Capture-Recapture Analysis and Model Selection for Closed Population Models

A Bayesian approach

Sujit Ghosh

(joint work with James Norris)

http://www.stat.ncsu.edu/~sghosh2/

NC State University
DEPARTMENT OF STATISTICS

Outline

1. Capture recapture models
(*closed populations* only)

2. A motivating example: *meadow vole data*

3. The M_{bh} model

4. Bayesian analysis

    (a) Noninformative and informative priors

    (b) Model selection (for non-nested models)

5. Results for the meadow vole data

6. Conclusions

Introduction

- Capture Recapture experiments are used to estimate the unknown size of a finite population

Introduction

- Capture Recapture experiments are used to estimate the *unknown size* of a finite population Used in various scientific fields
  Zoological studies: e.g., Seber (1987)
  Census undercount: e.g., Fienberg (1992)
  Software reliability: e.g., Eick et al. (1993)

Introduction

- Capture Recapture experiments are used to estimate the unknown size of a finite population

- A class of eight models is the most widely used framework for closed population capture recapture experiments

NC STATE UNIVERSITY

Introduction

- Capture Recapture experiments are used to estimate the unknown size of a finite population

- A class of eight models is the most widely used framework for closed population capture recapture experiments

- However parameter estimation in such a flexible framework may not be easy; especially in the presence of intrinsic heterogeneity between the individuals

NC STATE UNIVERSITY

Introduction

- Capture Recapture experiments are used to estimate the unknown size of a finite population

- A class of eight models is the most widely used framework for closed population capture recapture experiments

- However parameter estimation in such a flexible framework may not be easy; especially in the presence of intrinsic heterogeneity between the individuals

- We consider the model M_{bh} that accounts for heterogeneity (h) and behavioral (b) effects

Introduction

- Capture Recapture experiments are used to estimate the unknown size of a finite population

- A class of eight models is the most widely used framework for closed population capture recapture experiments

- However parameter estimation in such a flexible framework may not be easy; especially in the presence of intrinsic heterogeneity between the individuals

- We consider the model $\mathbf{M}_{bh}$ that accounts for heterogeneity (h) and behavioral (b) effects

- We utilize a flexible discrete mixture model to account for the h and b effects.

Capture Recapture Models

A capture-recapture study collects data on

$$
\begin{aligned}
X_{ij} &= 1 \text{ if individual } i \text{ captured on period } j \\
&= 0 \text{ otherwise}
\end{aligned}
$$

where $i = 1, \ldots, N$ and $j = 1, \ldots, k$.
Let $p_{ij} = \mathrm{Pr}[X_{ij} = 1] = 1 - \mathrm{Pr}[X_{ij} = 0]$

Capture Recapture Models

A capture-recapture study collects data on

$$X_{ij} = 1 \text{ if individual } i \text{ captured on period } j$$
$$= 0 \text{ otherwise}$$

where $i = 1, \ldots, N$ and $j = 1, \ldots, k$.

Let $p_{ij} = \Pr[X_{ij} = 1] = 1 - \Pr[X_{ij} = 0]$

- Eight models are based on making assumptions on the capture probabilities; p_{ij}

- Note that all X_{ij}'s are <u>not</u> observed

- N is the unknown population size

- We assume that the population for our study is closed

Closure assumptions

1. The population is closed (geographically and demographically)

2. Each sample is drawn randomly

3. A match (non-match) is considered correct

4. All matches are recorded correctly and reported on each capture period

Closure assumptions

1. The population is closed (geographically and demographically)

2. Each sample is drawn randomly

3. A match (non-match) is considered correct

4. All matches are recorded correctly and reported on each capture period

Open models are also available that permits the inclusion of *birth, deaths, immigration and emigration*.

Closure assumptions

1. The population is closed (geographically and demographically)

2. Each sample is drawn randomly

3. A match (non-match) is considered correct

4. All matches are recorded correctly and reported on each capture period

Open models are also available that permits the inclusion of *birth, deaths, immigration and emigration.*
(however see Kendall (1999) on the robustness of closure assumptions)

Closure assumptions

1. The population is closed (geographically and demographically)

2. Each sample is drawn randomly

3. A match (non-match) is considered correct

4. All matches are recorded correctly and reported on each capture period

Open models are also available that permits the inclusion of *birth, deaths, immigration and emigration.*
(however see Kendall (1999) on the robustness of closure assumptions)

In this talk we consider only closed populations

A motivating example

The Meadow Vole Data

- *Microtus* (meadow vole) were captured, marked and recaptured

A motivating example

The Meadow Vole Data

- *Microtus* (meadow vole) were captured, marked and recaptured

- During July 1999, the usual five consecutive capture days were employed (i.e., $k = 5$)

A motivating example

The Meadow Vole Data

- *Microtus* (meadow vole) were captured, marked and recaptured

- During July 1999, the usual five consecutive capture days were employed (i.e., $k = 5$)

- However, during August 1999 (on the same site), the last two of the five capture days were contaminated by a squirrel infestation
  Thus we only have three capture periods (i.e., $k = 3$)

A motivating example

The Meadow Vole Data

- *Microtus* (meadow vole) were captured, marked and recaptured

- During July 1999, the usual five consecutive capture days were employed (i.e., $k = 5$)

- However, during August 1999 (on the same site), the last two of the five capture days were contaminated by a squirrel infestation
  Thus we only have three capture periods (i.e., $k = 3$)

- Data provided by Jim Nichols at the *Patuxent National Wildlife Research*

NC STATE UNIVERSITY

Meadow vole data

- There is very little uncontaminated data available for August 1999

Meadow vole data

- There is very little uncontaminated data available for August 1999

- So we would especially like to consider incorporating prior information

Meadow vole data

- There is very little uncontaminated data available for August 1999

- So we would especially like to consider incorporating prior information
  i.e., including the use of July 1999's data, when we make inferences about August 1999

Meadow vole data

- There is very little uncontaminated data available for August 1999

- So we would especially like to consider incorporating prior information
  Jim Nichols states that there is a similar setting over the July to August 1999 sampling window

Meadow vole data

- There is very little uncontaminated data available for August 1999

- So we would especially like to consider incorporating prior information

- As the sampling was done on five consecutive days, we model a closed population with no time effects.

Meadow vole data

- There is very little uncontaminated data available for August 1999

- So we would especially like to consider incorporating prior information

- As the sampling was done on five consecutive days, we model a closed population with no time effects. However, admittedly this modeling is approximate since rainfall and other factors varied from one day to the next during this period

NC STATE UNIVERSITY

Meadow vole data

- There is very little uncontaminated data available for August 1999

- So we would especially like to consider incorporating prior information

- As the sampling was done on five consecutive days, we model a closed population with no time effects.

- The meadow voles could have different capture probabilities and an animal can become trap shy (or trap happy) after first capture

Meadow vole data

- There is very little uncontaminated data available for August 1999

- So we would especially like to consider incorporating prior information

- As the sampling was done on five consecutive days, we model a closed population with no time effects.

- The meadow voles could have different capture probabilities and an animal can become trap shy (or trap happy) after first capture

- To allow all such possibilities, we consider the class of M_{bh} models

The $\mathbf{M}_{bh}$ Model

- N: Unknown population size

The $\mathbf{M}_{bh}$ Model

- N: Unknown population size

- p_{ij}: capture probability of the ith individual at jth the period ($i = 1, \ldots, N$ and $j = 1, \ldots, k$)

The $\mathbf{M}_{bh}$ Model

- N: Unknown population size

- p_{ij}: capture probability of the ith individual at jth the period ($i = 1, \ldots, N$ and $j = 1, \ldots, k$)

- *The M_{bh} assumption:*

$$p_{ij} = \delta_i p_i^{(1)} + (1 - \delta_i) p_i^{(2)}$$

$p_i^{(1)}$: individual i's chance of capture if it has not been captured previously (i.e $\delta_i = 1$), and

$p_i^{(2)}$: individual i's chance of capture if it has been previously captured (i.e. $\delta_i = 0$)

The $\mathbf{M}_{bh}$ Model

- N: Unknown population size

- p_{ij}: capture probability of the ith individual at jth the period ($i = 1, \ldots, N$ and $j = 1, \ldots, k$)

- *The M_{bh} assumption:*
$$p_{ij} = \delta_i p_i^{(1)} + (1 - \delta_i) p_i^{(2)}$$
$p_i^{(1)}$: individual i's chance of capture if it has not been captured previously (i.e $\delta_i = 1$), and
$p_i^{(2)}$: individual i's chance of capture if it has been previously captured (i.e. $\delta_i = 0$)

- $\mathbf{p}_i = (p_i^{(1)}, p_i^{(2)})^T \sim G$
(an unknown 2-dimensional distribution)

Discrete mixture model

- Assume that G is discrete with r points of support.

NC STATE UNIVERSITY

Discrete mixture model

- Assume that G is discrete with r points of support. In other words,

$$dG(\mathbf{p}) = \sum_{m=1}^{r} \pi_m I(\mathbf{p} = \boldsymbol{\theta}_m),$$

$\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{21})^T \dots, \boldsymbol{\theta}_r = (\theta_{1r}, \theta_{2r})^T$
are the unknown support (mass) points, and
$\pi_1, \dots, \pi_r > 0$ s.t. $\sum_{m=1}^{r} \pi_m = 1$
are the unknown probabilities

Discrete mixture model

- Assume that G is discrete with r points of support. In other words,

$$dG(\mathbf{p}) = \sum_{m=1}^{r} \pi_m I(\mathbf{p} = \boldsymbol{\theta}_m),$$

$\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{21})^T \ldots, \boldsymbol{\theta}_r = (\theta_{1r}, \theta_{2r})^T$ are the unknown support (mass) points, and $\pi_1, \ldots, \pi_r > 0$ s.t. $\sum_{m=1}^{r} \pi_m = 1$ are the unknown probabilities

- For identifiability assume that,
$$0 \leq \theta_{11} < \ldots < \theta_{1r} \leq 1$$

Discrete mixture model

- Assume that G is discrete with r points of support. In other words,

$$dG(\mathbf{p}) = \sum_{m=1}^{r} \pi_m I(\mathbf{p} = \boldsymbol{\theta}_m),$$

$\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{21})^T \ldots, \boldsymbol{\theta}_r = (\theta_{1r}, \theta_{2r})^T$ are the unknown support (mass) points, and $\pi_1, \ldots, \pi_r > 0$ s.t. $\sum_{m=1}^{r} \pi_m = 1$ are the unknown probabilities

- For identifiability assume that,
$$0 \leq \theta_{11} < \ldots < \theta_{1r} \leq 1$$

- Notice that by choosing r large, we can approximate a nonparametric Dirichlet process prior (DPP) for G

Submodels of $\mathbf{M}_{bh}$

- $M_h : \theta_{1i} = \theta_{2i}, \forall i$

- $M_b : \boldsymbol{\theta}_i = \boldsymbol{\theta}_0 = (\theta_{10}, \theta_{20})$

- $M_0 :$ a M_h model with $r = 1$

Submodels of M$_{bh}$

- $M_h : \theta_{1i} = \theta_{2i}, \forall i$

- $M_b : \boldsymbol{\theta}_i = \boldsymbol{\theta}_0 = (\theta_{10}, \theta_{20})$

- $M_0 :$ a M_h model with $r = 1$

We compare all of these submodels to the full model, M_{bh}, using a rigorous model selection criterion

Submodels of $\mathbf{M}_{bh}$

- $M_h : \theta_{1i} = \theta_{2i}, \forall i$
- $M_b : \boldsymbol{\theta}_i = \boldsymbol{\theta}_0 = (\theta_{10}, \theta_{20})$
- $M_0 :$ a M_h model with $r = 1$

We compare all of these submodels to the full model, M_{bh}, using a rigorous model selection criterion

Although all of these submodels are nested within the M_{bh} models, the submodels themselves might not be nested, e.g., M_b and M_h are not nested

Submodels of $\mathbf{M}_{bh}$

- $M_h : \theta_{1i} = \theta_{2i}, \forall i$
- $M_b : \boldsymbol{\theta}_i = \boldsymbol{\theta}_0 = (\theta_{10}, \theta_{20})$
- $M_0 :$ a M_h model with $r = 1$

We compare all of these submodels to the full model, M_{bh}, using a rigorous model selection criterion

Although all of these submodels are nested within the M_{bh} models, the submodels themselves might not be nested, e.g., M_b and M_h are not nested

This motivates us to develop a model selection criterion which allows for non-nested mixture models.

The sufficient statistic

- Sufficient count statistics:
$$\mathbf{f}^{obs} = \{f_{jl}; j = 1, \ldots, k, l = 1, \ldots, k - j + 1\}$$

f_{jl}: the number of animals that were first caught on the j-th capture period and are caught a total of l times in the k capture periods.

The sufficient statistic

- Sufficient count statistics:
$$\mathbf{f}^{obs} = \{f_{jl}; j = 1, \ldots, k, l = 1, \ldots, k - j + 1\}$$

  f_{jl}: the number of animals that were first caught on the j-th capture period and are caught a total of l times in the k capture periods.

- The probability that an animal was first captured at period j and caught a total of l times:

$$
\begin{aligned}
P_{jl} &= \binom{k-j}{l-1} \int_0^1 \int_0^1 p_1(1-p_1)^{j-1} p_2^{l-1}(1-p_2)^{k-j-l+1} dG \\
&= \binom{k-j}{l-1} \sum_{i=1}^{r} \pi_i \theta_{1i}(1-\theta_{1i})^{j-1} \theta_{2i}^{l-1}(1-\theta_{2i})^{k-j-l+1}
\end{aligned}
$$

The likelihood

The multinomial likelihood function is given by,

$$\binom{N}{S} \prod_{j=1}^{k} \prod_{l=1}^{k-j+1} P_{jl}^{f_{jl}} \left(1 - \sum_{j=1}^{k} \sum_{l=1}^{k-j+1} P_{jl} \right)^{N-S}$$

where $S = \sum_{j=1}^{k} \sum_{l=1}^{k-j+1} f_{jl}$: the total number of distinct animals caught

The likelihood

The multinomial likelihood function is given by,

$$\binom{N}{S} \prod_{j=1}^{k} \prod_{l=1}^{k-j+1} P_{jl}^{f_{jl}} \left(1 - \sum_{j=1}^{k} \sum_{l=1}^{k-j+1} P_{jl} \right)^{N-S}$$

where $S = \sum_{j=1}^{k} \sum_{l=1}^{k-j+1} f_{jl}$: the total number of distinct animals caught

- Parameters: $N, \boldsymbol{\theta}, \boldsymbol{\pi}$

- Notice that the parameter G is characterized by the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$

The likelihood

The multinomial likelihood function is given by,

$$\binom{N}{S} \prod_{j=1}^{k} \prod_{l=1}^{k-j+1} P_{jl}^{f_{jl}} \left(1 - \sum_{j=1}^{k} \sum_{l=1}^{k-j+1} P_{jl} \right)^{N-S}$$

where $S = \sum_{j=1}^{k} \sum_{l=1}^{k-j+1} f_{jl}$: the total number of distinct animals caught

- Parameters: $N, \boldsymbol{\theta}, \boldsymbol{\pi}$

- Notice that the parameter G is characterized by the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$

- Our goal is to estimate N treating other parameters as nuisance.

Some Issues with ML approach

- Likelihood is difficult to maximize; there may not be an unique maxima!

Some Issues with ML approach

- Likelihood is difficult to maximize; there may not be an unique maxima!

- Sampling distributions of the MLEs are analytically complicated

Some Issues with ML approach

- Likelihood is difficult to maximize; there may not be an unique maxima!

- Sampling distributions of the MLEs are analytically complicated

- Large sample theory could be tricky!

Some Issues with ML approach

- Likelihood is difficult to maximize; there may not be an unique maxima!

- Sampling distributions of the MLEs are analytically complicated

- Large sample theory could be tricky!

  Should we consider a large sample as $N \rightarrow \infty$? (or $k \rightarrow \infty$)

Some Issues with ML approach

- Likelihood is difficult to maximize; there may not be an unique maxima!

- Sampling distributions of the MLEs are analytically complicated

- Large sample theory could be tricky!

Should we consider a large sample as $N \to \infty$? (or $k \to \infty$)

But,... N is the parameter of interest! and k is usually fixed by design and is often small.

Some Issues with ML approach

- Likelihood is difficult to maximize; there may not be an unique maxima!

- Sampling distributions of the MLEs are analytically complicated

- Large sample theory could be tricky!

  Should we consider a large sample as $N \to \infty$? (or $k \to \infty$)

  But,... N is the parameter of interest! and k is usually fixed by design and is often small.

Any other alternative?

Bayesian Analysis

- In order to complete the model specification from a Bayesian perspective, we specify a joint prior distribution for all parameters: viz., $N, \boldsymbol{\theta}, \boldsymbol{\pi}$

Bayesian Analysis

- In order to complete the model specification from a Bayesian perspective, we specify a joint prior distribution for all parameters: viz., $N, \boldsymbol{\theta}, \boldsymbol{\pi}$

- We consider both noninformative and informative prior distributions

Bayesian Analysis

- In order to complete the model specification from a Bayesian perspective, we specify a joint prior distribution for all parameters: viz., $N, \boldsymbol{\theta}, \boldsymbol{\pi}$

- We consider both noninformative and informative prior distributions

- To model the prior distributions comparatively insensitive to posterior estimates we specify a class of non-informative (Jeffreys, 1946) priors for all the model parameters.

NC STATE UNIVERSITY

Bayesian Analysis

- In order to complete the model specification from a Bayesian perspective, we specify a joint prior distribution for all parameters: viz., $N, \boldsymbol{\theta}, \boldsymbol{\pi}$

- We consider both noninformative and informative prior distributions

- To model the prior distributions comparatively insensitive to posterior estimates we specify a class of non-informative (Jeffreys, 1946) priors for all the model parameters.

- We fit the models (i.e., obtain posterior distributions) by the popular MCMC methods using `WinBUGS`

Noninformative Priors

$$p(N) \propto 1/N^{\delta} I(N = 1, 2, \ldots)$$

$$p(\theta_{11}, \ldots, \theta_{1r}) \propto \prod_{i=1}^{r} \theta_{1i}^{\beta_1 - 1} (1 - \theta_{1i})^{\beta_1 - 1}$$

$$I(\theta_0 < \theta_{11} < \ldots < \theta_{1r} \leq 1)$$

$$\theta_{2i} \sim Beta(\beta_2, \beta_2), \quad i = 1, \ldots, r$$

$$(\pi_1, \ldots, \pi_r) \sim Dirichlet(\frac{\alpha}{r}, \ldots, \frac{\alpha}{r})$$

$\delta, \beta_1, \beta_2, \theta_0, \alpha > 0$: arbitrary but known

Noninformative Priors

$$p(N) \propto 1/N^\delta I(N = 1, 2, \ldots)$$

$$p(\theta_{11}, \ldots, \theta_{1r}) \propto \prod_{i=1}^{r} \theta_{1i}^{\beta_1 - 1} (1 - \theta_{1i})^{\beta_1 - 1}$$

$$I(\theta_0 < \theta_{11} < \ldots < \theta_{1r} \le 1)$$

$$\theta_{2i} \sim Beta(\beta_2, \beta_2), \quad i = 1, \ldots, r$$

$$(\pi_1, \ldots, \pi_r) \sim Dirichlet(\frac{\alpha}{r}, \ldots, \frac{\alpha}{r})$$

$\delta, \beta_1, \beta_2, \theta_0, \alpha > 0$: arbitrary but known

Flat prior: $\delta = 0, \beta_1 = \beta_2 = 1$ and $\alpha = r$

Noninformative Priors

$$p(N) \propto 1/N^\delta I(N = 1, 2, \ldots)$$

$$p(\theta_{11}, \ldots, \theta_{1r}) \propto \prod_{i=1}^{r} \theta_{1i}^{\beta_1 - 1} (1 - \theta_{1i})^{\beta_1 - 1}$$

$$I(\theta_0 < \theta_{11} < \ldots < \theta_{1r} \le 1)$$

$$\theta_{2i} \sim Beta(\beta_2, \beta_2), \quad i = 1, \ldots, r$$

$$(\pi_1, \ldots, \pi_r) \sim Dirichlet(\frac{\alpha}{r}, \ldots, \frac{\alpha}{r})$$

$\delta, \beta_1, \beta_2, \theta_0, \alpha > 0$: arbitrary but known

Flat prior: $\delta = 0, \beta_1 = \beta_2 = 1$ and $\alpha = r$

Posterior distribution is proper even when $\delta \le 1$

Informative Priors

- There may be biologically motivated prior information about N and G

Informative Priors

- There may be biologically motivated prior information about N and G

- This prior information can be from a (previously run) similar study

Informative Priors

- There may be biologically motivated prior information about N and G

- This prior information can be from a (previously run) similar study

- We can use the posterior distribution from previous study as the prior for current study

Informative Priors

- There may be biologically motivated prior information about N and G

- This prior information can be from a (previously run) similar study

- We can use the posterior distribution from previous study as the prior for current study

- This is most easily accomplished by beginning with the, say, noninformative prior for the previous study

Informative Priors

- There may be biologically motivated prior information about N and G

- This prior information can be from a (previously run) similar study

- We can use the posterior distribution from previous study as the prior for current study

- This is most easily accomplished by beginning with the, say, noninformative prior for the previous study

- Then multiply it by the combined likelihood of the similar study and the current study to obtain the posterior of the current study (using the sequential version of Bayes theorem)

Model Selection

- Based on the decision-theoretic framework of Gelfand and Ghosh (1998)

Model Selection

- Based on the decision-theoretic framework of Gelfand and Ghosh (1998)

- Consider the posterior predictive distribution:

$$p(\mathbf{f}^{pred}|\mathbf{f}^{obs}) = \int p(\mathbf{f}^{pred}|N,G)p(N,G|\mathbf{f}^{obs})dNdG$$

$p(\mathbf{f}^{obs}|N,G)$: the sampling distribution (likelihood)

$p(N,G|\mathbf{f}^{obs})$: the posterior distribution

Model Selection

- Based on the decision-theoretic framework of Gelfand and Ghosh (1998)

- Consider the posterior predictive distribution:

$$p(\mathbf{f}^{pred}|\mathbf{f}^{obs}) = \int p(\mathbf{f}^{pred}|N, G)p(N, G|\mathbf{f}^{obs})dNdG$$

$p(\mathbf{f}^{obs}|N, G)$: the sampling distribution (likelihood)

$p(N, G|\mathbf{f}^{obs})$: the posterior distribution

- Log-transformed Mean Square Predicted Error (MSPE): $MSPE(\mathbf{f}^{pred}, \mathbf{f}^{obs}) =$

$$\frac{2}{k(k+1)} \sum_{j+l \leq k} \left[\log(f_{jl}^{pred} + 0.5) - \log(f_{jl}^{obs} + 0.5)\right]^2$$

Model Selection

- Based on the decision-theoretic framework of Gelfand and Ghosh (1998)

- Consider the posterior predictive distribution:

$$p(\mathbf{f}^{pred}|\mathbf{f}^{obs}) = \int p(\mathbf{f}^{pred}|N, G)p(N, G|\mathbf{f}^{obs})dNdG$$

$p(\mathbf{f}^{obs}|N, G)$: the sampling distribution (likelihood)

$p(N, G|\mathbf{f}^{obs})$: the posterior distribution

- Log-transformed Mean Square Predicted Error (MSPE): $MSPE(\mathbf{f}^{pred}, \mathbf{f}^{obs}) =$

$$\frac{2}{k(k+1)} \sum_{j+l \leq k} \left[\log(f_{jl}^{pred} + 0.5) - \log(f_{jl}^{obs} + 0.5) \right]^2$$

- $GGC = E\left[MSPE(\mathbf{f}^{pred}, \mathbf{f}^{obs})|\mathbf{f}^{obs} \right]$

Advantages of GGC

- Other loss functions (e.g., square-root transformed MSPE) can also be used

Advantages of GGC

- Other loss functions (e.g., square-root transformed MSPE) can also be used

- The proposed model selection method is applicable even for non-nested mixture models

NC STATE UNIVERSITY

Advantages of GGC

- Other loss functions (e.g., square-root transformed MSPE) can also be used

- The proposed model selection method is applicable even for non-nested mixture models

- The criterion is based on the finite sample predictive performances of the models, and

Advantages of GGC

- Other loss functions (e.g., square-root transformed MSPE) can also be used

- The proposed model selection method is applicable even for non-nested mixture models

- The criterion is based on the finite sample predictive performances of the models, and

- Hence, irrespective of the dimensions of the parameter space, it can be used to compare different models that may not be nested

Advantages of GGC

- Other loss functions (e.g., square-root transformed MSPE) can also be used

- The proposed model selection method is applicable even for non-nested mixture models

- The criterion is based on the finite sample predictive performances of the models, and

- Hence, irrespective of the dimensions of the parameter space, it can be used to compare different models that may not be nested

- Easy to compute using MCMC output (in `WinBUGS`)

Advantages of GGC

- Other loss functions (e.g., square-root transformed MSPE) can also be used

- The proposed model selection method is applicable even for non-nested mixture models

- The criterion is based on the finite sample predictive performances of the models, and

- Hence, irrespective of the dimensions of the parameter space, it can be used to compare different models that may not be nested

- Easy to compute using MCMC output (in `WinBUGS`)

- Gosky (2004) compares several model selection methods for closed population (*PhD Thesis*)

Results

Consider again the meadow vole data:

August data:

$f_{11} = 25, f_{12} = 30, f_{13} = 43,$

$f_{21} = 10, f_{22} = 10,$

$f_{31} = 4$

July data:

$f_{11} = 6, f_{12} = 9, f_{13} = 6, f_{14} = 6, f_{15} = 5,$

$f_{21} = 3, f_{22} = 6, f_{23} = 3, f_{24} = 2,$

$f_{31} = 5, f_{32} = 5, f_{33} = 2,$

$f_{41} = 7, f_{42} = 0,$

$f_{51} = 4$

Source: Patuxent National Wildlife Research station

(courtesy: Jim Nichols)

$\mathbf{M}_{bh}$ with noninformative priors

$r = 1(M_b)$	mean	sd	2.5%	median	97.5%
N	123.9	1.179	123.0	124.0	127.0
$MSPE$	**0.340**	0.138	0.070	0.295	0.848
$r = 2$	mean	sd	2.5%	median	97.5%
N	124.9	2.78	123.0	124.0	132.0
$MSPE$	**0.124**	0.137	0.014	0.087	0.422
$r = 3$	mean	sd	2.5%	median	97.5%
N	134.0	17.93	123.0	127.0	194.0
$MSPE$	**0.131**	0.135	0.016	0.095	0.431

$\mathbf{M}_h$ with noninformative priors

$r = 1(M_o)$	mean	sd	2.5%	median	97.5%
N	127.8	2.845	123.0	127.0	134.0
$MSPE$	**0.550**	0.038	0.101	0.472	0.530
$r = 2$	mean	sd	2.5%	median	97.5%
N	147.1	19.35	127.0	141.0	205.0
$MSPE$	**0.292**	0.027	0.024	0.213	1.000
$r = 3$	mean	sd	2.5%	median	97.5%
N	156.7	23.49	128.0	149.0	214.0
$MSPE$	**0.303**	0.028	0.022	0.221	1.040

$\mathbf{M}_{bh}$ with informative priors

$r = 1(M_b)$	mean	sd	2.5%	median	97.5%
N	124.2	1.082	123.0	124.0	127.0
$MSPE$	**0.320**	0.129	0.058	0.298	0.641
$r = 2$	mean	sd	2.5%	median	97.5%
N	125.3	2.22	**123.0**	**125.0**	**131.0**
$MSPE$	**0.121**	0.129	0.028	0.082	0.418
$r = 3$	mean	sd	2.5%	median	97.5%
N	132.0	16.13	123.0	128.0	193.0
$MSPE$	**0.129**	0.127	0.020	0.099	0.425

$\mathbf{M}_h$ with informative priors

$r = 1(M_o)$	mean	sd	2.5%	median	97.5%
N	127.9	2.372	123.0	127.0	134.0
$MSPE$	**0.502**	0.032	0.112	0.460	0.520
$r = 2$	mean	sd	2.5%	median	97.5%
N	146.1	18.61	125.0	139.0	201.0
$MSPE$	**0.250**	0.021	0.030	0.191	0.986
$r = 3$	mean	sd	2.5%	median	97.5%
N	150.5	20.23	124.0	135.0	209.0
$MSPE$	**0.283**	0.016	0.027	0.229	1.010

MCMC graphical output
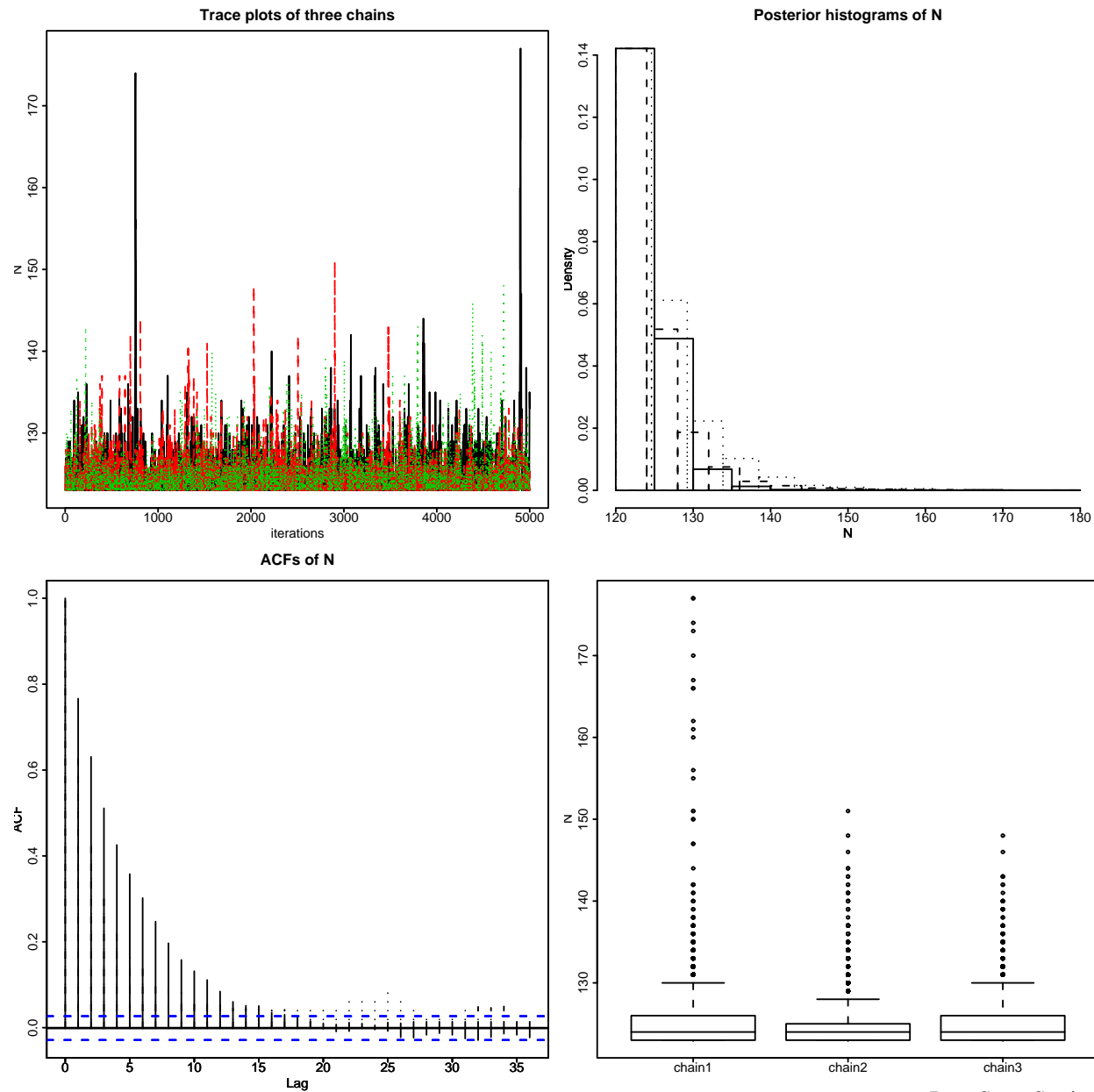
Conclusions

- Demonstrated how a flexible but user-friendly Bayesian analysis can be applied for M_{bh} models

Conclusions

- Demonstrated how a flexible but user-friendly Bayesian analysis can be applied for M_{bh} models

- Some of the major benefits are as follows:

   (i) the ability to obtain rigorous probability intervals

Conclusions

- Demonstrated how a flexible but user-friendly Bayesian analysis can be applied for M_{bh} models

- Some of the major benefits are as follows:

  (i) the ability to obtain rigorous probability intervals

  For example, for the meadow vole population we can say that there is a 95% chance that the population size is between 123 and 131

NC STATE UNIVERSITY

Conclusions

- Demonstrated how a flexible but user-friendly Bayesian analysis can be applied for M_{bh} models

- Some of the major benefits are as follows:

  (i) the ability to obtain rigorous probability intervals

  (ii) the ability to incorporate even complex information from previous studies

Conclusions

- Demonstrated how a flexible but user-friendly Bayesian analysis can be applied for M_{bh} models

- Some of the major benefits are as follows:

  (i) the ability to obtain rigorous probability intervals

  (ii) the ability to incorporate even complex information from previous studies
  For instance, we have used the July 1999 data to obtain an informative prior for August 1999 data and the subsequent analysis showed the gain in the information as compared to only using a noninformative prior

Conclusions

- Demonstrated how a flexible but user-friendly Bayesian analysis can be applied for M_{bh} models

- Some of the major benefits are as follows:

  (i) the ability to obtain rigorous probability intervals

  (ii) the ability to incorporate even complex information from previous studies

  (iii) the ability to use finite sample based model selection criteria even when competing models are not nested

Conclusions

- Demonstrated how a flexible but user-friendly Bayesian analysis can be applied for M_{bh} models

- Some of the major benefits are as follows:

  (i) the ability to obtain rigorous probability intervals

  (ii) the ability to incorporate even complex information from previous studies

  (iii) the ability to use finite sample based model selection criteria even when competing models are not nested

  e.g. GGC can be used to compare models like M_b and M_h, which are not nested

Conclusions

- Demonstrated how a flexible but user-friendly Bayesian analysis can be applied for M_{bh} models

- Some of the major benefits are as follows:

  (i) the ability to obtain rigorous probability intervals

  (ii) the ability to incorporate even complex information from previous studies

  (iii) the ability to use finite sample based model selection criteria even when competing models are not nested

  (iv) the ability to allow different estimates based on ecologically meaningful loss functions

Conclusions

- Demonstrated how a flexible but user-friendly Bayesian analysis can be applied for M_{bh} models

- Some of the major benefits are as follows:

  (i) the ability to obtain rigorous probability intervals

  (ii) the ability to incorporate even complex information from previous studies

  (iii) the ability to use finite sample based model selection criteria even when competing models are not nested

  (iv) the ability to allow different estimates based on ecologically meaningful loss functions e.g., use of the posterior median versus the posterior mean

Conclusions

- Demonstrated how a flexible but user-friendly Bayesian analysis can be applied for M_{bh} models

- Some of the major benefits are as follows:

  (i) the ability to obtain rigorous probability intervals

  (ii) the ability to incorporate even complex information from previous studies

  (iii) the ability to use finite sample based model selection criteria even when competing models are not nested

  (iv) the ability to allow different estimates based on ecologically meaningful loss functions

Conclusions

- Simple codes are provided which uses freely available software

Conclusions

- Simple codes are provided which uses freely available software

- This eliminates many of the barriers which obstruct the use of Bayesian analysis

Conclusions

- Simple codes are provided which uses freely available software

- This eliminates many of the barriers which obstruct the use of Bayesian analysis

- We have not considered models that allow time differences

Conclusions

- Simple codes are provided which uses freely available software

- This eliminates many of the barriers which obstruct the use of Bayesian analysis

- We have not considered models that allow time differences

- As a possible extension of our models, it would be worthwhile to study more general models, such as M_{th}, M_{tb} and M_{tbh}, that allows for time differences as well as heterogeneity and bait/trap effects.

Conclusions

- Simple codes are provided which uses freely available software

- This eliminates many of the barriers which obstruct the use of Bayesian analysis

- We have not considered models that allow time differences

- As a possible extension of our models, it would be worthwhile to study more general models, such as M_{th}, M_{tb} and M_{tbh}, that allows for time differences as well as heterogeneity and bait/trap effects.

THANKS!