makes it possible to use the quasi-likelihood ratio statistic (B.11) to construct quasi-likelihood profiles for confidence intervals. It also leads to an alternative estimate of $\hat{\sigma}^2$ since we can compute the normalized residual deviance (B.14). Namely,

$$\hat{\sigma}^2 = \frac{1}{n-p} D^*\left(\vec{y} \mid \vec{\hat{\mu}}\right) \qquad (3.18)$$

$$\begin{aligned} &= \frac{1}{n-p}\left\{-2\left[Q(\vec{\mu} \mid \vec{y}) - Q(\vec{y} \mid \vec{y})\right]\right\} \\ &= \frac{1}{n-p}\left\{-2\sum_{\omega}\left[x_{\omega}\log(\frac{NP_{\omega}}{x_{\omega}}) - (NP_{\omega} - x_{\omega})\right]\right\} \end{aligned} \tag{3.19}$$

The estimate $\hat{\sigma}^2$ can be easily calculated using the loglinear model setting proposed by Cormack (1993) with any software having analysis of deviance capability.

To evaluate and compare the performance of the quasi-likelihood techniques (3.9) and (3.17) with the log likelihood (3.7) and the classical model (3.3), we shall now turn to some simulated data.

3.4 Simulation Results

Data was simulated (Appendix C) for a 5 sample experiment (k=5) on a population containing 100 animals. Animal clusters or groups (not individual animals) were assigned a particular capture probability $P(S)$, and members from these located clusters were caught with a large probability $P(C|S)$. A very small probability of capture $P(C|S')$ was given to animals whose associated group had not yet been found. In particular, capture probabilities were preset at

$$\begin{aligned} P(S) &= 0.25 \\ P(C|S) &= 0.85 \\ P(C|S') &= 0.05 \end{aligned}$$

giving a true catchability of 0.25.

As in the removal experiment, two methods were used for confidence intervals for the unknown parameter N. The first relies on the observed (B.4) or quasi-observed (B.10) Fisher information via the Wald statistic and the second approach is to use profiles based on the likelihood (B.5) or quasi-likelihood (B.11) ratio statistic. The computations required to do both of these are discussed in Appendix D.

Approximate confidence intervals based on likelihood profiles (B.5) are accepted as having better coverage than those based on the Wald statistic, especially for small samples (Arnason et. al. 1991). Therefore, profiles were used to construct approximate confidence intervals for N using both the classical model (3.3) and the Pólya-Eggenberger model (3.7). Quasi-likelihood profiles (B.11) were based on the quasi-Poisson based quasi-likelihood function (3.17) which will allow us to compare profile-based confidence intervals for all models (which was not possible for the removal experiment).

Estimated standard errors for the classical and Pólya-Eggenberger models were calculated using the observed Fisher information (B.4). Approximate standard errors for the quasi-likelihood model are based on quasi-observed Fisher information (B.10) which has been multipled by the estimate of overdispersion σ^2.

The outcomes from the simulations are given in Table 3.4 and the model results are shown in Tables 3.2, 3.3, and 3.4. The value of the negative objective function at its minimum is provided as a coarse indicator of model fit.

	2 schools of size 50										
$x_{\omega\supset 1}$	32	4	0	0	8						$f_1 = 44$
$x_{\omega\supset 2}$	4	1	3	2	0	26	0	1	0	2	$f_2 = 39$
$x_{\omega\supset 3}$	0	1	1	0	0	1	3	0	2	0	$f_3 = 8$
$x_{\omega\supset 4}$	0	0	0	0	0						$f_4 = 0$
$x_{\omega\supset 5}$	0										$f_5 = 0$
											$r = 91$

	5 schools of size 20										
$x_{\omega\supset 1}$	20	4	3	1	3						$f_1 = 31$
$x_{\omega\supset 2}$	0	0	13	2	13	1	0	0	0	2	$f_2 = 31$
$x_{\omega\supset 3}$	0	1	1	0	0	16	0	0	0	0	$f_3 = 18$
$x_{\omega\supset 4}$	1	1	0	0	0						$f_4 = 2$
$x_{\omega\supset 5}$	0										$f_5 = 0$
											$r = 82$

	10 schools of size 10										
$x_{\omega\supset 1}$	16	10	5	1	3						$f_1 = 35$
$x_{\omega\supset 2}$	2	2	0	2	0	9	0	1	8	7	$f_2 = 31$
$x_{\omega\supset 3}$	5	1	0	0	0	1	0	0	0	0	$f_3 = 7$
$x_{\omega\supset 4}$	0	0	0	0	0						$f_4 = 0$
$x_{\omega\supset 5}$	0										$f_5 = 0$
											$r = 73$

	20 schools of size 5										
$x_{\omega\supset 1}$	16	8	13	10	7						$f_1 = 54$
$x_{\omega\supset 2}$	3	1	3	6	0	5	0	0	0	0	$f_2 = 18$
$x_{\omega\supset 3}$	0	1	2	0	0	1	0	0	2	0	$f_3 = 6$
$x_{\omega\supset 4}$	1	0	1	0	0						$f_4 = 2$
$x_{\omega\supset 5}$	1										$f_5 = 1$
											$r = 81$

	50 schools of size 2										
$x_{\omega\supset 1}$	15	3	10	6	8						$f_1 = 42$
$x_{\omega\supset 2}$	4	2	3	3	6	3	0	3	1	1	$f_2 = 26$
$x_{\omega\supset 3}$	0	1	1	0	0	1	1	0	1	0	$f_3 = 5$
$x_{\omega\supset 4}$	0	0	0	1	0						$f_4 = 1$
$x_{\omega\supset 5}$	0										$f_5 = 0$
											$r = 74$

	100 schools of size 1										
$x_{\omega\supset 1}$	9	5	7	8	8						$f_1 = 37$
$x_{\omega\supset 2}$	4	2	1	4	4	2	2	6	1	1	$f_2 = 27$
$x_{\omega\supset 3}$	1	1	1	0	0	1	1	2	1	0	$f_3 = 8$
$x_{\omega\supset 4}$	0	0	0	1	3						$f_4 = 4$
$x_{\omega\supset 5}$	0										$f_5 = 0$
											$r = 76$

Table 3.1: Data generated by simulation for a population consisting of a groups of size b. The true values of N and p are 100 and 0.25, respectively. $x_{\omega\supset i}$ is comprised of the possible x_ω's (the number of animals having capture history ω) for capture histories that involved being caught i times. The sum of each row is the capture frequency f_i (the number of animals captured exactly i times). The sum of the f_i's is r, the total number of different animals captured after the sampling has been completed.

a	b	$\hat{N}$	$S.\hat{E}.\left(\hat{N}\right)$	$\hat{p}$	$-l$
2	50	121.11 (106.02, 143.79)	6.40	0.24	–53.63
5	20	94.52 (86.62, 106.47)	3.89	0.33	–18.29
10	10	96.12 (83.30, 116.30)	5.59	0.25	–26.07
20	5	118.41 (99.76, 147.86)	7.46	0.20	–49.57
50	2	104.66 (88.53, 130.46)	6.65	0.22	–35.39
100	1	93.87 (83.49, 109.88)	4.78	0.28	–21.75

Table 3.2: Maximum Likelihood Estimates with approximate 95% Confidence Intervals underneath for classical model (3.3) using simulated data. The value of the negative log likelihood at its minimum is represented by $-l$. Confidence Intervals are based on the likelihood ratio statistic (B.5) and Standard Errors result from observed information (B.4).

a	b	$\hat{N}$	$S.\hat{E}.\left(\hat{N}\right)$	$\hat{p}$	$\hat{\rho}$	$-l$
2	50	91.00* (91.00, 110.39)	#	0.36	0.119	–112.71
5	20	82.00* (82.00, 110.45)	#	0.40	0.117	–73.44
10	10	91.29 (73.01, 138.61)	10.53	0.25	0.039	–42.85
20	5	110.79 (86.34, 146.27)	8.68	0.23	0.007	–50.85
50	2	79.74 –	2.60	0.22	0.000	–4096.00
100	1	77.62 –	1.49	0.12	0.000	–3072.00

Table 3.3: Maximum Likelihood Estimates with approximate 95% Confidence Intervals underneath for Pólya-Eggenberger model (3.7) using simulated data. The value of the negative log likelihood at its minimum is represented by $-l$. Confidence Intervals are based on the likelihood ratio statistic (B.5) and Standard Errors result from observed information (B.4).

Note:	*	indicates estimator $\hat{N}$ is at the boundary of admissible values (i.e. $\hat{N}$ is equal to the total number of different animals captured r)
	#	indicates result not applicable because of *
	–	indicates result not available due to algorithm failure (i.e. no well defined extrema within parameter space)

a	b	$\hat{N}$	$S.\hat{E}.\left(\hat{N}\right)^m$	$S.\hat{E}.\left(\hat{N}\right)^P$	$\hat{p}$	$\breve{\sigma}^2$	$\hat{\sigma}^2$	$-Q(\vec{\hat{\mu}} \mid \vec{y})$
2	50	122.19			0.24			−52.94
		(91, 257.90)	68.29	135.18		9.291		
		(91, 222.36)	45.69	90.46			6.217	
5	20	95.35			0.33			−17.30
		(82, 183.66)	35.53	103.00		9.424		
		(82, 161.10)	22.51	65.26			5.971	
10	10	97.19			0.24			−25.37
		(73, 155.97)	20.81	42.28		3.257		
		(73, 152.18)	18.77	38.14			2.938	
20	5	119.68			0.20			−49.00
		(81, 183.54)	22.47	39.25		2.393		
		(82.05, 176.77)	18.88	32.98			2.011	
50	2	105.86			0.21			−34.79
		(80.31, 139.98)	8.38	15.29		1.029		
		(79.23, 141.93)	9.21	16.81			1.131	
100	1	94.82			0.28			−20.93
		(76, 123.00)	5.96	13.75		1.173		
		(76, 121.10)	4.91	11.32			0.966	

Table 3.4: Quasi-Likelihood Estimates with approximate 95% Confidence Intervals based on (3.9) using simulated data. Confidence intervals are based on the value of σ^2 for that row and are calculated using the the quasi-likelihood ratio statistic (B.10). $S.\hat{E}.\left(\hat{N}\right)^m$ is the estimated standard error derived from the quasi-multinomial model and $S.\hat{E}.\left(\hat{N}\right)^P$ is derived from the quasi-Poisson model. The value of the negative log quasi-likelihood at its minimum is represented by $-Q(\vec{\hat{\mu}} \mid \vec{y})$.

The classical model (Table 3.2) provided very reasonable estimates of N and p and only in the case of extreme clustering (2 schools of size 50) did the approximate confidence interval for N not encompass the true value of 100.

The results of the Pólya-Eggenberger model (Table 3.3) yielded much more conservative results and estimates of N in the first two cases were actually the total catch r. The estimate of ρ properly conveyed the degree of clustering in all cases, however (i.e. the value of $\hat{\rho}$ decreases as one reads down Table 3.3). Confidence intervals for N are broader than those from the classical model and numerical difficulties prohibited their calculation in the last two situations when $\hat{\rho}$ was very close to zero. The estimated standard errors for the last two simulations using the Pólya-Eggenberger model are clearly unrealistic.

The parameter estimates using quasi-likelihood (Table 3.4) are very similar to those obtained using the classical model. However, because quasi-likelihood profiles for N have been adjusted by $\breve{\sigma}^2$, the true value of 100 is covered on every simulation (unlike the classical method results). For all cases, the estimated confidence intervals for N using quasi-likelihood are broader than those obtained using the classical model, because of the overdispersion adjustment. It is interesting to note that the quasi-multinomial standard errors are similar to those of the classical model whereas the quasi-Poisson standard errors seem to be consistently larger.

Although the method of modelling clumpiness is somewhat crude (recall that our model is based on the premise that the probability of observing counts of animals with particular capture histories will be affected if the animals are clustered), the estimates of σ^2 accurately reflect the degree of overdispersion (i.e. the estimates generally decline as one reads down the rows of Table 3.4). Notice further that $\hat{\sigma}^2$ reflects the degree of overdispersion slightly better than $\breve{\sigma}^2$. Nonetheless, all estimates of N and σ^2 are reasonable, and so the quasi-likelihood model seems more promising than the Pólya-Eggenberger model because of its simplicity and lack of numerical problems.

3.5 An Application to Real Data: Red Backed Voles

The above models were also applied to real data that was kindly provided by Mr. Doug Steventon of the BC Forest Service and graduate student Todd Mahon. Small mammals were live-trapped in the Date Creek Silvicultural Systems Project within 2 ha grids each consisting of 49 trap stations (20m x 20m spacing). Data involving red-backed voles (*Clethrionomys gapperi*) captured within the two grids having the best capture success during a three night trapping session (October 5, 6, 7, 1994) is presented. The raw data are given in Table 3.5 and the estimation results are shown in Tables 3.6, 3.7 and 3.8. The value of the negative objective function at its minimum is provided as a coarse indicator of model fit.

x_ω	Grid 1	Grid 2
x_{100}	9	6
x_{010}	13	17
x_{001}	8	13
x_{110}	8	6
x_{101}	11	7
x_{011}	4	7
x_{111}	10	8
r	63	64

Table 3.5: Red-Backed Vole Capture Data from the Date Creek Silvicultural Systems Project. The number of animals having capture history ω is given by x_ω, and $r = \sum_\omega x_\omega$ indicates the total number of different animals captured.

Grid	$\hat{N}$	$\hat{S.E.}\left(\hat{N}\right)$	$\hat{p}$	$-l$
1	72.15 (65.41, 83.57)	3.34	0.49	–76.49
2	78.62 (69.11, 94.72)	4.32	0.42	–80.02

Table 3.6: Maximum Likelihood Estimates with approximate 95% Confidence Intervals underneath for classical model (3.3) using Red-Backed Vole data. The value of the negative log likelihood at its minimum is represented by $-l$. Confidence Intervals are based on the likelihood ratio statistic (B.5) and Standard Errors result from observed information (B.4).

Grid	$\hat{N}$	$S.\hat{E}.\left(\hat{N}\right)$	$\hat{p}$	$\hat{\rho}$	$-l$
1	71.64 $(64.80, 83.95)$	3.24	0.49	0.000	–76.49
2	77.62 $(66.51, 96.64)$	4.30	0.42	0.001	–80.02

Table 3.7: Maximum Likelihood Estimates with approximate 95% Confidence Intervals underneath for Pólya-Eggenberger model (3.7) using Red-Backed Vole data. The value of the negative log likelihood at its minimum is represented by $-l$. Confidence Intervals are based on the likelihood ratio statistic (B.5) and Standard Errors result from observed information (B.4).

Grid	$\hat{N}$	$S.\hat{E}.\left(\hat{N}\right)^m$	$S.\hat{E}.\left(\hat{N}\right)^P$	$\hat{p}$	$\breve{\sigma}^2$	$\hat{\sigma}^2$	$-Q(\vec{\mu} \mid \vec{y})$
1	73.06			0.48			-75.49
	$(63, 94.70)$	4.79	10.15		1.042		
	$(63, 95.81)$	5.24	11.09			1.139	
2	79.68			0.42			-79.20
	$(64, 110.32)$	9.40	16.02		1.459		
	$(64, 110.54)$	9.51	16.22			1.477	

Table 3.8: Quasi-Likelihood Estimates with approximate 95% Confidence Intervals based on (3.9) using Red-Backed Vole data. Confidence intervals are based on the value of σ^2 for that row and are calculated using the the quasi-likelihood ratio statistic (B.10). $S.\hat{E}.\left(\hat{N}\right)^m$ is the estimated standard error derived from the quasi-multinomial model and $S.\hat{E}.\left(\hat{N}\right)^P$ is derived from the quasi-Poisson model. The value of the negative log quasi-likelihood at its minimum is represented by $-Q(\vec{\mu} \mid \vec{y})$.

The resultant parameter estimates and approximate confidence intervals for all three models from the red-backed vole study are very similar. In fact, even the values of the negative log likelihood are identical (to the first two decimal places) for both the classical model (Table 3.6) and the Pólya-Eggenberger model (Table 3.7). This is because ρ is estimated to be so close to zero.

The quasi-Poisson estimated standard errors are about twice as large as the quasi-multinomial standard errors (Table 3.8) and this is consistent with the simulation results (Table 3.4). However, emphasis should be placed on the profile-based confidence intervals of Table 3.8 because of the asymmetry of the objective function.

There is slight indication of overdispersion in grid 2 as indicated by a value of (0.001) for $\hat{\rho}$ in Table 3.7 and (1.459) for $\breve{\sigma}^2$ in Table 3.8. Since the degree of clustering appears quite low, the classical model appears adequate for both cases. There is however, no apparent 'penalty' for using either of the alternatives.

Chapter 4

Discussion

The preliminary results from the simulations and the real datasets indicate that quasi-likelihood is a better approach than using the Pólya-Eggenberger distribution for both the removal and mark-recapture experiments. The Pólya-Eggenberger models apparently yield inaccurate estimates of population size N (on the conservative side) and the likelihoods suffered unsatisfactory operating characteristics (e.g. no well defined extrema within the parameter space), especially for very small values of the contagion parameter ρ, which often made estimation of parameters and approximate confidence intervals impossible. In the removal experiment, the Dirichlet likelihood seemed to only perform adequately for cases of moderate to low animal clustering and was not pursued as a model for the mark-recapture experiment because it is unsuitable for data with empty cells (i.e. capture histories having zero animals).

For both experiments, the quasi-likelihood estimated standard errors are different than those obtained using a complete likelihood. Confidence intervals based on the likelihood ratio statistic and Fisher information were easily adjusted by an estimate of overdispersion which was much simpler to derive than trying to estimate ρ simultaneously. This result is in agreement with accepted beliefs of the effect of heterogeneity or overdispersion on confidence intervals. In our case, we are attributing this heterogeneity to animal clustering.

For the mark-recapture experiment, quasi-likelihood estimating equations were developed based on a probability scheme that was quasi-multinomial or equivalently a product of quasi-Poissons. Parameter estimates for the quasi-multinomial model are exactly equivalent to the maximum likelihood estimates derived from the quasi-Poisson model, which are also the same as Poisson likelihood estimates.

The main distinguishing feature between the quasi-Poisson and the quasi-multinomial models is that we can calculate an explicit quasi-likelihood function for the Poisson model; this enables the construction of quasi-likelihood profiles and a residual deviance-based estimator for the overdispersion parameter σ^2. Cormack (1993) dismissed an explanation for σ^2 in his log linear model estimators saying on page 1191: "there are no structural replicates and no obvious clusters to be modelled." Although this may be true within the general log linear model setting (where the catchability structure can take on a variety of forms), for the case when the probability of capture is assumed to be constant for all animals and all sampling occasions, the residual deviance *does* indeed reflect the degree of heterogeneity/clustering. Our quasi-multinomial results have shown that the conventional moment estimator based on the residual vector also works. Minta and Mangel (1989) and Mace et. al. (1994) utilize a similar idea when they compare the Monte Carlo variance to the nominal binomial variance to estimate the degree of heterogeneity among the sighting probabilities in mark-resight studies.

For the removal study, two estimates of σ^2 were examined. The first was the conventional moment estimator and the second was based on the normalized residual deviance taken from an exact likelihood. At the present time, this latter technique appears to model overdispersion slightly better.

Is quasi-likelihood better than the classical approach for the mark-recapture and removal experiments? There is little difference between the two in terms of parameter estimation and the only 'penalty' for using quasi-likelihood in the absence of overdispersion is the possibility of slightly broader confidence intervals. The results of the simulations give evidence that when the cluster size of the population is greater than five percent of the total population size (e.g. $b > 5$ for a population consisting of 100 animals), the classical model's confidence intervals are inappropriate. It is for these cases that quasi-likelihood is recommended.

The effect of the sample sizes, population size, probability of capture, etc. on the properties of the estimators is presently uncertain and is an area for future research. For example, it may be possible that the poor performance of some of the methods will disappear with increased sample sizes.

However, it should be noted that the conclusions are based only in analyzing single datasets, albeit ones in which the true parameter values are known (at least for the simulated data). For example, on the basis of a single dataset it is impossible to say anything about coverage probabilities of confidence intervals. To do this, one needs to artificially generate multiple samples from the same model with known parameter values. The main purpose of this thesis has been to examine the possibility of applying various methods for accommodating contagion or clumpiness in animal abundance estimation, and to derive the appropriate likelihoods, quasi-likelihoods, estimating equations, etc. It remains an important objective of future research to investigate fully by Monte Carlo methods the performance of the 'classical' methods, and the proposed alternatives when contagion is present. At present, the most that can be concluded with any confidence is that the quasi-likelihood approach seems to offer the most promising alternative to existing methods.

Bibliography

[1] Arnason, N. A., Schwarz, C. J. and J. M. Gerrard. 1991. Estimating closed population size and number of marked animals from sighting data. J. Wildl. Manage. 55(4): 716-730.

[2] Bishop, Y. M. M., Fienberg, S. E. and P. W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice.* MIT Press. Cambridge. 557 pp.

[3] Burnham, K. P. 1972. *Estimation of Population Size in Multiple Capture Studies When Capture Probabilities Vary Among Animals.* Unpublished Ph.D. dissertation. Oregon State University, Department of Statistics. Corvallis.

[4] Burnham, K. P. and W. S. Overton. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. Biometrika 65: 625-633.

[5] Burnham, K. P. and W. S. Overton. 1979. Robust estimation of population size when capture probabilities vary among animals. Ecology 60: 927-936.

[6] Carle, F. L. and M. R. Strub. 1978. A new method for estimating population size with removal data. Biometrics 34: 621-630.

[7] Chao, A. 1989. Estimating population size for sparse data in capture-recapture experiments. Biometrics 45: 427-438.

[8] Chao, A., Lee, S-M. and S-L. Tang. 1992. Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. Biometrics 48: 201-216.

[9] Chapman, D. G. 1952. Inverse, multiple and sequential sample censuses. Biometrics 8: 286-306.

[10] Charnes, A., Frome, E. L. and P. L. Yu. 1976. The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. J. Am. Stat. Assoc. 71: 169-171.

[11] Cormack, R. M. 1989. Loglinear models for capture-recapture. Biometrics 45: 395-413.

[12] Cormack, R. M. 1993. Variance of mark-recapture estimates. Biometrics 49: 1188-1193.

[13] Darroch, J. N. 1958. The multiple-recapture census I: Estimation of a closed population. Biometrika 45: 343-359.

[14] Eggenberger, F. and G. Pólya. 1923. Über die statistik verketteter vorgänge. Zeitschrift für angewandte Mathematik und Mechanik 1: 279-289.

[15] Ishii, G. and R. Hayakawa. 1960. On the compound binomial distribution. Annals of the Institute of Statistical Mathematics Tokyo 12: 69-80. (Errata 12:208).

[16] Johnson, N. L. and S. Kotz. 1969. *Discrete Distributions.* Houghton Mifflin. Boston. 328 pp.

[17] Johnson, N. L. and S. Kotz. 1977. *Urn Models and Their Applications: An Approach to Discrete Probability Theory.* John Wiley & Sons. New York. 402 pp.

[18] Legendre, P. and M-J. Fortin. 1989. Spatial pattern and ecological analysis. Vegetatio 80: 107-138.

[19] Liang, K. Y. and S. L. Zeger. 1986. Longitudinal data analysis using generalized linear models. Biometrika 73: 13-22.

[20] Mace, R. D., Minta, S. C., Manley, T. L. and K. E. Aune. 1994. Estimating grizzly bear populations using camera sightings. Wildl. Soc. Bull. 22: 74-83.

[21] Minta, S. C. and M. Mangel. 1989. A simple population estimate based on simulation for capture-recapture and capture-resight data. Ecology 70(6): 1738-1751.

[22] McCullagh, P. and J. A. Nelder. 1989. *Generalized Linear Models.* 2nd ed. Chapman & Hall. London. 511 pp.

[23] Moran, P. A. P. 1951. A mathematical theory of animal trapping. Biometrika 38: 307-311.

[24] Otis, D. L., Burnham, K. P., White, G. C. and D. R. Anderson. 1978. Statistical Inference for capture-recapture data in closed animal populations. Wildl. Monogr. No. 62. 135 pp.

[25] Pollock, K. H. 1991. Modeling capture, recapture and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present and future. J. Amer. Statist. Assoc. 86: 225-238.

[26] Schnute, J. 1983. A new approach to estimating populations by the removal method. Can. J. Fish. Aquat. Sci. 40: 2153-2169.

[27] Seber, G. A. F. 1982. *The Estimation of Animal Abundance and Related Parameters.* 2nd ed. Charles W. Griffin. London. 654 pp.

[28] Seber, G. A. F. 1986. A review of estimating animal abundance. Biometrics 42: 267-292.

[29] Seber, G. A. F. 1992. A review of estimating animal abundance II. International Statistical Review 60(2): 129-166.

[30] Seber, G. A. F. and J. F. Whale. 1970. The removal method for two and three samples (Corr. V27 p.1104). Biometrics 26: 93-400.

[31] Seber, G. A. F. and C. J. Wild. 1989. *Nonlinear Regression.* John Wiley & Sons. New York. 768 pp.

[32] Sprott, D. A. 1981. Maximum likelihood applied to a capture-recapture model. Biometrics 37: 371-375.

[33] Wedderburn, R. W. M. 1974. Quasilikelihood functions, generalized linear models and the Gauss-Newton method. Biometrika 61: 439-447.

[34] White, G. C., Anderson, D. R., Burnham, K. P. and D. L. Otis. 1982. *Capture-recapture and Removal Methods for Sampling Closed Populations.* Los Alamos National Laboratory. LA 8787-NERP. Los Alamos. 235 pp.

[35] Yip, P. 1991. A method of inference for a capture-recapture experiment in discrete time with variable capture probabilities. Commun. Statist. - Stochastic Models 7(3): 343-362.

[36] Zeger, S. L. and K. Y. Liang. 1988. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42: 121-130.

Appendix A

Contagious Distributions

A.1 Univariate Distributions

The Pólya urn model (Eggenberger and Pólya 1923) considers an urn originally consisting of c_f balls of colour F and c_g balls of colour G. A ball is randomly removed from the urn and replaced with s additional balls of the same colour. After η such sequences the distribution of the random variable x, representing the number of times a ball of colour F is drawn, has the Pólya-Eggenberger distribution with probability mass function

$$P(X=x) \quad = \quad \binom{\eta}{x} \frac{\Gamma\left(\frac{c_f+c_g}{s}\right)}{\Gamma\left(\eta+\frac{c_f+c_g}{s}\right)} \frac{\Gamma\left(x+\frac{c_f}{s}\right)}{\Gamma\left(\frac{c_f}{s}\right)} \frac{\Gamma\left(\eta-x+\frac{c_g}{s}\right)}{\Gamma\left(\frac{c_g}{s}\right)} \tag{A.1}$$

$$x=0,1,\ldots,\eta \quad \text{and} \quad \Gamma(\alpha)=\int_0^\infty t^{\alpha-1}e^{-t}dt \quad .$$

When $s = 0$, this distribution becomes binomial and when $s = -1$, hypergeometric. This feature allows us to easily generalize binomial and hypergeometric distributions into this form.

The Pólya-Eggenberger distribution is also referred to as a Beta-Binomial distribution (Ishii and Hayakawa 1960) because it arises as the mixture or compound distribution, namely the marginal distribution of X when

$$X \,|\, \pi \;\sim\; Binomial(\eta, \pi) \quad \text{and} \quad \pi \;\sim\; Beta(\frac{c_f}{s}, \frac{c_g}{s}) \quad .$$

However, instead of working with the parameter s we define a new parameter $\rho = \frac{s}{c_f + c_g}$, indicating the number of additional balls added to the urn on each drawing, expressed as a fraction of the initial endowment of balls. After reparameterizing, (A.1) becomes

$$P(X = x) \;=\; \binom{\eta}{x} \frac{\Gamma\left(\frac{1}{\rho}\right)}{\Gamma\left(\eta + \frac{1}{\rho}\right)} \frac{\Gamma\left(x + \frac{\pi}{\rho}\right)}{\Gamma\left(\frac{\pi}{\rho}\right)} \frac{\Gamma\left(\eta - x + \frac{(1-\pi)}{\rho}\right)}{\Gamma\left(\frac{(1-\pi)}{\rho}\right)} \tag{A.2}$$

$$x = 0, 1, \ldots, \eta \quad \text{and} \quad \Gamma(\alpha) = \int_0^\infty t^{\alpha - 1} e^{-t} dt$$

where

$$\pi = \frac{c_f}{c_f + c_g} \quad \text{and} \quad \rho = \frac{s}{c_f + c_g} \quad .$$

The mean and variance of (A.2) can be shown to be

$$E(x) = \eta\pi \quad \text{and} \quad V(x) = \left[1 + (\eta - 1)\left(\frac{1}{1 + \rho^{-1}}\right)\right] \eta\pi(1 - \pi) \quad .$$

Notice that although the mean is unaffected by ρ, $V(x)$ becomes larger than the usual binomial variance when $\rho > 0$, being inflated by the factor in square brackets. In other words, a large value of ρ corresponds to high degree of *overdispersion*, as indeed one would expect from the Pólya urn model.

It can also be shown that the limiting distribution as the number of drawings η becomes very large for the *proportion* Y of balls remaining in the urn of colour F is beta (a special case of Dirichlet) (Johnson and Kotz 1977). Namely,

$$P(Y=y) \quad = \quad \frac{\Gamma\left(\dfrac{1}{\rho}\right)}{\Gamma\left(\dfrac{\pi}{\rho}\right)\Gamma\left(\dfrac{(1-\pi)}{\rho}\right)}\, y^{\left(\frac{\pi}{\rho}-1\right)}\,(1-y)^{\left(\frac{(1-\pi)}{\rho}-1\right)} \qquad \text{(A.3)}$$

$$0 \le y \le 1$$

with mean and variance

$$E(y) = \pi \quad \text{and} \quad V(y) = \left(\frac{1}{1+\rho^{-1}}\right)\pi(1-\pi) \quad .$$

A.2 Multivariate Distributions

The multivariate Pólya urn scheme assumes an urn initially contains c_j balls of colour C_j and at each drawing, the ball is returned with s additional balls of its same colour. The multivariate generalization of the Pólya-Eggenberger distribution (Johnson and Kotz 1977, p.194) describes the joint distribution of the number $x_1, x_2, \ldots, x_n$ of times balls of colour $C_1, C_2, \ldots, C_n$ are selected on the ηth drawing from the urn. The joint probability mass function is given by

$$P(\vec{X} = \{x_1, x_2, \ldots, x_n\}) \quad = \quad \frac{\eta!}{x_1!, x_2!, \ldots, x_n!} \frac{\Gamma\left(\sum_{j=1}^{n} \frac{c_j}{s}\right)}{\Gamma\left(\eta + \sum_{j=1}^{n} \frac{c_j}{s}\right)} \prod_{j=1}^{n} \frac{\Gamma\left(x_j + \frac{c_j}{s}\right)}{\Gamma\left(\frac{c_j}{s}\right)} \tag{A.4}$$

on the simplex: $x_1 + x_2 + \ldots + x_n = \eta \qquad j = 0, 1, \ldots, n \qquad 0 \le x_j \le \eta$

and $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

The multivariate Pólya-Eggenberger distribution can be derived as a compound distribution (Ishii and Hayakawa 1960), namely the marginal distribution of $\vec{X}$ when

$$\vec{X} \mid \{\pi_j\} \sim \textit{Multinomial}\,(\eta, \{\pi_j\}) \quad \text{and} \quad \{\pi_j\} \sim \textit{Dirichlet}(\left\{\frac{c_j}{s}\right\}) \quad .$$

Not surprisingly, the marginal distribution of each component of $\vec{X}$ is Beta-Binomial (Johnson and Kotz 1969).

We shall reparameterize our model by setting $\rho = \frac{s}{\sum_{j=1}^{n} c_j}$ and $\pi_j = \frac{c_j}{\sum_{j=1}^{n} c_j}$. Equation (A.4) then becomes

$$P(\vec{X} = \{x_1, x_2, \ldots, x_n\}) \quad = \quad \frac{\eta!}{x_1!, x_2!, \ldots, x_n!} \frac{\Gamma\left(\frac{1}{\rho}\right)}{\Gamma\left(\eta + \frac{1}{\rho}\right)} \prod_{j=1}^{n} \frac{\Gamma\left(x_j + \frac{\pi_j}{\rho}\right)}{\Gamma\left(\frac{\pi_j}{\rho}\right)} \tag{A.5}$$

on the simplex: $x_1 + x_2 + \ldots + x_n = \eta \qquad j = 0, 1, \ldots, n \qquad 0 \le x_j \le \eta$

and $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

The mean, variance, and covariance of (A.5) can be shown to be

$$E(x_j) = \eta\pi_j$$

$$Cov(x_j, x_k) = \begin{cases} \left[1 + (\eta - 1)\left(\frac{1}{1+\rho^{-1}}\right)\right] \eta\pi_j(1-\pi_j), & j = k \\ -\left[1 + (\eta - 1)\left(\frac{1}{1+\rho^{-1}}\right)\right] \eta\pi_j\pi_k, & j \neq k \end{cases} .$$

Notice that although ρ does not affect $E(x_j)$, it does change the variances and covariances from the nominal multinomial form by the multiplicative factor in square brackets. As mentioned in section (A.1), ρ reflects the degree of *overdispersion*.

It can also be shown that in the Pólya urn scheme, the limiting distribution as the number of drawings η becomes very large for the *proportion* $\vec{Y}$ of balls remaining in the urn of colour c_j is again Dirichlet (Johnson and Kotz 1977). Namely,

$$P\left(\vec{Y} = \vec{y}\right) = \frac{\Gamma\left(\frac{1}{\rho}\right)}{\prod_{j=1}^{n} \Gamma\left(\frac{\pi_j}{\rho}\right)} y_1^{\left(\frac{\pi_1}{\rho}-1\right)} y_2^{\left(\frac{\pi_2}{\rho}-1\right)} \cdots y_n^{\left(\frac{\pi_n}{\rho}-1\right)} \qquad \text{(A.6)}$$

on the simplex: $y_1 + y_2 + \ldots + y_n = 1 \qquad j = 0, 1, \ldots, n \qquad 0 \leq y_j \leq 1$

with mean, variance, and covariance

$$E(y_j) = \pi_j$$

$$Cov(y_j, y_k) = \begin{cases} \left(\frac{1}{1+\rho^{-1}}\right) \pi_j(1-\pi_j), & j = k \\ -\left(\frac{1}{1+\rho^{-1}}\right) \pi_j\pi_k, & j \neq k \end{cases} .$$

Appendix B

Quasi-likelihood Theory

Rather than specifying explicit likelihoods, another much simpler approach is to use a quasi-likelihood function for an *overdispersed* model. The term overdispersion describes the situation when the nominal variance of the observed data becomes inflated, due to clustering or other causes. The presence of overdispersion seems quite natural; therefore the ability to address the phenomenon has great utility. To use a quasi-likelihood function, only information about the first two moments of the process being modelled is required. Quasi-likelihood is based on the idea that although the potential presence of overdispersion may change the distribution of our data from its original form, it will not change associated expectations, and will affect variances (and covariances) only insofar as they are inflated by the same constant (but unknown) factor. The following section is loosely based on Wedderburn (1974) and chapter 9 of McCullagh and Nelder (1989).

Suppose that observed data are arranged into a response vector $\vec{y}$ containing n **independent** elements and $\vec{\theta}$ is an unknown parameter vector of dimension p to be estimated. Suppose also that

$$E(y_i) = \mu_i = g_i\left(\vec{\theta}\right) \quad \text{and} \quad Var(y_i) = V_i(\mu_i) = h_i\left(\vec{\theta}\right).$$

The variance of the y_i is written as a function of the μ_i; this is possible since the μ_i themselves are functions of the unknown parameters. If the form of the distribution for each y_i can be specified by a probability mass or density function f, then the log likelihood is given by

$$l\left(\vec{\theta} \mid \vec{y}\right) = \sum_{i=1}^{n} \log f\left(\vec{\theta} \mid y_i\right) \tag{B.1}$$

The score function or statistic is defined as

$$\vec{U}_l\left(\vec{\theta} \mid \vec{y}\right) = \frac{\partial l\left(\vec{\theta} \mid \vec{y}\right)}{\partial \vec{\theta}} \tag{B.2}$$

and has variance-covariance matrix

$$i_l\left(\vec{\theta}\right) = -E\left[\frac{\partial^2 l\left(\vec{\theta} \mid \vec{y}\right)}{\partial \vec{\theta}\, \partial \vec{\theta}'}\right] = \sum_{i=1}^{n} E\left[\frac{\partial \log f\left(\vec{\theta} \mid y_i\right)}{\partial \vec{\theta}}\right]^2, \tag{B.3}$$

which is known as the Fisher information matrix. Maximum likelihood estimates $\vec{\hat{\theta}}$ are obtained by solving $\vec{U}_l\left(\vec{\theta} \mid \vec{y}\right) = \vec{0}$ with respect to $\vec{\theta}$, and further statistical inference can be made by utilizing asymptotic results involving (B.2) and (B.3) under certain regularity conditions. For example, assuming that the model is correct, asymptotically

$$\left(\vec{\hat{\theta}} - \vec{\theta}\right) \sim N_p\left(\vec{0}, i_l(\vec{\theta})^{-1}\right) \tag{B.4}$$

and

$$-2\left[l\left(\hat{\vec{\theta}}\mid\vec{y}\right)-l\left(\vec{\theta}\mid\vec{y}\right)\right]\sim\chi^2_p\,. \tag{B.5}$$

Now let us consider an overdispersed form of the previous model. Assume still that $E(\vec{y})=\vec{\mu}=g\left(\vec{\theta}\right)$ but now suppose that the variances and covariances, still functions of $\vec{\mu}$, are inflated by an unknown overdispersion factor σ^2. In other words, the covariance matrix of $\vec{y}$ is $Var(\vec{y})=\sigma^2\,\mathbf{V}(\vec{\mu})=\sigma^2\,\mathbf{h}\left(\vec{\theta}\right)$, say.

The quasi-score function defined by

$$\vec{U}(\vec{\theta})=\vec{U}_Q\left(\vec{\theta}\mid\vec{y}\right)=\sum_{i=1}^{n}\frac{y_i-\mu_i}{\sigma^2V_i\left(\mu_i\right)} \tag{B.6}$$

has several similar properties to the likelihood score statistic (B.2) including

$$E\left(\vec{U}(\vec{\theta})\right)=\vec{0}.$$

Furthermore, when it exists, the so called log quasi-likelihood given by

$$Q(\vec{\mu}\mid\vec{y})=\sum_{i=1}^{n}\mathcal{Q}_i(\mu_i\mid y_i) \tag{B.7}$$

where

$$\mathcal{Q}_i(\mu_i\mid y_i)=\int_{y_i}^{\mu_i}\frac{y_i-t_i}{\sigma^2V_i\left(t_i\right)}dt_i \tag{B.8}$$

behaves in a similar fashion to the log likelihood $l\left(\vec{\theta}\mid\vec{y}\right)$ in (B.1). In fact, the quasi-score function (B.6) is actually just one specific example of a *generalized estimating function* which is characterized by having zero mean for all parameter values. Furthermore, (B.6) can be written in a more usable and compact form with quasi-likelihood estimates $\vec{\breve{\theta}}$ obtained by solving $\vec{U}\left(\vec{\theta}\right) = \vec{0}$. Namely,

$$\begin{aligned} \vec{U}(\vec{\theta}) &= \frac{\partial Q(\vec{\mu} \mid \vec{y})}{\partial \vec{\theta}} \\ &= \left(\frac{\partial \mu_i}{\partial \vec{\theta}'}\right)' \frac{\partial Q(\vec{\mu} \mid \vec{y})}{\partial \mu_i} \\ &= \frac{\mathbf{D}'\,\mathbf{V}(\vec{\mu})^{-1}\,(\vec{y} - \vec{\mu})}{\sigma^2} \end{aligned} \tag{B.9}$$

where $\mathbf{D}$ is an $n \times p$ matrix with elements consisting of the derivative of the mean vector $\vec{\mu}$ taken with respect to each element of the unknown parameter vector $\vec{\theta}$. It will be denoted as $\left[\frac{\partial \mu_i}{\partial \theta_j}\right]$.

The covariance matrix of $\vec{U}(\vec{\theta})$, denoted by $i_Q\left(\vec{\theta}\right)$ can be shown to be $\frac{\mathbf{D}'\,\mathbf{V}(\vec{\mu})^{-1}\,\mathbf{D}}{\sigma^2}$. Under suitable regularity conditions and assumptions, limiting results similar to the maximum likelihood case follow. Namely,

$$\left(\vec{\breve{\theta}} - \vec{\theta}\right) \sim N_p\left(\vec{0}, i_Q(\vec{\theta})^{-1}\right) \tag{B.10}$$

and

$$-2\frac{\left[Q\left(\vec{\hat{\mu}} \mid \vec{y}\right) - Q(\vec{\mu} \mid \vec{y})\right]}{\sigma^2} \sim \chi^2_p. \tag{B.11}$$

Notice that any inference on $\vec{\theta}$ will now involve σ^2. The estimation of σ^2 is performed independently of (B.9) and there are two standard methods available, both based on model lack-of-fit. The first way is a conventional moment estimator based on the residual vector $\vec{y} - \vec{\hat{\mu}}$ which is known as the Pearson estimate and here denoted as $\breve{\sigma}^2$. It is given by

$$\breve{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{k}\frac{(y_i - \hat{\mu}_i)^2}{V_i(\hat{\mu}_i)} \tag{B.12}$$

$$= \frac{\left(\vec{y}-\vec{\hat{\mu}}\right)' \mathbf{V}(\vec{\hat{\mu}})^{-1}\left(\vec{y}-\vec{\hat{\mu}}\right)}{n-p} \quad . \tag{B.13}$$

The other approach is based on the *scaled residual deviance* $D^*\left(\vec{y} \mid \vec{\hat{\mu}}\right)$ which can be obtained from the quasi-likelihood function, if available. Denoted as $\hat{\sigma}^2$, it is given by

$$\hat{\sigma}^2 = \frac{1}{n-p} D^*\left(\vec{y} \mid \vec{\hat{\mu}}\right) \tag{B.14}$$

where $D^*\left(\vec{y} \mid \vec{\hat{\mu}}\right)$ is the scaled residual deviance defined by

$$D^*\left(\vec{y} \mid \vec{\hat{\mu}}\right) = -2\left[Q(\vec{\hat{\mu}} \mid \vec{y}) - Q(\vec{y} \mid \vec{y})\right]. \tag{B.15}$$

The estimator (B.14) is known to perform poorly in *binary* situations when *sparse* data is used (McCullagh and Nelder 1989, p.128).

Once an estimate for the overdispersion parameter has been obtained, approximate confidence intervals for the model parameter(s) $\vec{\theta}$ are are available by exploiting (B.10) or (B.11). Estimates of σ^2 less than one are certainly valid but are often treated as unity since *underdispersion* is rarely contemplated.

When the components of $\vec{y}$ are not independent, $\mathbf{V}(\vec{\mu})^{-1}$ will no longer be diagonal but most of the previously described results remain valid. In particular, (B.9), (B.13), and (B.14) may be used for parameter estimation and (B.10) and (B.11) still hold. However, because explicitly obtaining the log quasi-likelihood (B.7) can become cumbersome in these situations, utilizing (B.11), (B.14), and related hypothesis testing tools may not be available.

Another possible situation is the case when multiple multivariate observations are collected. An approach that is consistent with the previous notation is to concatenate the n k-variate response vectors into a single vector of nk elements. In this case, $\mathbf{V}(\vec{\mu})^{-1}$ will be of order nk (and probably block diagonal). The general results remain unchanged but n in the denominator of (B.13) and (B.14) should be replaced by the corresponding rank of $\mathbf{V}(\vec{\mu})$.

For the quasi-score function (B.6) to be the gradient vector of a quasi-likelihood function, it is required that the derivative matrix of $\vec{U}(\vec{\theta})$ with respect to $\vec{\theta}$ be symmetric. In other words, for $v \neq w$,

$$\frac{\partial \vec{U}_v(\vec{\theta})}{\partial \vec{\theta}_w} = \frac{\partial \vec{U}_w(\vec{\theta})}{\partial \vec{\theta}_v}.$$

Satisfying this restriction is not always trivial and McCullagh and Nelder (1989) provide conditions to help meet this requirement which also facilitate the calculation of an explicit quasi-likelihood function (pp. 334-336). Nonetheless, as long as these derivative matrices are equal in expectation to $-i_Q\left(\vec{\theta}\right)$, it is not imperative for (B.6) to be the quasi-likelihood's gradient vector and we may still employ (B.10) to make statistical inference; for example, see example 9.3.3 of McCullagh and Nelder (1989).

A very useful generalization of quasi-likelihood arises when $Var(\vec{y})$ is not a function of $E(\vec{y})$ alone; for instance, see example 9.3.3 of McCullagh and Nelder 1989. Even though it becomes impossible to construct an explicit quasi-likelihood function as in (B.7), we may still use (B.9) to estimate the model parameters, (B.13) to estimate the degree of overdispersion, and (B.10) to make statistical inference. It is this generalization which we shall rely upon in order to tackle the problem of estimating animal abundance based on capture data in the presence of overdispersion.

In conclusion, quasi-likelihood is a convenient method for providing estimates and standard errors of the parameters of a model, which includes overdispersion but only requires specification of means and covariances, rather than of complete distributions.

Appendix C

Simulating a Clustered Population

If the animals are truly distributed in a clustered fashion, we would expect to find entire clusters (not individual animals) with a particular probability, and then catch members from these located clusters with a very high probability. Furthermore, there will always exist the remote possibility of catching an animal, even though its associated group has not been found. Data were simulated in this fashion for a population consisting of a clusters of size b. The following probabilities were prespecified and animals were successfully 'captured' when a randomly generated number between zero and one was less than the value of the prespecified capture probability.

1. $P(S)$ = the probability of locating a group.
2. $P(C|S)$ = the probability of catching a member of a group that has been located.
3. $P(C|S')$ = the probability of catching an animal even though its group has not been located.

Using the Law of Total Probability, catchability can easily be calculated for our artificial population as $p = P(C) = P(C \cap S) + P(C \cap S') = P(C|S)P(S) + P(C|S')P(S')$. In order to simulate various degrees of clustering, a and b were adjusted and the resultant data then analyzed by the various models. Unfortunately, this physical model of clustering does not give rise to a simple expression that reflects the increase in variance of the observed catches in either the removal or mark-recapture studies. This means that we are unable to calculate a true value of contagion or overdispersion for the simulations.

Appendix D

Computational Resources

All computations were performed on an HP-715 unix workstation running HP-UX 9.05. The computer was generally required in four different areas: maximum likelihood estimation, likelihood profile evaluation, simulation, and quasi-likelihood estimation.

D.1 Maximizing the Likelihood

Programs written in HP FORTRAN/9000 were used to find **all** maximum likelihood estimates. In each case, the objective function was optimized using the NAG (Numerical Algorithms Group, Fortran Library Mark15) routine, E04JAF. This routine minimizes the objective function using a *quasi-Newton* algorithm subject to fixed upper and lower bounds. The routine requires the user to specify bounds and starting values but function derivatives need not be provided. The algorithm stops when the relative change occurring between successive iterations is less than some small prescribed value. The loggamma function ($\log \Gamma(\cdot)$) was calculated using the NAG routine S14ABF, and factorials were dealt with by using the NAG gamma function S14AAF.

D.2 Obtaining Approximate Standard Errors and Confidence Intervals

D.2.1 Observed Information

Standard errors based on information (for example, Fisher information (B.3)) were estimated using Maple V (Release 3). The second derivative of the likelihood was taken with respect to the parameter (in this case, N) symbolically using the command **diff**. This negative of this function was then evaluated at the parameter estimate using the function **evalf**. Terms involving factorials were handled using Maple's Gamma functions **GAMMA** or **lnGAMMA**.

D.2.2 Likelihood Profile Evaluation

To calculate confidence intervals for N, two different techniques were employed. For the classical model(s), true likelihood profiles were available because the maximum likelihood estimate of the nuisance parameter (catchability p) was an explicit function of N. This meant that likelihood contour extremes in the direction of N were sitting on a curve that could be explicitly expressed; this curve is the p-isocline of the likelihood. The profiles were reconstructed using Maple V (Release 3) by plotting against N the height of the likelihood along this isocline. Plots of the profile were examined and the command **fsolve** was then used to find the value(s) where the likelihood intersected with the appropriate horizontal line ($\frac{\chi^2}{2} = 1.92$ units above the minimum value of the negative log-likelihood). However, for the likelihoods based on the contagious distributions, HP FORTRAN/9000 was required. At fixed values of N placed equidistant apart, the negative log-likelihood was minimized with respect to the remaining parameters using the NAG routine E04JAF and the value at this minimum retained. The NAG routine E01BAF was then employed in order to fit a cubic spline through this set of points. Since this spline represented an estimate of the likelihood profile, the NAG interpolation algorithm E02BBF was used to find the points were the appropriate horizontal line intersected it. Although this technique worked quite well (it was tested using the known likelihood profiles from the classical model), the nature (i.e. complexity in shape) of the contagious likelihoods often made confidence interval approximation impossible. This was especially common when the contagion parameter ρ was extremely close to zero. Figures D.1 and D.2 show typical examples of likelihood profiles. Notice that the profile for the removal model (Figure D.1) is far from quadratic. By employing likelihood profiles to obtain confidence sets, we avoid the task of trying to find the correct (parameter) transformation that justifies using the 'usual' normal approximation method based on the Fisher information matrix (i.e. the Wald statistic).

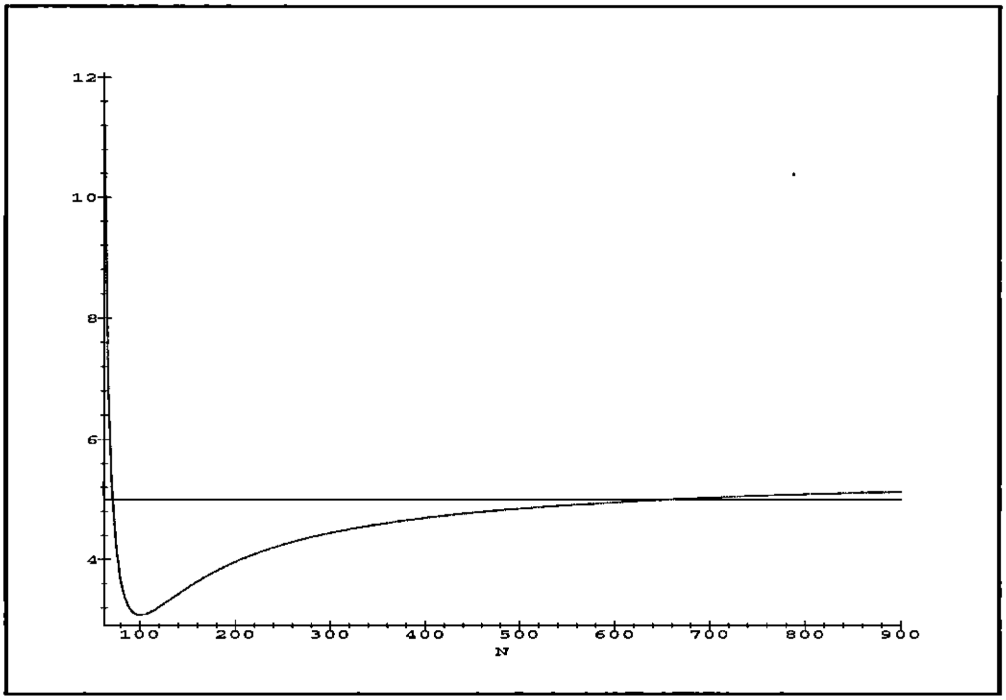

Figure D.1: Example likelihood profile for classical removal model (2.2) using simulated data representing 20 groups of size 5. The 95% confidence interval specifies the amount of change in $-l$ (1.92 units) above the minimum required to reject a value of N.

D.3 Simulation Programs

Both removal and mark-recapture experiments were simulated using programs written in HP FORTRAN/9000. Please refer to Chapter C of the Appendix for further details. However, the actual random number generator employed was taken from a linked HP C script. The HP C random number generator utilized was **drand48** with 'seed' provided by **seed48**. The function **drand48** generates a non-negative double-precision pseudo-random floating-point number that is uniformly distributed

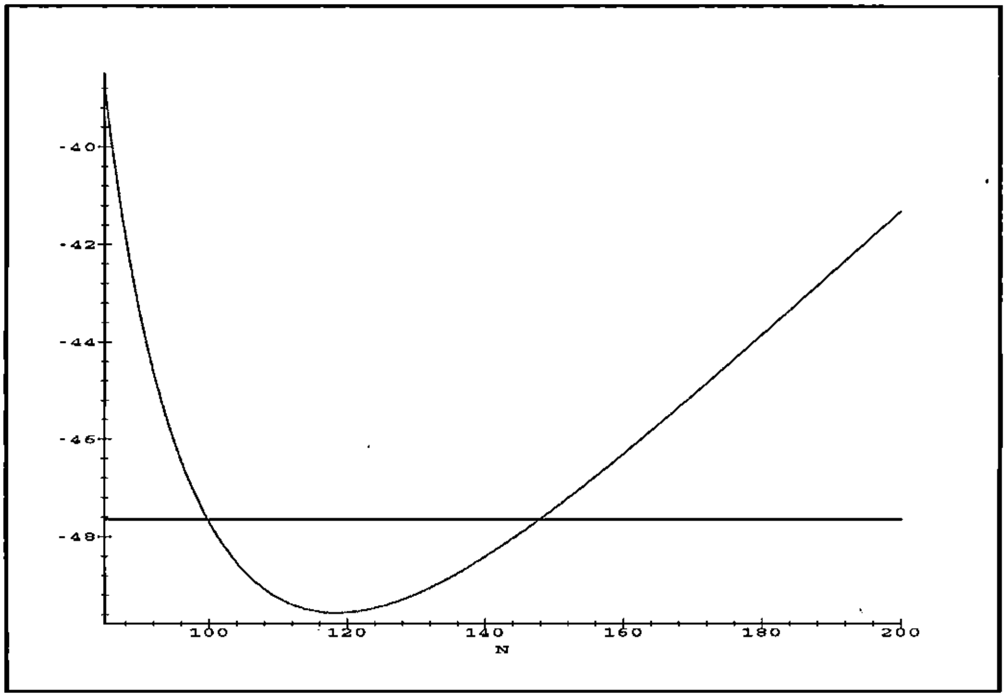

Figure D.2: Example likelihood profile for classical mark-recapture Model (3.1) using simulated data representing 20 groups of size 5. The 95% confidence interval specifies the amount of change in $-l$ (1.92 units) above the minimum required to reject a value of N.

between zero and one, using a linear congruential method. The main fortran program made use of the nonstandard intrinsic function SECNDS to initiate the random number generator. Pre-specified capture probabilities were compared with the randomly generated number and animals were considered to be 'captured' when this probability was greater than the random number. Samples were generated by first looping through the groups and "deciding" which groups were located by comparing the group-capture probability with a generated random number. Individual captures were then generated by looping through each group and using a high (low) probability of capture if the group had been located (missed). For the removal experiment, the size of the group was updated after every sample.

D.4 Quasi-Likelihood Estimation

All numerical work related to quasi-likelihood was performed using Maple V (Release 3). The command **fsolve** was used to solve the quasi-score function and calculations such as obtaining standard errors or the overdispersion parameter utilized symbolic formulae.

VITA

Surname: Ott

Given Names: Peter Kurt

Place of Birth: Canada

Date of Birth: October 10, 1969

Educational Institutions Attended:

University of Victoria:	1993 to 1996
University of Guelph:	1988 to 1993

Degree Awarded:

B.Sc (Honours)	University of Guelph	1992
Diploma	University of Guelph	1993

Honours and Awards:

University of Victoria Fellowship	1994-1995

Publications:

Beamish, F. W. H., Ott, P. K. and S. L. Roe. 1994. Interpopulational variation in fecundity and egg size in southern brook lamprey, *Ichthyomyzon gagei.* Copeia 3:718-25.

Krasowski, M. J., Hawkins, C. D. B., Coates, H. and P. K. Ott. 1996. Static tests of lodgepole pine stability in the central interior of British Columbia. Can.J.For.Res. In Press.

Krasowski, M. J., Letchford, T., Caputa, A., Bergerud, W. A. and P. K. Ott. 1996. The role of morphological properties of planting stock in susceptibility to freeze-desiccation injury. New Forests. In Press.

PARTIAL COPYRIGHT LICENSE

Title of Thesis/Dissertation: **ESTIMATING THE ABUNDANCE OF NON-SOLITARY ANIMALS - THE EFFECTS OF CONTAGION ON MARK-RECAPTURE AND REMOVAL METHODS**

Author [redacted]

PETER KURT OTT
(Name in Block Letters)

DEC. 6, 1996
(Date)