Estimating occupancy and fitting models

Natalie Karayarsamis

 $\begin{array}{c} {\bf Department~of~Mathematics~and~Statistics} \\ {\bf La~Trobe~University} \end{array}$

RMIT, October 3, 2019

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motiv

Occupancy an detectability

Presence absence

Full likelihood function — BOI

Edge solutions

Exact mean and

Summar

wo-stage

omogeneous cas

sults

Heterogeneous ca

Numeric maximisation

WLS

esults

Results

esults

ımmary

stage R ckage 1/61

Motivation

Motivation

Occupancy and detectability

Presence-absence data

Full likelihood function — BOD model

Edge solutions

Exact mean and variance

Summary

Two-stage approach

Homogeneous case

Results

Heterogeneous case

Numeric maximisation

IWLS

Iterative method

Results

GAMs

Results

Summary

twostage R package

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivatio

Occupancy an detectability

Presence absence

Full likelihood function — BOE

Edge solutions

Exact mean and variance

ummary

'wo-stage pproach

> iogeneous cas ilts

umeric

maximisation

IWLS

Results

GAMs

Summary

ostage R ackage 2/61

Motivation

Motivation

Estimating occupancy and fitting models

Motivation

Motivation

Why monitor populations?

▶ invasive species - foxes: threat to native wildlife e.g. lyrebirds



endangered species - e.g. growling grass frog



Estimating occupancy and fitting models

Natalie Karavarsami

Motivation

Motivation

Occupancy and detectability

Presence absence data

Full likelihood function — BOI model

Edge solutions

Exact mean and
variance

Summary

wo-stage oproach

Homogeneous case Results

Heterogeneous Numeric

Numeric maximisation

IWLS

terative tesults

GAMs

Summar

wostage R ackage 4/6

Motivation

Motivation

Occupancy and detectability

Presence-absence data

Full likelihood function — BOD model

Edge solutions

Exact mean and variance

Summary

Two-stage approach

Homogeneous case

Results

Heterogeneous case

Numeric maximisation

IWLS

Iterative method

Results

GAMs

Results

Summary

twostage R package

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

1110011401011

Occupancy and detectability

Presence absence lata

Full likelihood unction — BOI

Edge solutions

ariance

wo-stage

mogeneous ca

terogeneous umeric

Numeric maximisation

IWLS

Iterative

AMs

AMs

.....

ostage R ickage 5/6:

Objective and challenge

Probability of presence

Probability of detection

 ψ

)

Estimate Occupancy in the presence of Imperfect detection

Objective

Challenge

- site = patch of land, fixed area of stream bank etc.
- site occasion = visit to a site.

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motiv

Occupancy and detectability

> resence bsence ata

Full likelihood function — BOD model

Edge solu

xact mean s ariance

Summary

Гwo-stage ipproach

mogeneous ca

eterogeneous Numeric

Numeric maximisation

IWLS

erative n sults

AMs

Results

stage R

Presence—absence data

A typical detection matrix

Site	Survey.1	Survey.2	Survey.3	Survey.4	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	
5	0	0	0	0	
6	0	0	0	0	
7	0	0	0	0	
8	1 ←	x_{ij} 1	1	1	
9	0	0	1	1	
10	1	1	1	1	
11	0	1	0	0	
12	0	1	0	0	$x_i \neq 0$
13	1	1	1	1	
14	1	1	1	1	
15	1	1	1	1	_
16	0	0	0	0	$x_i \cdot = 0$
17	0	0	0	0	
18	1	1	1	1	
19	0	0	0	0	
20	1	1	1	1	
21	1	1	1	1	
22	0	0	0	0	
23	0	1	1	1	
24	0	1	1	1	
25	0	0	0	1	
26	1	1	1	1	
27	0	0	0	0	

 $\begin{array}{l} \textbf{Table: Capture histories for the growling grass frog. The 27 independent sites each} \\ \textbf{were surveyed on 4 occasions at night within the 2002-2003 season (Heard et al., 2006)}. \end{array}$

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivat

Occupancy and detectability

Presenceabsence data

Full likelihood function — BOD model

model Edge solutions

Exact mean variance

Γwo-stage

Homogeneous case Results

leterogeneous ca

Numeric maximisation

maximisation IWLS

WLS terative me

esults AMs

esults

Summary

stage R

The number of detections X_i , at site i is distributed as

$$X_{i} \stackrel{d}{=} \begin{cases} 0, & \text{with probability } (1 - \psi), \\ \overline{\text{Bi}(T, p)}, & \text{with probability } \psi. \end{cases}$$

$$\Pr(X_{i}. = x_{i}.) = \begin{cases} \psi(1-p)^{T} + (1-\psi), & x_{i}. = 0; \\ \psi\begin{pmatrix} T \\ x_{i}. \end{pmatrix} p^{x_{i}} \cdot (1-p)^{T-x_{i}}, & x_{i}. = 1, 2, \dots, T. \end{cases}$$

Two states and three possible outcomes: detection $(x_{ij}) = 0 \rightarrow \text{not present}$, OR present but not detected detection $(x_{ij}) = 1 \rightarrow \text{present}$

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivati

Occupancy and

Presence absence lata

Full likelihood function — BOD model

Edge solutions Exact mean and variance

wo-stage

Iomogeneous case tesults

Vumeric

maximisation IWLS

> erative me sults

GAMs

Summary

wostage R

Full likelihood (ZIB) under constant $\psi \& p$

(MacKenzie et al., 2002)

$$L(\psi, p \mid \boldsymbol{X}) = \prod_{i=1}^{N} L_i(\psi, p \mid x_{i.})$$

$$= \left(\psi(1-p)^T + (1-\psi)\right)^{N-k} \psi^k p^x (1-p)^{NT-x}$$
nondetections
detections

number of detected sites

$$k = \sum_{i=1}^{N} I(x_i > 0)$$

total number of detections

$$x = \sum_{i=1}^{N} x_i.$$

k and x sufficient for ψ , p

Estimating occupancy and fitting models

Karavarsamis

Full likelihood function — BOD

model

Score equations – Full likelihood under constant $\psi~\&~p$

$$\psi_s = \frac{k}{N\theta_s}$$
 and $p_s = \frac{x\theta_s}{kT}$

 $\theta_s = 1 - (1 - p_s)^T$ (prob. of at least one detected site.)

$$\operatorname{Var}_{\operatorname{Mac}}\left(\hat{\psi}\right) = \frac{\psi}{N}\left((1-\psi) + \frac{1-\theta}{\theta - Tp(1-p)^{T-1}}\right).$$

BUT do not always give MLEs!!!

As $N, T \to \infty$ these are the MLE.

If N, T small then these will not apply.

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Occupancy an

resence bsence

Full likelihood function — BOD model

Edge solutions Exact mean and

Γwo-stage

oproach omogeneous case

Resurts Heterogeneous ca

Numeric

maximisation

lterative n lesults

GAMs

Summary

wostage R

Limitations – Full likelihood (BOD)

- 1. Convergence direct maximisation BOD does not always converge, no closed form solutions for ψ_s, p_s
- 2. Boundary issues $-\hat{\psi} > 1, \hat{p} > 1$
- 3. Standard errors no closed form solutions for Var_{Mac} , hessian not always available

(Karavarsamis et al. (2013); Karavarsamis and Huggins (2019b), Karavarsamis and Huggins (2019a), Karavarsamis and Watson (b), Karavarsamis and Watson (a))

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motiv

Occupancy a

Presence absence data

Full likelihood function — BOD model

Edge solutions

Exact mean and
variance

vo-stage

mogeneous case

sults

Numeric

maximisation

WLS

sults

AMs

nmarv

stage R ckage 11/61

Full likelihood function — BOD model

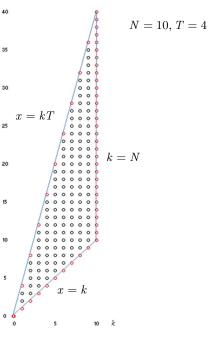
Edge solutions

Estimating occupancy and fitting models

2. Boundary problem

 $\psi_s = k/N\theta_s$ $p_s = x\theta_s/kT$

x



Estimating occupancy and fitting models

Natalie Karavarsamis

Motivatio

Motivation Occupancy and

> resencebsence ata

full likelihood unction — BOD

inction — BOL iodel Edge solutions

riance ımmary wo-stage

wo-stage pproach Iomogeneous ca

Results Heterogeneous c

umeric aximisation

WLS terative me

AMs

nmary

twostage R package 13/61

2. Boundary problem

'Edge' solutions

$$N = 10, T = 4$$

$$\hat{\psi} = \frac{\kappa}{N} \qquad \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array}$$

$$x = kT$$

x



000

$$x = k$$

$$\psi_s = k/N\theta_s$$
$$p_s = x\theta_s/kT$$

MLE is 1 but soln to score eqn > 1

$$k/N\theta_s$$

 $x\theta_s/kT$

$$x\theta_s/kT$$

fitting models

Estimating occupancy and

Limitations – Full likelihood (BOD)

This caused

- 1. non–convergence of the likelihood (too flat, multiple local maxima)
- 2. estimates that were greater than 1 i.e. $\hat{\psi} > 1$ or $\hat{p} > 1$
- 3. problems with interval estimators i.e. standard errors

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motiv

Occupancy a detectability

Presence absence

Full likelihood function — BOD

Edge solutions

Exact mean and variance

Summary

Two-stage

omogeneous case

esults eterogeneous car

Numeric maximisation

maximisation

Iterative :

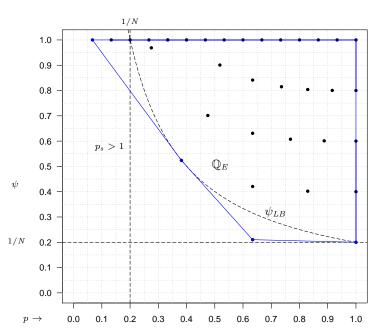
esults

Results

ummary

ostage R ackage 14/61

Plausible region – MLEs always exist (N = 5, T = 3)



Estimating occupancy and fitting models

Natalie Karavarsamis

Motivatio

Motiva

Occupancy and detectability

resencebsence lata

Full likelihood function — BOD model

Edge solutions

xact mean and

wo-stage

oproach Iomogeneou

sults

Numeric maximisation

maximisation IWLS

IWLS Iterativ

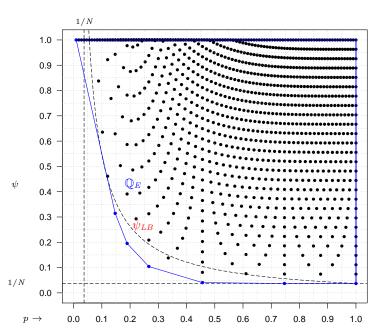
Results

GAMS Results

Summary

twostage R package 15/61

Plausible region – MLEs always exist (N = 27, T = 4)



Estimating occupancy and fitting models

Natalie Karavarsamis

Motivati

Occupancy and detectability

absence lata

function — BOD model

model
Edge solutions

wo-stage proach

pproach Iomogeneous

Results

Numeric maximisation

maximisatior IWLS

Results

GAMs Results

Summary

twostage R

Full likelihood function — BOD model

Exact mean and variance

Estimating occupancy and fitting models

Exact mean and

Problem 3. Standard errors – exact mean and variance

We derived an expression for the joint pmf of (X, K).

This allowed us to evaluate the

- \blacktriangleright bias of $\hat{\psi} = \hat{\psi}(x,k)$
- ightharpoonup exact variance of $\hat{\psi}$

Results

- bias—corrections for $\hat{\psi}$ not so effective for small N, T, or p because not enough information in (x, k)
- asymptotic variance underestimates actual variance

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivo

Occupancy a

Presence bsence

Full likelihood function — BOD

Edge solutions Exact mean and

ariance

wo-stage

Homogeneous case Results

Numeric

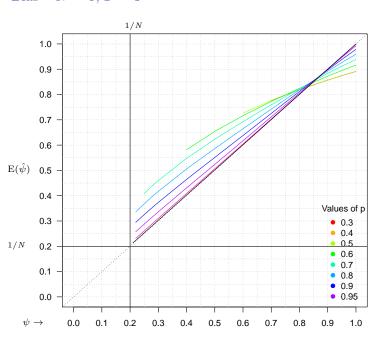
maximisatio

erative m

esults AMs

Results

wostage R oackage 18/61



Estimating occupancy and fitting models

Natalie Karavarsamis

Motivati

Motivati

Occupancy an letectability

absence data Full likelihood

function — BOI model

Edge solutions
Exact mean and

wo-stage

omogeneous case

Heterogeneous ca Numeric

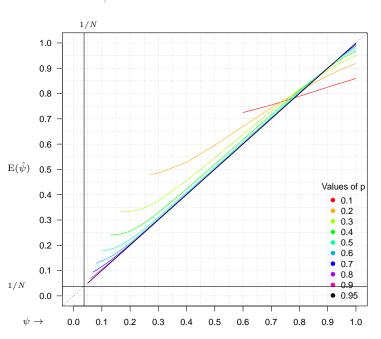
Numeric maximisation IWLS

IWLS

Results GAMs

ummary

twostage R package 19/6



Estimating occupancy and fitting models

Natalie Karavarsamis

Motivati

Madiana

Occupancy an

resencebsence lata

Full likelihood function — BOD model

model Edge solutions

Exact mean and variance Summary

> wo-stage oproach

lomogeneous

tesuits Ieterogeneous c

Numeric maximisation

IWLS

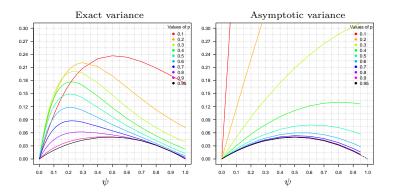
Iterative n Results

GAMs

Summary

twostage R backage 20/61

Asymptotic and exact variance for $\hat{\psi}$: N = 5, T = 3



Estimating occupancy and fitting models

Natalie Karavarsamis

Motivatio

Motivo

Occupancy a

resencebsence

Full likelihood function — BOD

Edge solutions Exact mean and

variance Summary

wo-stage

proacn mogeneous ca

sults

umeric aximisation

WLS

terative m esults

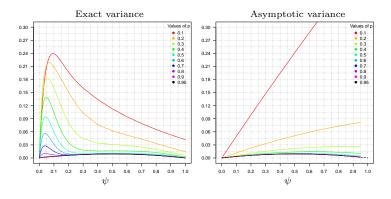
esults.

ımmary

wostage R ackage 21/61

Asymptotic and exact variance for $\hat{\psi}$: N = 27, T = 4

Frogs $\hat{\psi} = 0.557, \hat{p} = 0.782, ase = 0.096, se = 0.095$



Estimating occupancy and fitting models

Natalie Karavarsamis

Motivatio

Motiv

Occupancy a detectability

> resencebsence

Full likelihood function — BOD model

model

Edge solutions

Exact mean and

Summary

wo-stage

omogeneous ca

eterogeneous Jumeric

Numeric maximisation

WLS

Results

ummarv

twostage R package 22/61

Full likelihood function — BOD model

Summary

Estimating occupancy and fitting models

Summary

Full likelihood limitations and solutions — summary

1. Non-convergence of the likelihood (identifiability)

- 1.1 next...
- 2. Boundary issues $(\psi_s, p_s > 1)$
 - 2.1 Edge solutions and Plausible region (Karavarsamis & Watson, 2019 in prep.)
- 3. Standard errors
 - 3.1 Exact variance showed asymptotic not good (Karavarsamis *et al.*, 2013)
- 4. Too hard to include covariates
 - 4.1 next...

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

34

Occupancy a detectability

Presence absence

Full likelihood function — BOD

Edge solutions

Summary

Two-stage

mogeneous cas

eterogeneous

Numeric maximisation

IWLS

terative in esults

GAMs

,....

twostage R package 24/61

Two-stage approach

Homogeneous case

Estimating occupancy and fitting models

Homogeneous case

Presence-absence data

Characteristics

- Repeated visits to a site (introduces heterogeneity may be solved with covariates)
- ▶ Observe presence—absence of a species
- Covariate information
 - ▶ site characteristics geographic...
 - species characteristics
 - ψ : habitat type, patch size, age, gender...
 - p: weather, site accessibility, detection methods...

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motiva

Occupancy ar detectability

Presence absence data

Full likelihood function — BOI model

Edge solutions Exact mean and

ummary

Two-stage

Homogeneous case

esults

Numeric

maximisation

IWLS

esults

AMs esults

mmary

twostage R

Full and partial likelihoods

Full likelihood

- estimates highly variable
- ▶ too hard to include covariates e.g. Welsh et al. (2013), and not a GLM

Existing methods

▶ unmarked, bootstrap, Bayesian methods

Partial likelihood (two-stage approach)

- ▶ easy to include non–linear covariates i.e. GAMs!
- resolves limitations e.g. efficient closed form variance approximations
- reduces the dimension of models

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivati

Occupancy and detectability

Presence ibsence lata

Full likelihood function — BOD

Edge solutions Exact mean an

Summary

I'wo-stage approach

Homogeneous case

Heterogeneous Numeric

maximisation

IWLS

esults

AMs Lesults

ummary

ostage R .ckage **27/61**

Partial likelihoods – benefits

What we want:

▶ to include non–linear covariates with GAMs and have full use of GLM methodology

How to achieve goals:

- \triangleright partials suit this well and allow to consider ψ and p separately
- repeated observations at each site give more info on detections
- ▶ more info on detections encourages us to consider (i.e. to estimate) detections separately from occupancy which ignores info on 1st detections
- ▶ achieve this with partial likelihoods (they simplify complex likelihoods and deal with nuisance params) need to ignore info on first detections but no great loss of efficiency
- \blacktriangleright two–stage estimation for ψ and p gives full use of GLMs etc
- now can get variance approximations too
- ▶ standard errors are readily obtainable, unlike those obtained from inverting the hessian of the full likelihood that may fail especially near the boundaries of the parameter space in about 5% 20% of cases

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivatio

Motivation Occupancy as

> Presence absence lata

Full likelihood function — BOD model

Edge solutions Exact mean and variance

wo-stage pproach

Homogeneous case Results

Numeric

maximisation IWLS

terative m esults

esults

ummarv

twostage R

Homogenous case – ψ and p constant

Partial likelihood

$$L(\psi, p) \propto (1 - \psi + \psi(1 - p)^{T})^{N - k} \prod_{i=1}^{k} \psi \left[p^{x_{i}} \cdot (1 - p)^{T - x_{i}} \cdot \right]$$

$$= (1 - \psi\theta)^{N - k} \psi^{k} \left[\prod_{i=1}^{k} (1 - p)^{a_{i} - 1} p \right] \left[p^{x - k} (1 - p)^{b - (x - k)} \right]$$

$$= L_{1}(\psi, p) L_{2}(p)$$

- \bullet omit first detections, a_i
- total number of occasions after a_i is b

Now estimate ψ and p SEPARATELY!

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivatio:

Motiva

Occupancy a detectability

> resence bsence

Full likelihood function — BOD

model Edge solutions

Exact mean variance

Two-stage approach

Homogeneous case Results

Results Heterogeneous

Numeric

maximisat IWLS

Iterative

GAMs

esults

ummary

wostage R ackage 29/61

Homogenous case – ψ and p constant

Partial likelihood

$$L(\psi, p) \propto (1 - \psi + \psi(1 - p)^{T})^{N - k} \prod_{i=1}^{k} \psi \frac{p^{x_{i}} \cdot (1 - p)^{T - x_{i}}}{p^{x - k} (1 - p)^{k - k}}$$

$$= (1 - \psi\theta)^{N - k} \psi^{k} \times \boxed{p^{x - k} (1 - p)^{b - (x - k)}}$$

$$= L_{1}(\psi, p) \times L_{2}(p)$$

- omit first detections, a_i
- total number of occasions after a_i is b

Now estimate ψ and p SEPARATELY!

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motiva

Occupancy a detectability

> Presence absence

Full likelihood function — BOD

Edge solutions Exact mean an

Summary

approach

Homogeneous case Results

Heterogeneous Numeric

maximisatio

terative m

GAMs

ımmary

ostage R ackage 29/61

Score equations — homogeneous partial likelihood

Stage 1: $L_2(p)$ gives

$$\hat{p} = \frac{x-k}{b},$$
 $\operatorname{Var}(\hat{p}) = \hat{p}(1-\hat{p})/b$

Stage 2: $L_1(\psi, \hat{p})$ gives

$$\hat{\psi} = \frac{k}{N\hat{\theta}}$$

and

$$\begin{aligned} \operatorname{Var}(\hat{\psi}) &= \operatorname{Var}\left\{\operatorname{E}\left(\hat{\psi} \mid b, k\right)\right\} + \operatorname{E}\left\{\operatorname{Var}\left(\hat{\psi} \mid b, k\right)\right\} \\ &\approx \frac{\psi(1 - \psi\theta)}{N\theta} + \left(\frac{\psi(1 - \psi\theta)}{N\theta} + \psi^2\right) \frac{T^2(1 - p)^{2(T - 1)}}{\theta^2} \frac{p(1 - p)}{b} \end{aligned}$$

Now we have closed form solutions, yipee!!!

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motiva

Occupancy and detectability

Presence absence

Full likelihood function — BOD

Edge solutions

variance

Two-stage

Homogeneous case Results

Heterogeneous cas

Numeric maximisation

IWLS

Results

AMs Lesults

Summary

wostage R ackage 30/61

Two-stage approach

Results

Estimating occupancy and fitting models

Results

Full versus partial likelihood - homogeneous case

Simulations – small p, large NExamine efficiency of partial likelihood

	Partial		Full	
	\hat{p}	$\hat{\psi}$	\hat{p}	$\hat{\psi}$
N = 1000, T = 5	0.050	0.400	0.050	0.400
Med. est.	0.049	0.407	0.049	0.407
Median s.e.	0.016	0.124	0.015	0.120
Mad	0.016	0.127	0.015	0.123
Efficiency	1.021	0.988		
N = 100, T = 5	0.200	0.600	0.200	0.600
Med. est.	0.198	0.609	0.197	0.609
Median s.e.	0.040	0.106	0.037	0.101
Mad	0.051	0.103	0.036	0.101
Efficiency	0.843	0.909		

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivatio

Motiv

Occupancy detectability

Presence absence data

Full likelihood function — BOD model

model

Edge solutions

Summar

'wo-stage pproach

omogeneous cas

Results

Numeric

maximisatio

IWLS

esults

GAMs

lesults

wostage R oackage 32/61

Full versus partial likelihood - homogeneous case

Simulations for N = 27, T = 4.

	Par	tial	Full	
	\hat{p}	$\hat{\psi}$	\hat{p}	$\hat{\psi}$
Value	0.600	0.600	0.600	0.600
Median estimate	0.600	0.604	0.600	0.604
Mad	0.078	0.104	0.065	0.105
Median s.e.	0.078	0.097	0.066	0.097
Efficiency	0.709	0.991		

Application – Frogs

$$\hat{\psi}_{Part} = 0.556$$
, s.e. $= 0.098$

$$\hat{p}_{Part} = 0.889$$
, s.e. = 0.052

$$\hat{\psi}_{Full} = 0.557$$
, s.e. = 0.096

$$\hat{p}_{Full} = 0.780$$
, s.e. = 0.054

Estimating occupancy and fitting models

Results

package 33/61

 Eventhough we ignored 1st detections, the partial works well, no significant loss of efficiency

► Analytic forms for the estimators means more stable than full likelihood

► These results are encouraging to pursue with the two-stage approach for heterogeneous case and for including covariates

Natalie Karavarsamis

Motivation

Motiv

Occupancy : detectability

> resence bsence

Full likelihood function — BOD

Edge solutions
Exact mean and

variance Summary

> proach omogeneous case

Results

Heterogeneous Numeric

maximisation IWLS

terative m

esults AMs

tesults

ostage R

Outline

Motivation

Motivation

Occupancy and detectability

Presence—absence data

Full likelihood function — BOD model

Edge solutions

Exact mean and variance

Summary

Two-stage approach

Homogeneous case

Results

Heterogeneous case

Numeric maximisation IWLS

Iterative method

Results

CAMe

Results

Summary

twostage R package

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivation

Occupancy as

Presence bsence

Full likelihood function — BOD

dge solutions

ariance

wo-stage oproach

omogeneous ca esults

Heterogeneous case Numeric

maximisation

IWLS

terative 1

AMs

sults

tesuits

wostage R backage 35/63

Site inhomogeneity

• ψ , p vary between sites but constant within sites.

Contribution of a single site is

$$L_{i}(\psi_{i}, p_{i}) = \left\{1 - \psi_{i} + \psi_{i}(1 - p_{i})^{\tau}\right\}^{1 - w_{i}} \left\{ \begin{pmatrix} \tau \\ x_{i} \end{pmatrix} \psi_{i} p_{i}^{x_{i}} \cdot (1 - p_{i})^{\tau - x_{i}} \right\}^{w_{i}}$$

$$\propto (1 - \psi_{i}\theta_{i})^{1 - w_{i}} \psi_{i}^{w_{i}} \left\{ p_{i}(1 - p_{i})^{(a_{i} - 1)} \right\}^{w_{i}} \left\{ p_{i}^{(x_{i} - 1)} (1 - p_{i})^{b_{i} - x_{i}} \cdot + 1 \right\}^{w_{i}}$$

$$= L_{1i}(\psi_{i}, p_{i}) L_{2i}(p_{i})$$

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

......

Occupancy and

resencebsence ata

Full likelihood function — BOD

model

Edge solutions

ariance Summary

wo-stage proach

mogeneous case sults

Heterogeneous case

naximisation WLS

erative m sults

AMs

ummary

twostage R package 36/61

Site inhomogeneity

• ψ , p vary between sites but constant within sites.

Contribution of a single site is

$$L_{i}(\psi_{i}, p_{i}) = \left\{1 - \psi_{i} + \psi_{i}(1 - p_{i})^{\tau}\right\}^{1 - w_{i}} \left\{ \begin{pmatrix} \tau \\ x_{i} \end{pmatrix} \psi_{i} \frac{p_{i}^{x_{i}} \cdot (1 - p_{i})^{\tau - x_{i}}}{p_{i}^{x_{i}} \cdot (1 - p_{i})^{\tau - x_{i}}} \right\}^{w_{i}}$$

$$\propto (1 - \psi_{i}\theta_{i})^{1 - w_{i}} \psi_{s}^{w_{i}} \times \left\{ p_{i}^{(x_{i}, -1)} (1 - p_{i})^{b_{i} - x_{i}, +1} \right\}^{w_{i}}$$

$$= L_{1i}(\psi_{i}, p_{i}) \times L_{2i}(p_{i})$$

...ignore first detections

$$L_i(\eta_i, p_i) = L_{1i}(\eta_i) \times L_{2i}(p_i), \qquad \eta_i = \psi_i \theta_i$$

Now we can easily include covariates...

Dimension of models to check reduces significantly! $(w_i = presence)$

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivatio

Occupancy and detectability

'resencebsence ata

Full likelihood function — BOD model

Edge solution Exact mean : variance

wo-stage

omogeneous case

Heterogeneous case Numeric

maximisation

w*L3* terative me

lesults AMs

sults

vostage R

Benefits of two-stage approach

- not impacted by boundary conditions
- full likelihood numerically unstable eg without constraints non-convergence, local maxima, or ests outside parameter space or extreme ests
- Bayes method may underestimate variance of posterior distribution
- penalized likelihood methods to help with instability eg occuPEN, occuPEN_CV, two-stage
- ▶ faster than full likelihood (occu in unmarked via optim) as it reduces dimension of parameter space
- covariates may be related to detection or occupancy to be associated with each site separately
- covariates may vary with time
- ▶ full access to R glm machinery, vglm and vgam etc

(Karavarsamis & Huggins (2019) (CSDA))

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivatio

Occupancy as detectability

Presence absence lata

Full likelihood function — BOD model

Edge solutions Exact mean an variance

approach Homogeneous case

Results Heterogeneous case

Numeric

maximisation

IWLS

esults

tesults

Summary

ostage R

We model detections with a conditional likelihood

$$L(\eta_s, \boldsymbol{p}_s) = (1 - \eta_s)^{z_s} \eta_s^{1 - z_s} \times \left\{ \frac{\prod_{j=1}^{\tau} p_{sj}^{y_{sj}} (1 - p_{sj})^{1 - y_{sj}}}{\theta_s} \right\}^{1 - z_s}$$
$$= L_1(\eta_s) L_2(\boldsymbol{p}_s).$$

The contribution of site s to the log-likelihood is then

$$\ell(\eta_s, \boldsymbol{p}_s) = z_s \log(1 - \eta_s) + (1 - z_s) \log(\eta_s) \quad (1)$$

+
$$(1-z_s)$$
 $\left\{\sum_{j=1}^{\tau} y_{sj} \log(p_{sj}) + \sum_{j=1}^{\tau} (1-y_{sj}) \log(1-p_{sj}) - \log(\theta_s)\right\}.$ (2)

Use (2) to get $\hat{\beta}$ (\hat{p}) then use these ests to obtain $\hat{\alpha}$ ($\hat{\psi}$) from (1).

Replace η_s by $\tilde{\eta}_s = \psi_s \hat{\theta}_s$ in the log-partial likelihood (1) and maximise this to estimate α .

$$(z_s = 1 - w_s \text{ indicator of no detections})$$

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivati

detectability

'resence bsence lata

Full likelihood function — BOD model

Edge solutions Exact mean and variance

'wo-stage

omogeneous case

Heterogeneous case

maximisation

IWLS

esults

AMs esults

ummarv

twostage R

Stage 1: Estimate detection p and coefficients

$$L_2(\beta) = \prod_{s=1}^{O} \frac{p_s^{y_s} (1 - p_s)^{\tau - y_s}}{\theta_s},$$

a function of the number of detections at each site where there was at least one detection, i.e. y_s , $s = 1, \ldots, O$.

- model redetections with logistic regression (positive binomial family in vgam)
- simple binomial function e.g. GLMs, GAMs, VGAMs etc.
- $p(\mathbf{u}_s, \boldsymbol{\beta}) = h(\mathbf{u}_s^T \boldsymbol{\beta}), j = 1, \dots, \tau$, vector of covariates \mathbf{u}_s and coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^q$. $h(x) = (1 + \exp(-x))^{-1}$, logistic function.

Now can use GLM family and covariates to get

- $\widehat{\beta}$ unlike the simple homogeneous model the conditional likelihood estimators will not be the mle's
- estimated covariance \hat{V}_{β} for $\hat{\beta}$

 - $\begin{array}{ll}
 \bullet & \hat{p} \\
 \bullet & \operatorname{Var}(\hat{p}) = \widehat{\beta}^T \widehat{V}_{\beta} \widehat{\beta}
 \end{array}$

Estimating occupancy and fitting models

Natalie Karavarsamis

Heterogeneous case

Stage 1: Estimate detection p and coefficients

- ▶ Fitting the detection model for time homogeneous covariates.
- ▶ With the univariate response *Y*, the implementation is very similar to glm.
- ► Term omit.constant=TRUE does not affect the fitting but removes the constant terms from the computation of the AIC.
- ▶ Data frame 'data' is a reduced data frame that contains data from the sites where occupancy was detected.
- ▶ Variable Y is the number of times the species was detected at each occupied site,
- au $\tau (=3)$ is the number of visits to each site $s, (S=1, \dots 656 \text{ sites})$
- ➤ Site covariates are vegcov1,vegcov2,...,vegcov6.

```
> V.out=vglm(cbind(Y,3-Y) vegcov1+vegcov2+vegcov3+vegcov4
+vegcov5+vegcov6, family=posbinomial(omit.constant=TRUE),data=data)
```

```
> coef(V.out) (Intercept) vegcov1 vegcov2 vegcov3 vegcov4
1.5590909 0.5493825 -0.2512287 -0.1048756 0.1656597
vegcov5 vegcov6
0.1186192 -0.1277806
```

Hutchinson et al. (2015) avian data

Estimating occupancy and fitting models

Natalie Karavarsamis

Aotivatior Motivation

Occupancy and letectability

resencebsence .ata

Full likelihood function — BOD model

Exact mean and variance

i wo-stage approach Homogeneous case

Results Heterogeneous case

Numeric

maximisatio

Iterative i

GAMs

Summary

wostage R ackage 40/6

- \triangleright distinct probabilities p_{sj} , $j=1,\ldots,\tau$ for different visits to site s.
- detections form a sequence of independent Bernoulli trials, we observe the outcome if there is at least one detection ie redetections
- \triangleright covariate vector u_{sj} contains an indicator of the visit time
- modelled by allowing the intercept to vary with the visit j, and easily implemented in VGAM package.
- ▶ four time dependent covariates measured for each visit to each site: time, temp, cloud, julian
- ▶ time1,..., time2, ..., julian2, julian3

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivat

Occupancy ar detectability

> resence bsence lata

Full likelihood function — BOD model

Edge solutions

Exact mean and variance

wo-stage oproach

omogeneous case esults

Heterogeneous case

Numeric maximisation

IWLS Iterative me

terative me esults

esults

wostage R oackage 41/61

Stage 1: Time Dependent Covariates for Detection

- ▶ fitting these models is more complex as many more models are available and the response consists of the detections on each visit to the site and is hence multivariate.
- ▶ time dependent intercepts and the relationship with the site covariates remains independent of time
- specified through the parallel.t argument to the posbernoulli.t family.
- ▶ parallel.t=FALSE~1 is the default for the posbernoulli.t family

```
> V.out=vglm(cbind(survey1,survey2,survey3) \sim vegcov1+vegcov2+vegcov3+vegcov4+vegcov5+vegcov6, family=posbernoulli.t(parallel.t=FALSE \sim 1), data=data) > coef(V.out) (Intercept):1 (Intercept):2 (Intercept):3 vegcov1 vegcov2 1.8583766 1.5130892 1.3527893 0.5515551 -0.2522520 vegcov3 vegcov4 vegcov5 vegcov6 -0.1052631 0.1663444 0.1190595 -0.1282785
```

```
Estimating
occupancy and
fitting models
```

Natalie Karavarsamis

Motivatio

Occupancy a detectability

Presence absence lata

Full likelihood function — BOD model

> Edge solutions Exact mean and variance

> 'wo-stage pproach

omogeneous cas esults

Heterogeneous case Numeric

Numeric maximisatio

IWLS Iterative n

Results

Results .

ımmary -

ostage R ackage 42/63

Stage 1: Time Dependent Covariates for Detection

- time dependent covariates are time, temp, cloud and julian measured for each visit to each site
- included in the data data frame as time1, time2, ... julian2, julian3
- fit time varying covariates but constant intercept
- requires use of the xij and form2 arguments in VGAM

```
> V.out=vglm(cbind(survev1.survev2.survev3)
  ~vegcov1+vegcov2+vegcov3+vegcov4
    +vegcov5+vegcov6+time.tij+temp.tij+cloud.tij+julian.tij,
 data=Data.all.
 xij=list(time.tij~time1+time2+time3-1,temp.tij~temp1+temp2+temp3-
1.
 cloud.tij~cloud1+cloud2+cloud3-1, julian.tij~julian1+julian2
    +julian3-1),
 family=posbernoulli.t(parallel.t=FALSE~0),
 form2=~vegcov1+vegcov2+vegcov3+vegcov4+vegcov5+vegcov6+time.tij
    +temp.tij+cloud.tij+julian.tij+time1+time2+time3+temp1+temp2
    +temp3+cloud1+cloud2+cloud3+julian1+julian2+julian3)
> coef(V.out)
 (Intercept) vegcov1 vegcov2 vegcov3
                                                 vegcov4
  1.60651791
              0.54525171 -0.24061702 -0.08727207
                                                  0.16955603
 vegcov5 vegcov6 time.tij temp.tij cloud.tij julian.tij
 0.108527 - 0.112085 - 0.069456 - 0.238605 - 0.161028 - 0.264658
Figure: Fitting a model using with time varying covariates but constant intercept for
```

the two-stage approach in vglm.

Estimating occupancy and fitting models

Heterogeneous case

Estimating

Ms

nmary

twostage R

Three methods:

1) Direct maximisation of the 1st partial likelihood, $L_{1s}(\psi_s, p_s)$ as a function of ψ_s , $(1 - \psi_s \theta_s)^{1-w_s} \psi_s^{w_s}$

2) Iterative Weighted Least Squares $\,$

3) Iterative method

$$\prod_{s=1}^{S} L_{1s}(\widetilde{\eta}_s)$$

where p_s and hence θ_s has been replaced by its estimator from Stage 1: $\hat{p}_s = p_s(\hat{\beta})$.

$$L_1(\boldsymbol{\alpha}) = \prod_{s=1}^S L_{1s}(\widetilde{\eta}_s) \propto \prod_{s=1}^S (1 - \psi_s \widehat{\theta}_s)^{z_s} \psi_s^{1-z_s}$$

Let $w_s = 1 - z_s$, then the log-partial likelihood is

$$\ell(\boldsymbol{\alpha}) = \sum_{s=1}^{S} \left\{ (1 - w_s) \log(1 - \psi_s \widehat{\theta}_s) + w_s \log(\psi_s) \right\}.$$

This may be maximised numerically using the optim function in R. However, there are two other possible approaches.

$$\psi_s = h(x_s^T \boldsymbol{\alpha})$$

where x_s is a vector of covariates associated with site s and $\boldsymbol{\alpha} \in \mathbb{R}^p$ is a vector of coefficients.

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Occupancy and detectability

resenceosence ata

Full likelihood function — BOD model

Exact mean and variance

wo-stage pproach

Results

Numeric maximisation

WLS

sults

Results

oummary

For a logistic model, let matrix X have sth column x_s

- $\mathbf{v} = (w_1, \dots, w_S)^T, E(w_s) = \eta_s = \theta_s \psi_s, \, \boldsymbol{\eta} = (\eta_1, \dots, \eta_S)^T$
- Then, as θ_s is not a function of α , maximising the partial log-likelihood is equivalent to maximising $\ell(\eta) = \sum_{s=1}^{S} \{(1 w_s) \log(1 \eta_s) + w_s \log(\eta_s)\}.$
- $ightharpoonup \eta(\alpha)$ be η evaluated at α .
- ► Set $V = \operatorname{diag}\{(1 \boldsymbol{\eta})\boldsymbol{\eta}\}$ and $U = \operatorname{diag}\{\theta_s\psi_s(1 \psi_s)\}.$
- $\boldsymbol{\alpha}^{(k)}$ is estimate at the kth step and let $\boldsymbol{Z} = UX\boldsymbol{\alpha}^{(k)} + \boldsymbol{w} \boldsymbol{\eta}(\boldsymbol{\alpha}^{(k)}).$
- $ightharpoonup u(\alpha)$ are the partial score equations

Then the estimate at the (k+1)th is

$$\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} + J(\boldsymbol{\alpha})^{-1} \boldsymbol{u}(\boldsymbol{\alpha}^{(k)})$$
$$= \left(XUV^{-1} UX^{T}\right)^{-1} XUV^{-1} U\boldsymbol{Z}$$

The IWLS estimate is obtained by repeating this step until convergence.

An estimate of the expected Fisher information corresponding to the partial likelihood, $E\{I(\boldsymbol{\alpha},\boldsymbol{\beta})\}$, is given by $\tilde{I}(\boldsymbol{\alpha},\boldsymbol{\beta}) = XUV^{-1}UX^{T}$. Derivations in Karavarsamis & Huggins (2015), CSDA

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivatio:

Occupancy as detectability

Presence absence data

full likelihood function — BOD model

Eage solutions
Exact mean an
variance

Γwo-stage pproach

lomogeneous cas tesults

Numeric

IWLS

Iterative n

GAMs

limmary

ostage R

Method 3: Iterative method

Under the logistic model

$$\psi_i(x) = \frac{\exp(\alpha^T x_i)}{1 + \exp(\alpha^T x_i)}$$

$$\psi_i(x)\theta_i = \frac{\exp(\alpha^T x_i + \log(\theta_i))}{1 + \exp(\alpha^T x_i)}.$$

If

$$a_i = \log(\theta_i) - \log\{1 + \exp(\alpha^T x_i)(1 - \theta_i)\}\$$

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

3.6-41----

Occupancy ar

resencebsence

Full likelihood function — BOI

Edge solutions Exact mean and

wo-stage

pproach Iomogeneous case

esults

Vumeric naximisation

maximisatio IWLS

Iterative method Results

GAMs Results

ummary

twostage R

Method 3: Iterative method

Under the logistic model

$$\psi_i(x) = \frac{\exp(\alpha^T x_i)}{1 + \exp(\alpha^T x_i)}$$

$$\psi_i(x)\theta_i = \frac{\exp(\alpha^T x_i + \log(\theta_i))}{1 + \exp(\alpha^T x_i)}.$$

If

$$a_i = \log(\theta_i) - \log\{1 + \exp(\alpha^T x_i)(1 - \theta_i)\}$$

∴ Offset

then

$$\psi_i \theta_i = \frac{\exp(\alpha^T x_i + a_i)}{1 + \exp(\alpha^T x_i + a_i)}$$

• a_i is function of linear predictor $\alpha^T x_i$.

 \therefore Iterative approach

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motiva

Occupancy a

Presenc absence

Full likelihood function — BOD

Edge solutions Exact mean and

wo-stage

pproach Iomogeneous case

esults eterogeneous cas

Numeric maximisation

IWLSIterative method

esults

esults

ummary

ostage R ackage 46/61

Method 3: Iterative method

- $\widehat{\alpha}_0$: initial estimate for α from GLM without offset
 - 1. $\widehat{\alpha}_{s-1}$ is estimate of α from previous step, s-1, then

$$a_i^{(s)} = \log(\theta_i) - \log\{1 + \exp(\alpha_{s-1}^T x_i)(1 - \theta_i)\}.$$

2. Fit GLM to the w_i with offset $a_i^{(s)}$ to produce a new $\widehat{\alpha}_i$.

Repeat steps 1 and 2 until convergence

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

.

Occupancy a

Presence absence

Full likelihood function — BOD

Edge solution Exact mean a

Summary

ipproach Homogeneous case

Ieterogeneous Numeric

maximisation

IWLS Iterative method

Results

Results

ostage R

Variance for $\widehat{\alpha}$ and $\widehat{\psi}$

$$\widehat{\operatorname{Var}}\{\widehat{\alpha}(\widehat{\beta})\} \approx$$

$$I\{\widehat{\alpha}(\widehat{\beta}),\widehat{\beta}\}^{-1} + I\{\widehat{\alpha}(\widehat{\beta}),\widehat{\beta}\}^{-1}\widetilde{B}\{\widehat{\alpha}(\widehat{\beta}),\widehat{\beta}\}\}\widehat{V}_{\beta}\widetilde{B}\{\widehat{\alpha}(\widehat{\beta}),\widehat{\beta}\}^{T}I\{\widehat{\alpha}(\widehat{\beta}),\widehat{\beta}\}^{-1},$$

gives

$$\widehat{\operatorname{Var}}(\widehat{\psi}_i) = \{\widehat{\psi}_i(1 - \widehat{\psi}_i)\}^2 x_i^T \widehat{\operatorname{Var}}\{\widehat{\alpha}(\widehat{\beta})\} x_i$$

- \widehat{V}_{β} covariance for $\widehat{\beta}$
- $Q(\alpha, \beta) = \partial l(\alpha, \beta)/\partial \alpha$ partial score function
- $I(\alpha, \beta) = -\partial Q(\alpha, \beta)/\partial \alpha$
- $\widetilde{B}\{\alpha(\beta),\beta\} = \partial Q(\alpha,\beta)/\beta$

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivation

resencebsence

Full likelihood function — BOI

Edge solutions Exact mean and

wo-stage

omogeneous case esults

Numeric

maximisation IWLS

Iterative method Results

esults.

twostage R

Outline

Two-stage approach

Results

Estimating occupancy and fitting models

Method 3: Iterative method - Results

1000 simulations for large N and small T with 2 covariates

- Relatively unbiased
- s.e. reasonable

	\widehat{lpha}_0	\widehat{lpha}_1	$\widehat{\alpha}_2$
N = 1000, T = 4	1.000	0.500	0.500
Median	1.007	0.509	0.508
mad	0.207	0.095	0.094
Med s.e.	0.199	0.093	0.097

	\widehat{lpha}_0	$\widehat{\alpha}_1$	\widehat{lpha}_2
N = 1000, T = 6	1.000	0.500	0.500
Median	1.016	0.503	0.504
mad	0.120	0.075	0.075
Med s.e.	0.132	0.076	0.076

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivoti

Occupancy as detectability

Presence absence

Full likelihood function — BOI

Edge solutio

Summary

wo-stage

omogeneous ca

sults

Numeric

maximisatio

terative m

Results

GAMs

ummary

ostage R ackage 50/61

Fitting the Full Likelihood with occu

- matrix of factors, Visit corresponding to the three visits
- list Obs that contains data frames of the time varying covariates

```
> Obs=list(time=as.data.frame(Model.out@T.ij[,c(1,5,9)]),
    temp=as.data.frame(Model.out@T.ij[,c(2,6,10)]),
    cloud=as.data.frame(Model.out@T.ij[,c(3,7,11)]),
    julian=as.data.frame(Model.out@T.ij[,c(4,8,12)]),
    Visit=as.data.frame(Visit))
```

- > 0.5.out=occu(~Visit+vegcov1+vegcov2+vegcov3+vegcov4+vegcov5+vegcov6 +time+temp+cloud+julian-1~vegcov1+vegcov2+vegcov3+vegcov4+vegcov5 +vegcov6,data=D,engine=c("C"))
- > 0.5.out@estimates

Figure: Fitting a model with occu for time varying covariates on the full model.

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Occupancy ai detectability

resencebsence ata

Full likelihood function — BOD

Edge solutions Exact mean an ariance

wo-stage

Homogeneous case

erogeneous c

naximisation WLS

fterative n

Results

GAMs Results

ummary

ostage R ackage **51/6**

Iterative and unmarked

Estimating occupancy and fitting models

Single simulated data for N = 150 and T = 4

• Estimates roughly the same

		β	s.e. β	t_{eta}	α	s.e. α	t_{α}
Iterative	Inter.	-0.273	0.177	-1.540	0.206	0.229	0.900
	x	0.442	0.169	2.620	-0.075	0.224	-0.335
	x.1	0.283	0.158	1.798	0.623	0.208	2.995
Unmarked	Inter.	-0.228	0.153	-1.489	0.198	0.224	0.882
	x	0.459	0.142	3.221	-0.081	0.221	-0.368
	x.1	0.085	0.126	0.673	0.710	0.210	3.380

IWLS and occu

Parameter

(a) Unstandardised

Estimate

- Default settings in occu the estimates did not converge but two-stage was fine
- "Nelder-Mead" method set to a maximum of 2000 iterations
- For the standardised data, occu with the default options did converge

Full Likelihood

Table: Occupancy and detection estimates for full likelihood and two-stage approaches for the (a) unstandardised and (b) standardised brook trout data. For each covariate, we report its: estimate (Estimate) standard error (so) Student's t statistic (t) and

Estimating	
occupancy and	
fitting models	

Two-stage

se

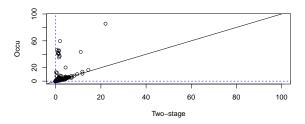
Summary

	or the two-stage approach estimated with IWLS method.	
,	· · ·	

Occupancy ψ

Intercept	-3.9716	0.6858	-5.7914	0.0000	-4.0452	1.1218	-3.6060	0.0003
Ele	0.0013	0.0003	4.5338	0.0000	0.0013	0.0004	3.6441	0.000vo-stage
			De	etection p				approach
Intercept	0.0580	0.7352	0.0788	0.9372	-0.1609	1.2397	-0.1298	0.8988mogeneous case
Ele	0.0004	0.0002	1.9697	0.0489	0.0004	0.0003	1.2516	0.2167 ^{sults}
CSA	-0.8325	0.2822	-2.9503	0.0032	-0.7438	0.2873	-2.5888	0.0096 terogeneous cas
(b) Standard	ised							Numeric
			Occ	cupancy ψ				maximisation
Intercept	-0.19	0.36	-0.52	0.60	-0.34	0.32	-1.04	0.30^{IWLS}
Ele	1.53	0.45	3.42	0.00	1.48	0.40	3.71	0.00 method
			De	etection p				Results
Intercept	-0.14	0.35	-0.38	0.70	-0.16	0.36	-0.44	$0.66^{ m AMs}$
Ele	0.36	0.35	1.04	0.30	0.43	0.37	1.18	$0.24^{ m sults}$
CSA	-0.82	0.28	-2.97	0.00	-0.80	0.28	-2.81	$0.90_{ m mmary}$

Estimate



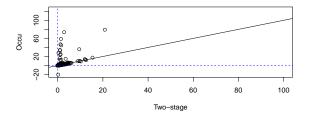


Figure: Comparison of estimated occupancy parameters $(\hat{\boldsymbol{\alpha}})$ between occu and two-stage (IWLS) for 1000 simulations with $\boldsymbol{\alpha}=(1,1), \, \boldsymbol{\beta}=(-1.5,-0.5,-0.5)$. Top figure shows intercept estimates and bottom figure estimates for the slope parameter.

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Motivat

Occupancy a detectability

Presence ibsence lata

Full likelihood function — BOI model

model

Exact mean

variance Summary

 Γ wo-stage

wo-stage oproach

nogeneous cas

sults torogonoous se

Numeric

naximisation

WLS terative metho

Results

GAMs

ummary

wostage R ackage 52/61

Agreement

Agreement for intercept estimates greater, or less, than three when the actual value to be estimated is $\alpha_1 = 1$.

• number of estimates that are either both or neither greater than three $(\hat{\alpha}_1 > 3)$, less than or equal to three $(\hat{\alpha}_1 \leqslant 3)$, or when these disagree • occu gives estimates $\hat{\alpha}_1 > 3$ that are large four times more often than IWLS (36 vs 12) • no universal best method for finding estimates for occupancy • if IWLS then try optim (or occu)

	occu method		
Two-stage IWLS	$\hat{\alpha}_1 \leqslant 3$	$\hat{\alpha}_1 > 3$	
$\hat{\alpha}_1 \leqslant 3$	832	36	
$\hat{\alpha}_1 > 3$	12	37	

Estimating occupancy and fitting models

Karavarsamis

Motivation

Motivation

Occupancy detectability

> Presence ibsence lata

Full likelihood function — BOD model

Exact mean a

ummary

wo-stage proach

mogeneous case

leterogeneous Numeric

lumeric naximisation

WLS

Results

AMs esults

ummary

ostage R ackage **52/6**:

Outline

Two-stage approach

GAMs

Estimating occupancy and fitting models

Stage 1: \hat{p}_i

 \triangleright GAMS to redetections r_i

Stage 2: $\widehat{\psi}_i$

 \triangleright GAMS to presence–absences w_i via iterative method with offset

$$a_i^{(s)} = \log(\hat{\theta}_i) - \log\{1 + \exp(\nu_i^{(s-1)})(1 - \hat{\theta}_i)\}$$

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivatior

Motivation

Occupancy a detectability

resence ibsence lata

full likelihood function — BOD model

Edge solutions Exact mean and

Summary

Two-stage approach

Homogeneous case Results

Results Heterogeneous case

Numeric

maximisatio IWLS

erative me

GAMs

esults

ostage R

GAMs - variances

$$\widehat{\operatorname{Var}}(\widehat{\alpha}_{\lambda}(\widehat{\beta})) \approx I_{\lambda}\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\}^{-1} I\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\} I_{\lambda}\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\}^{-1} + I_{\lambda}\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\}^{-1} \widetilde{B}\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})\} \widehat{V}_{\beta}^{*} \widetilde{B}\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\}^{T} I_{\lambda}\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\}^{-1},$$

gives

$$\widehat{\operatorname{Var}}(\widehat{\psi}_s^{(\lambda)}) = \{\widehat{\psi}_s(1 - \widehat{\psi}_s)\}^2 x_s^T \widehat{\operatorname{Var}}\{\widehat{\alpha}_{\lambda}(\widehat{\beta})\} x_s$$

where $I_{\lambda}(\alpha, \widehat{\beta}) = I(\alpha, \widehat{\beta}) + \lambda_S \mathcal{P}^*$.

- V_{β}^* covariance for $\widehat{\beta}$
- $Q(\alpha, \beta) = \partial l(\alpha, \beta) / \partial \alpha$
- $I(\alpha, \widehat{\beta}) = -\partial Q(\alpha, \beta)/\partial \alpha$
- $\widetilde{B}\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\} = \partial Q(\alpha, \beta)/\beta$
- λ_S smooth
- P* penalty matrix

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivatio

Motivatio

Occupancy detectabilit

> Presence ibsence lata

Full likelihood function — BOD model

Exact mean variance

Summary

pproach Homogeneous cas

> suits terogeneous c

Numeric maximisation

IWLS

GAMs

Results

ımmary

Outline

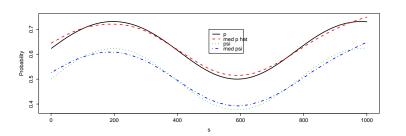
75 /5			

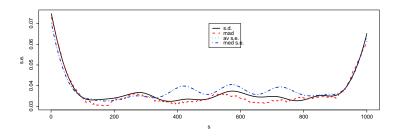
Two-stage approach

Results

Estimating occupancy and fitting models

GAMs – simulations





Estimating occupancy and fitting models

Natalie Karavarsami

Motivatio

Motive

Occupancy a

resencebsence ata

Full likelihood function — BOD model

Edge solutions

variance Summary

Two-stage

omogeneous ca

tesults Heterogeneous ca:

Numeric maximisation

maximisation IWLS

IWLS Iterative me

AMs

esults

ımmary

ostage R ickage **57/61**

Results - GAMs

	beta	se.beta	t.beta	alpha	se.alpha	t.alpha
(Intercept)	-2.700	11.350	-0.238	147.663	1242.713	0.119
ele	1.896	31.177	0.061	319.120	_	_
forest	0.673	0.650	1.035	-0.719	19.815	-0.036
s(ele).1	-18.100	30.030	-0.603	-102.503	2116.133	-0.048
s(ele).2	2.749	58.341	0.047	-472.991	1207.377	-0.392
s(ele).3	21.185	64.824	0.327	-563.946	2289.131	-0.246
s(ele).4	12.077	47.065	0.257	-603.291	3555.318	-0.170
s(ele).5	15.019	27.374	0.549	-725.366	2411.729	-0.301
s(ele).6	6.219	29.648	0.210	-689.764	4113.145	-0.168
s(ele).7	29.652	35.357	0.839	-775.511	1446.579	-0.536
s(ele).8	-14.317	26.192	-0.547	-768.387	2917.795	-0.263
s(ele).9	-49.528	73.198	-0.677	2314.131	10896.921	0.212
int	-1.742	0.238	-7.311	3.950	1.931	2.045
ele	0.829	0.392	2.112	2.141	0.862	2.484
ele.sq	0.745	0.445	1.672	-3.549	1.533	-2.316
forest	0.131	0.187	0.702	0.609	0.737	0.826
	ele forest s(ele).1 s(ele).2 s(ele).3 s(ele).3 s(ele).5 s(ele).6 s(ele).6 int ele).9 int ele ele.sq	(Intercept) -2.700 ele 1.896 forest 0.673 s(ele).1 -18.100 s(ele).2 2.749 s(ele).3 21.185 s(ele).4 12.077 s(ele).5 15.019 s(ele).6 6.219 s(ele).6 6.219 s(ele).8 -14.317 s(ele).9 -49.528 int -1.742 ele 0.829 ele.sq 0.745	$ \begin{array}{c cccc} (Intercept) & -2.700 & 11.350 \\ ele & 1.896 & 31.177 \\ forest & 0.673 & 0.650 \\ s(ele).1 & -18.100 & 30.030 \\ s(ele).2 & 2.749 & 58.341 \\ s(ele).3 & 21.185 & 64.824 \\ s(ele).4 & 12.077 & 47.065 \\ s(ele).5 & 15.019 & 27.374 \\ s(ele).6 & 6.219 & 29.648 \\ s(ele).7 & 29.652 & 35.357 \\ s(ele).7 & 29.652 & 35.357 \\ s(ele).8 & -14.317 & 26.192 \\ s(ele).9 & -49.528 & 73.198 \\ \hline int & -1.742 & 0.238 \\ ele & 0.829 & 0.392 \\ ele.sq & 0.745 & 0.445 \\ \hline \end{array} $	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$

Estimating occupancy and fitting models

Conclusions

- Occupancy models appear simple but are harder than expected
 - We resolved problems with construction of estimators and interval estimators
- ► Full likelihood possible but does not allow easy access to GLM machinery
- ▶ Welsh et al. (2013) show that problems with the full likelihood feed through to the covariate model
 - Partial likelihood allows full access to GLM machinery at both stages
 - Estimators from both stages are probabilities (so naturally constrained to between 0 and 1)

Estimating occupancy and fitting models

Natalie Karavarsamis

Motivation

Occupancy an

Presence absence lata

Full likelihood function — BOD

Edge solutions Exact mean as variance

wo-stage

omogeneous case

leterogeneous ca

Numeric maximisation

maximisation IWLS

terative met

GAMs

Summary

twostage R package 59/61

twostage R package

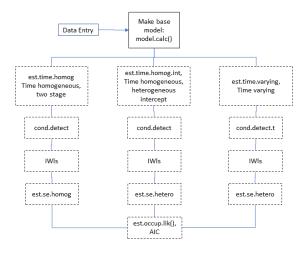


Figure: Flowchart of twostage algorithm

Estimating occupancy and fitting models

Natalie Karavarsami

Motivatio

Motiv

Occupancy a

resence ibsence lata

Full likelihood function — BOI model

Exact mean variance

Γwo-stage

mogeneous cas sults

Heterogeneou Numeric

maximisatio

IWLS

Results

esults

.....

twostage R package 60/61

Thank you

About me and contact info at

https://natalie-karavarsamis.github.io

Estimating occupancy and fitting models

twostage R package 61 / 61

References

- Heard, G., Robertson, P., and Scroggie, M. (2006). Assessing detection probabilities for the endangeared growling grass frog in Southern Victoria. Wildlife Research, 33:557–564.
- Karavarsamis, N. and Huggins, R. M. (2019a). A two-stage approach to the analysis of occupancy data II. The Heterogeneous case. Computational Statistics and Data Analysis, 133:195–207.

Karavarsamis, N. and Huggins, R. M. (2019b). Two-stage

- approaches to the analysis of occupancy data I: The homogeneous case. *Communications in Statistics Theory and Methods*, page DOI: 10.1080/03610926.2019.1607385.

 Karavarsamis, N., Robinson, A. P., Hepworth, G., Hamilton, A., and
- Heard, G. (2013). Comparison of four bootstrap-based interval estimators of species occupancy and detection probabilities.

 Australian and New Zealand Journal of Statistics, 55(3):235–252.
- Karavarsamis, N. and Watson, R. Bias of occupancy estimator. (In preparation).
- Karavarsamis, N. and Watson, R. Plausible region for estimating occupancy. (In preparation).
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J., and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8):2248–2255.

Estimating occupancy and fitting models

Natalie Karayarsamis

Motivatio

occupancy and

resence bsence

Full likelihood function — BOD

Edge solutions Exact mean ar variance

> vo-stage proach

omogeneous cas

terogeneous (umeric

Numeric maximisation

IWLS Iterative n

esults

GAMs Results

ummary

ostage R ackage 61/6 Welsh, A. H., Lindenmayer, D. B., and Donnelly, C. F. (2013). Fitting and interpreting occupancy models. *PLoS ONE*, 8(1):e52015. doi:10.1371/journal.pone.0052015.s001.

Estimating occupancy and fitting models

Natalie

Motivation

Motivation

etectability

resencebsence .ata

> ll likelihood action — BO adel

lge solutions

act mean and riance mmary

o-stage oroach mogeneous case

mogeneous case sults serogeneous case

erogeneous imeric aximisation

terative n esults AMs

Results

age R.

stage R ckage **61/6**1