Partial Stratification in Two-Sample Capture-Recapture Experiments

Lasantha Premarathna

Department of Statistics and Actuarial Science,
Simon Fraser University,
Burnaby, BC.

May 24, 2014

Outline

Introduction

- **Capture-recapture Method**
  - Capture-recapture is a method commonly used in ecology to estimate an animal population size.
- **Lincoln-Petersen Method.**
  - A simple two-sample capture-recapture method
  - Lincoln-Petersen estimte for population abundance

$$\hat{N} = n_1 n_2 / m$$

  where n_1 is a sample of animals captured, marked and released in the first capture occasion and n_2 is the total number of animals captured in the second occasion and m is the number of marked animals captured in the second occasion.
- **Partial Stratification**
  - Animals may vary in capture probability due to sex,size or many other factors.
  - Heterogenity in catchability is know to lead bias in two-sample Lincoln-Petersen estimate of population size.
  - stratification adress the capture heterogenity.
  - Importance of partial stratification

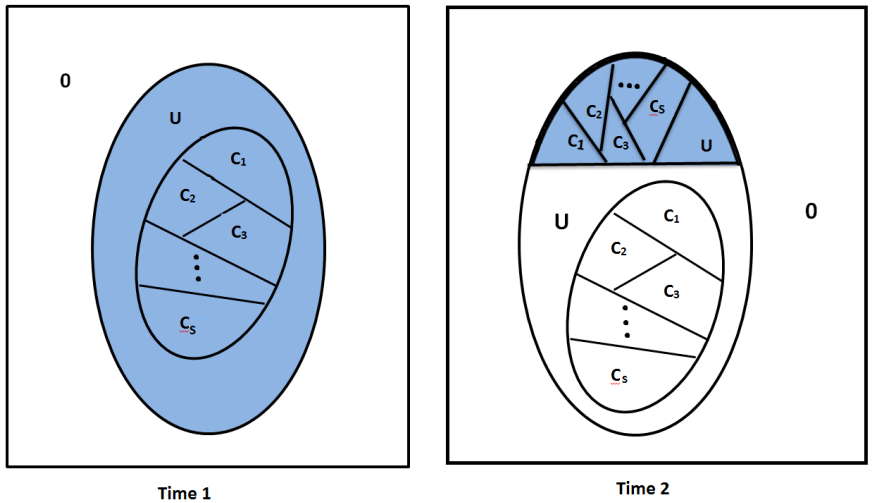
Model Development

Notation

Partial Stratification in Two-Sample Capture-Recapture Experiments

T	Number of capture occasions, where $T = 2$
t	index for capture occasion, where $t = 1, T$
s	Number of categories
k	index for category, where $k = 1, 2, ..., s$
0	animal is not captured
U	animal is captured but not stratified
C_k	animal is captured and identified the category k

Sampling protocol



Model Development

Assumptions

- The population is **closed** (geographically and demographically). The number of individuals does not change during the study through birth or immigration and/or death or emigration
- Mark status is correctly identified at recovery and unique to each animal
- Marks are not lost between sampling occasions
- Capture and marking does not affect subsequent catchability of an animal
- The population can be divided into non-overlapping categories
- Sub-sample at each occasion is a random sample of animals that are not marked
- categories of the animals in the sub-sample are successfully identified
- Animal captures are independent

Model Development

Capture Histories

$U0, UU, 0U, C_k0, C_kC_k, 0C_k$ and 00

- There are $3s + 4$ capture histories

- Capture history 00 is unobservable

Statistics

$n_{U0}, n_{UU}, n_{0U}, n_{C_k0}, n_{C_kC_k}, n_{0C_k}$ are the number of animals related to the *observable* capture histories and n is the number of animals caught in the study

$$n = n_{U0} + n_{UU} + n_{0U} + \sum_k n_{C_k0} + \sum_k n_{C_kC_k} + \sum_k 0C_k$$

Model Parameters

p_{tk}	Capture probability of animals belong to category k at sampling occasion t
λ_k	Proportion of category k animals in the population
θ_t	sub-sample proportion at sampling occasion t
N	Population size

$$\lambda_s = 1 - \sum_{k=1}^{s-1} \lambda_k$$

Model Development

Probability Statements of Capture History

$$
\begin{aligned}
P_{U0} &= \sum_{k}^{s} \lambda_k p_{1k}(1-\theta_1)(1-p_{2k}) \\
P_{UU} &= \sum_{k}^{s} \lambda_k p_{1k}(1-\theta_1)p_{2k} \\
P_{0U} &= \sum_{k}^{s} \lambda_k (1-p_{1k})p_{2k}(1-\theta_2) \\
P_{C_k 0} &= \lambda_k p_{1k}\theta_1(1-p_{2k}) \\
P_{C_k C_k} &= \lambda_k p_{1k}\theta_1 p_{2k} \\
P_{0C_k} &= \lambda_k (1-p_{1k})p_{2k}\theta_k \\
P_{00} &= \sum_{k}^{s} \lambda_k (1-p_{1k})(1-p_{2k})
\end{aligned}
$$

- $P_{U0} + P_{UU} + P_{0U} + \sum_k P_{C_k 0} + \sum_k P_{0C_k} + \sum_k P_{0C_k} + P_{00} = 1$
- Even though the capture history 00 is unobservable, the probability statement cab be explicitly given

Model Development

The Model

Multinomial distribution with unknown index

$$f(.) = \frac{N!}{n_{U0}! \quad n_{UU}! \quad n_{0U}! \quad \prod_{k} n_{C_k 0}! \quad \prod_{k=1} n_{C_k C_k}! \quad \prod_{k} n_{0 C_k}! \quad (N-n)!} \times$$
$$(P_{U0})^{n_{U0}} (P_{UU})^{n_{UU}} (P_{0U})^{n_{0U}} \times$$
$$\prod_{k=1}^{s} (P_{C_k 0})^{n_{C_k 0}} \times$$
$$\prod_{k=1}^{s} (P_{C_k C_k})^{n_{C_k C_k}} \times$$
$$\prod_{k=1}^{s} (P_{0 C_k})^{n_{0 C_k}} (P_{00})^{N-n}.$$

Parameter Estimation

Closed form solution

Numerical solution

- Standard numerical methods are used to maximize the likelihood

Model Constraints, Link Functions

- Parameters can be constrained by using the design matrix and offset vectors
- Used parameter index matrices (PIM) as implemented in MARK program to describe the three separate design matrix for the model parameters p_{tk}'s, λ_k's and θ_t's
- logit link functions were used to restrict parameter estimates. For the capture probabilities

$$logit(p_{tk}) = X\beta + offset$$

Where X is the corresponding design matrix

- **Example:** Consider a two-sample experiment with two categories(male (M) and female (F) and capture probabilities for two categories are equal and category proportions in the population are fixed(say proportion of males=0.4)

$p_{1M} = p_{1F}$, $p_{2M} = p_{2F}$, $\lambda_M = 0.4$ and $\lambda_F = 0.6$

  - Design matrix and offset for capture probabilities are

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

  - Design matrix for category proportions is a matrix with zero column and one row and offset vector is $\begin{bmatrix} 0.4 \end{bmatrix}$

Model Selection

- Best model is selected based on the Akaike's information criterion (AIC) (Burnham and Anderson, 2002)

$$AIC = -2ln(L) + 2n_p$$

and the Corrected AIC

$$AICc = AIC + \frac{2n_p(n_p + 1)}{n - n_p - 1}$$

Where,

n_p is the number of parameters to be estimated

Planning Experiments

Precision and Bias of Estimates

- Assessing the performances of methods often requires addressing the question about bias and precision about of estimates.
- Bias and the precision assessments based on the expected counts for each sample capture history (Devineau (2006)).

Optimal Allocation

- Fixed amount of funds available for the study (C_0)
  - The objective is to find out the optimal number of fish should be captured at the both occasions and the size of the sub samples to be stratified so that the variance of the estimated population size ($Var(\hat{N})$) is minimum.

Planning Experiments

Optimal Allocation

- Cost of the partial stratified two-sample capture recapture depend on

c_t = Cost to catch an animal at occasion t
c_t^* = Cost to stratify an animal at occasion t
n_t = Number of animals captured at occasion t
n_t^* = Number of animals stratified at occasion t

- The total cost(C) of the experiment can be considered as a linear function of

$$C = n_1 c_1 + n_1^* c_1^* + n_2 c_2 + n_2^* c_2^* \leq C_0$$

- Find the optimal allocation of n_t and n_t^* where $t = 1, 2$, such that minimise the $Var(\hat{N})$ with respect the linear constraint

Example

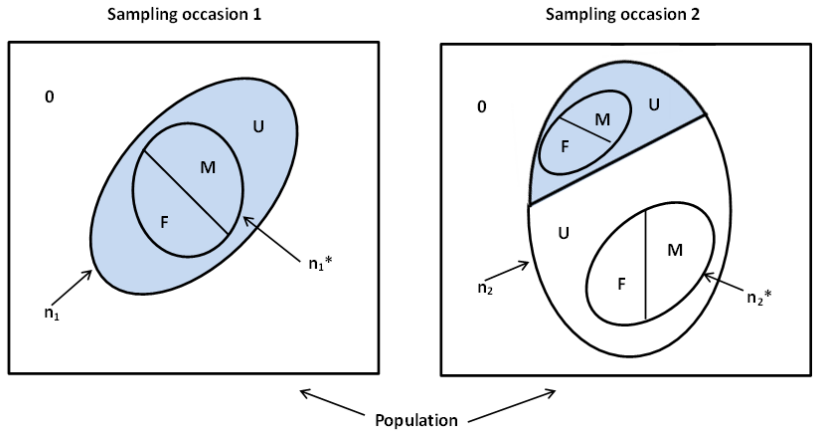
Walleye Data-Mille Lacs, MN

- Walleye are captured on the spawning grounds. Almost all the fish can be sexed in the first occasion
- All the captured fish are tagged and released and the recapture occurred 3 weeks later using gill-nets
- From a sample of fish captured at second occasion that are not tagged, a random sample is selected and sexed

Capture History	statistics
$U0$	40
UU	1
$M0$	5067
MM	40
$F0$	1551
FF	33
$0M$	41
$0F$	237
$0U$	3075

Sampling protocol

0	Fish is not captured
U	Fish is captured but not stratified
M	Fish is captured and identified as male
F	Fish is captured and identified as female



Example - Model Selection

Model Notation

- Let k stands for category and t stands for time
- Capture probability p depends on k and t, sampling fraction θ depends on t and category proportion λ depends on k
- For Example,
  - p_{k*t} implies that capture-probability varies by category and time
  - p_t implies capture-probabilities vary by time but not by category
  - $p_.$ implies capture-predictabilities do not vary by category or by time

model	np	$max\{ln(L)\}$	$\hat{N}$ (in 10^3)	$s.e.(\hat{N})$ (in 10^3)	$AICc$	$\Delta AICc$
$(p_{k*t}, \theta_t, \lambda_k)$	8	30.57	205.5	26.1	77.1	0.0
$(p_{k*t}, \theta_t, \lambda_.)$	7	33.95	208.1	24.3	81.9	4.7
$(p_k, \theta_t, \lambda_k)$	6	813.37	399.3	54.6	1638.7	1561.5
$(p_., \theta_t, \lambda_k$	5	819.58	348.5	39.9	1649.1	1572.0
$(p_k, \theta_t, \lambda_.)$	5	5840.31	399.3	54.6	11690.6	11613.4
$(p_., \theta_., \lambda_.)$	3	6673.99	348.5	39.9	13353.9	13276.8

Example

Parameter Estimation

- Parameter estimation in partial stratified two-sample capture-recapture model and simple Lincoln-Petersen estimate for population size ($\hat{N}_{LP}$)

Parameter	MLE	S.E. of the MLE	S.E. ($p_{k*t}, \theta_k \lambda_.^{MLE}$)
p_{1M}	7.58×10^{-2}	1.50×10^{-2}	9.43×10^{-3}
p_{1F}	1.15×10^{-2}	0.20×10^{-2}	1.04×10^{-3}
p_{2M}	0.78×10^{-2}	0.12×10^{-2}	1.04×10^{-3}
p_{2F}	2.09×10^{-2}	0.36×10^{-2}	2.82×10^{-3}
λ_M	3.29×10^{-1}	6.05×10^{-2}	0
λ_F	6.70×10^{-1}	6.05×10^{-2}	0
θ_1	9.93×10^{-1}	0.94×10^{-3}	9.48×10^{-3}
θ_2	0.82×10^{-1}	4.76×10^{-3}	4.76×10^{-3}
N	2055.05×10^{2}	261.38×10^{2}	254.22×10^{2}
N_M	677.84×10^{2}	133.91×10^{2}	83.85×10^{2}
N_F	1377.21×10^{2}	237.29×10^{2}	170.36×10^{2}
$\hat{N}_{LP}$	3117.64×10^{2}	347.17×10^{2}	

- Last column of the table is the estimates when the sex ratio is fixed at MLE

Example

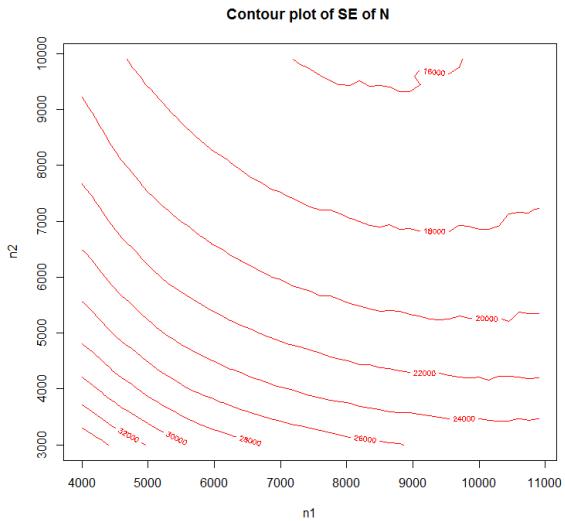
Optimal Allocation of effort

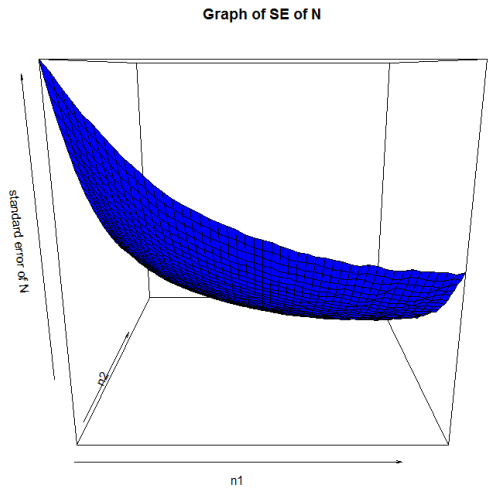
- Use MLE's of N, λ_M, r_1, and r_2 (ratios of male to female at time 1 and 2) as guestimates
- Considered $c_1 = 2$, $c_2 = c_1/2$, $c_1^* = 4$, $c_2^* = c_1^*/2 = 4$, $C_0 = 40000$

Optimal allocation
$n_1 = 8199$
$n_1^* = 3003$
$n_2 = 9511$
$n_2^* = 997$

$SE(\hat{N}) = 15904$

- Conditional Contour plot and Graph for $SE(\hat{N})$





Summary

- **Problem**
  - Capture heterogeneity is case bias in estimates in two-sample capture-recapture experiments
  - Stratification is not possible for all captured animals in each occasion
  - Stratification is costly
- **Solution**
  - A method developed using partial stratification
  - Given the relative cost of sampling for a simple capture and for processing the sub-sample, optimal allocation of effort for a given cost can be determined.
  - Still the optimal allocation method has to be fine tuned.
  - Several methods used for finding the analytical solutions for MLE, but not plausible
- **Future work**
  - Bayesian solution
    - Why intersted in bayesian approach.
  - Adding individual covariates
  - Extend the method for continuous covariates

Sincere appreciation to my Supervisor Prof. Carl James Schwarz

THANK YOU.