Estimating occupancy and fitting models

Natalie Karavarsamis

Department of Mathematics and Statistics
La Trobe University

RMIT, October 3, 2019

Outline

Outline

Motivation

Why monitor populations?

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

- endangered species - e.g. growling grass frog

Outline

Objective and challenge

Probability of presence

ψ

Probability of detection

p

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.

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 $\leftarrow x_{ij}$	1	1	1	
9	0	0	1	1	
10	1	1	1	1	
11	0	1	0	0	
12	0	1	0	0	$x_{i\cdot} \neq 0$
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	

Table: Capture histories for the growling grass frog. The 27 independent sites each were surveyed on 4 occasions at night within the 2002-2003 season (Heard et al., 2006).

The basic occupancy model (BOD) – ZIB

- The detection (p) and occupancy (ψ) probabilities remain constant over all N sites and T survey occasions.

The number of detections $X_{i\cdot}$ at site i is distributed as

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

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

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

Full likelihood (ZIB) under constant ψ & p

(MacKenzie et al., 2002)

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

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

nondetections detections

- number of detected sites

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

- total number of detections

  $x = \sum_{i=1}^{N} x_{i\cdot}$

k and x sufficient for ψ, p

Score equations – Full likelihood under constant ψ & p

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

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

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

BUT do not always give MLEs !!!

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

If N, T small then these will not apply.

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 – $\hat{\psi} > 1, \hat{p} > 1$

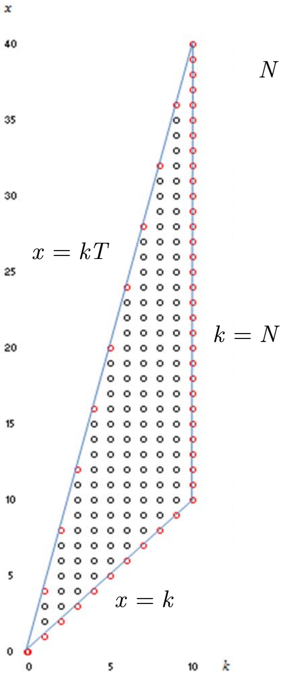
3. Standard errors – no closed form solutions for Var_{Mac}, hessian not always available

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

Outline

2. Boundary problem

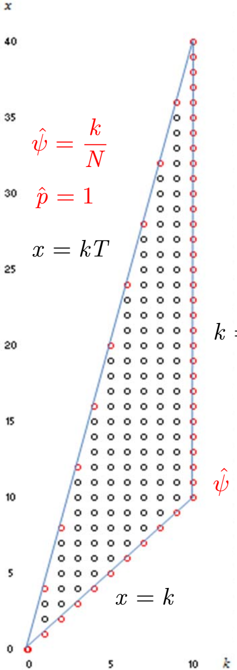
$N = 10, T = 4$



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

2. Boundary problem

'Edge' solutions



$N = 10,\ T = 4$

$\hat{\psi} = \dfrac{k}{N}$

$\hat{p} = 1$

$x = kT$

$k = N$

$\hat{\psi} = 1,\ \hat{p} = \dfrac{x}{NT}$

$x = k$

MLE is 1 but
soln to score eqn > 1

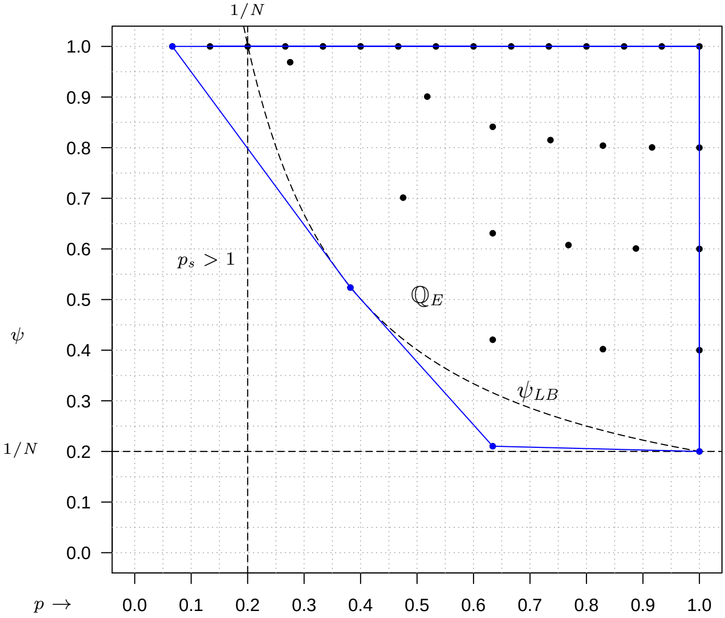
$\psi_s = k/N\theta_s$
$p_s = x\theta_s/kT$

Limitations – Full likelihood (BOD)

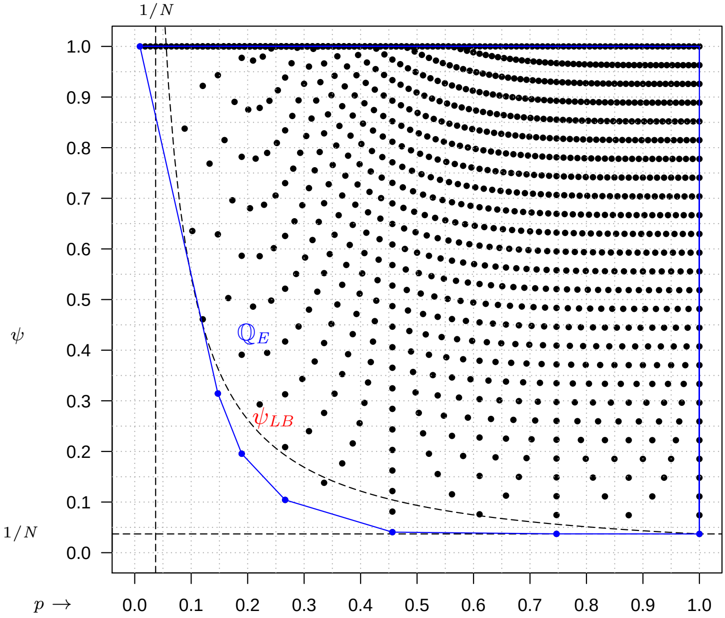
This caused

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

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



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



Outline

Problem 3. Standard errors – exact mean and variance

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

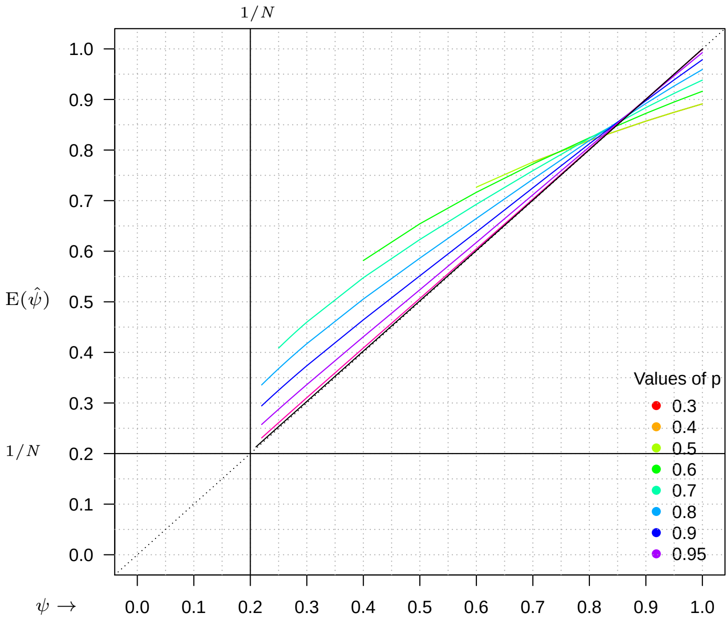
This allowed us to evaluate the

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

Results

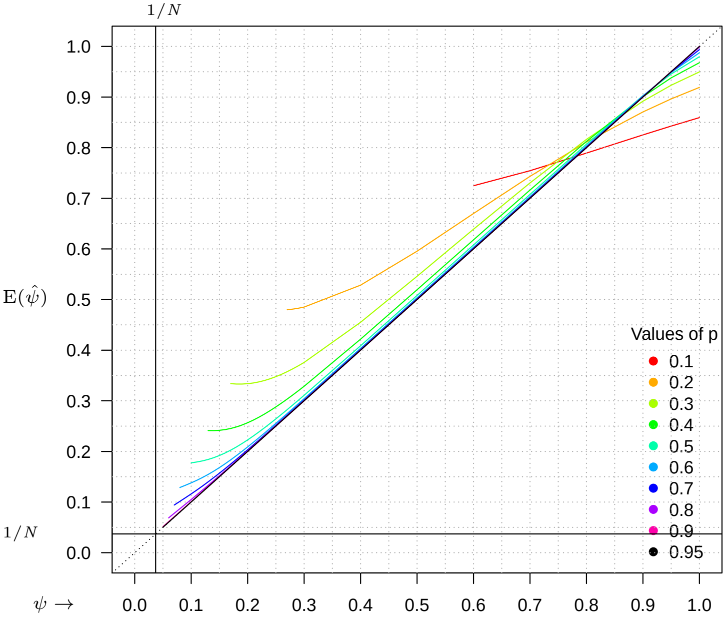
- bias–corrections for $\hat{\psi}$ not so effective for small N, T, or p because not enough information in (x, k)
- asymptotic variance underestimates actual variance

Bias – $N = 5, T = 3$



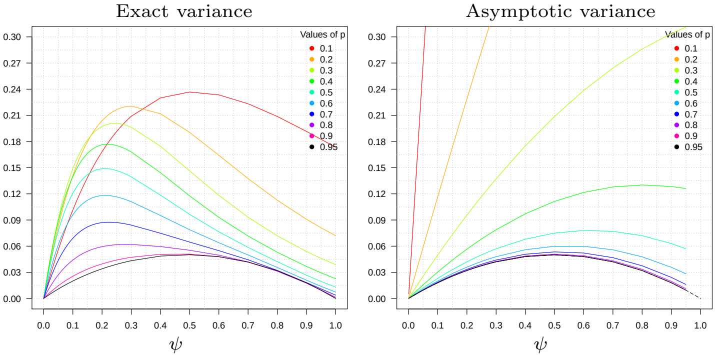
1/N
1.0
0.9
0.8
0.7
0.6
E($\hat{\psi}$)
0.5
0.4
0.3
1/N
0.2
0.1
0.0
$\psi \rightarrow$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Values of p
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.95

Bias – $N = 27$, $T = 4$



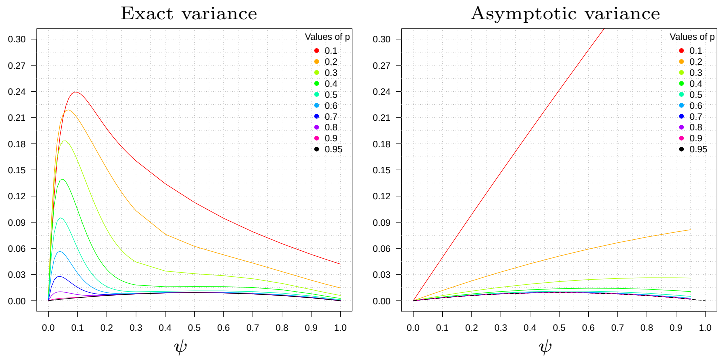
1/N
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
E(ψ̂)
1/N
ψ →
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Values of p
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.95

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



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

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



Outline

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...

Outline

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...

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

Partial likelihoods – benefits

What we want:

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

How to achieve goals:

- 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 1^{st} 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
- 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

Homogenous case – ψ and p constant

Partial likelihood

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

$$= (1-\psi\theta)^{N-k}\psi^k \boxed{\prod_{i=1}^{k}(1-p)^{a_i-1}p} \; \boxed{p^{x-k}(1-p)^{b-(x-k)}}$$

$$= L_1(\psi, p)\; L_2(p)$$

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

Now estimate ψ and p SEPARATELY !

Homogenous case – ψ and p constant

Partial likelihood

$$\mathrm{L}(\psi, p) \propto (1-\psi+\psi(1-p)^T)^{N-k} \prod_{i=1}^{k} \psi \; p^{x_i} \cdot (1-p)^{T-x_i} \cdot$$

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

$$= \mathrm{L}_1(\psi, p) \times \mathrm{L}_2(p)$$

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

Now estimate ψ and p SEPARATELY !

Score equations — homogeneous partial likelihood

Stage 1: $L_2(p)$ gives

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

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

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

and

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

Now we have closed form solutions, yipee!!!

Outline

Full versus partial likelihood - homogeneous case

Simulations – small p, large N
Examine 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		

Full versus partial likelihood - homogeneous case

Simulations for $N = 27$, $T = 4$.

	Partial		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. $= 0.098$ $\hat{\psi}_{Full} = 0.557$, s.e. $= 0.096$
- $\hat{p}_{Part} = 0.889$, s.e. $= 0.052$ $\hat{p}_{Full} = 0.780$, s.e. $= 0.054$

Summary

- Equations are straight forward and can be used directly for estimation (R code). Full likelihood uses optim.
- 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

Outline

Site inhomogeneity

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

Contribution of a single site is

$$L_i(\psi_i, p_i) = \left\{1 - \psi_i + \psi_i(1-p_i)^{\tau}\right\}^{1-w_i} \left\{\binom{\tau}{x_{i\cdot}} \psi_i \, p_i^{x_{i\cdot}} (1-p_i)^{\tau - x_{i\cdot}}\right\}^{w_i}$$

$$\propto (1-\psi_i\theta_i)^{1-w_i} \psi_i^{w_i} \left\{ p_i (1-p_i)^{(a_i - 1)} \right\}^{w_i} \left\{ p_i^{(x_{i\cdot}-1)} (1-p_i)^{b_i - x_{i\cdot}+1} \right\}^{w_i}$$

$$= L_{1i}(\psi_i, p_i) \; L_{2i}(p_i)$$

Site inhomogeneity

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

Contribution of a single site is

$$L_i(\psi_i, p_i) = \left\{1 - \psi_i + \psi_i(1-p_i)^\tau\right\}^{1-w_i} \left\{\binom{\tau}{x_{i\cdot}} \psi_i \, p_i^{x_{i\cdot}} (1-p_i)^{\tau - x_{i\cdot}}\right\}^{w_i}$$

$$\propto (1-\psi_i\theta_i)^{1-w_i}\psi_s^{w_i} \times \left\{ \boxed{p_i^{(x_{i\cdot}-1)}(1-p_i)^{b_i - x_{i\cdot}+1}} \right\}^{w_i}$$

$$= L_{1i}(\psi_i, p_i) \times L_{2i}(p_i)$$

...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)

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))

Conditional likelihood for detection

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

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

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

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

Replace η_s by $\widetilde{\eta}_s = \psi_s\widehat{\theta}_s$ in the log-partial likelihood (1) and maximise this to estimate $\boldsymbol{\alpha}$.

($z_s = 1 - w_s$ indicator of no detections)

Stage 1: Estimate detection p and coefficients

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

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.
• $\mathrm{p}(\mathrm{u}_s, \boldsymbol{\beta}) = h(u_s^T \boldsymbol{\beta})$, $j = 1, \ldots, \tau$, vector of covariates 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 $\widehat{V}_\beta$ for $\hat{\beta}$
  - $\hat{p}$
  - $\mathrm{Var}(\hat{p}) = \widehat{\beta}^T \widehat{V_\beta} \widehat{\beta}$

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,
- $\tau (= 3)$ is the number of visits to each site s, $(S = 1, \ldots 656$ 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

Stage 1: Time Dependent Covariates for Detection

$$L_2(\boldsymbol{p}_s) = \frac{\prod_{j=1}^{\tau} p_{sj}^{y_{sj}} (1 - p_{sj})^{1-y_{sj}}}{\theta_s}.$$

- 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
- 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`

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) ~
vegcov1+vegcov2+vegcov3+vegcov4+vegcov5+vegcov6,
family=posbernoulli.t(parallel.t=FALSE ~ 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
```

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(survey1,survey2,survey3)
  ~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`.

Stage 2: Estimate occupancy and coefficients

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

Stage 2: Numeric maximisation

To estimate $\boldsymbol{\alpha}$ maximise the partial likelihood

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

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

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

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.

Stage 2: IWLS

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

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

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

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

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 $\widetilde{I}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = XUV^{-1}UX^T$.

Derivations in Karavarsamis & Huggins (2015), CSDA

Method 3: Iterative method

Under the logistic model

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

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

If

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

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

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

$\therefore$ 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

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

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

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

gives

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

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

Outline

Method 3: Iterative method - Results

1000 simulations for large N and small T with 2 covariates

- Relatively unbiased
- s.e. reasonable

	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$
$N = 1000,\ T = 4$	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

	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$
$N = 1000,\ T = 6$	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

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

```
> Visit=matrix(as.factor(c(rep("a",656),rep("b",656),rep("c",656))),
   ncol=3)
> 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))
> D=unmarkedFrameOccu(y=Model.out@Detect,
    siteCovs=as.data.frame(Model.out@X[,-1]),obsCovs=Obs)
> 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.

Iterative and unmarked

Single simulated data for $N = 150$ and $T = 4$

- Estimates roughly the same

		β	s.e. β	t_β	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

- 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

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 (se), Student's t-statistic (t), and p-value (p). Occupancy for the two-stage approach estimated with IWLS method.

	Full Likelihood				Two-stage			
Parameter	Estimate	se	t	p	Estimate	se	t	p
(a) Unstandardised								
				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	0.0003
				Detection p				
Intercept	0.0580	0.7352	0.0788	0.9372	-0.1609	1.2397	-0.1298	0.8968
Ele	0.0004	0.0002	1.9697	0.0489	0.0004	0.0003	1.2516	0.2107
CSA	-0.8325	0.2822	-2.9503	0.0032	-0.7438	0.2873	-2.5888	0.0096
(b) Standardised								
				Occupancy ψ				
Intercept	-0.19	0.36	-0.52	0.60	-0.34	0.32	-1.04	0.30
Ele	1.53	0.45	3.42	0.00	1.48	0.40	3.71	0.00
				Detection p				
Intercept	-0.14	0.35	-0.38	0.70	-0.16	0.36	-0.44	0.66
Ele	0.36	0.35	1.04	0.30	0.43	0.37	1.18	0.24
CSA	-0.82	0.28	-2.97	0.00	-0.80	0.28	-2.81	0.00

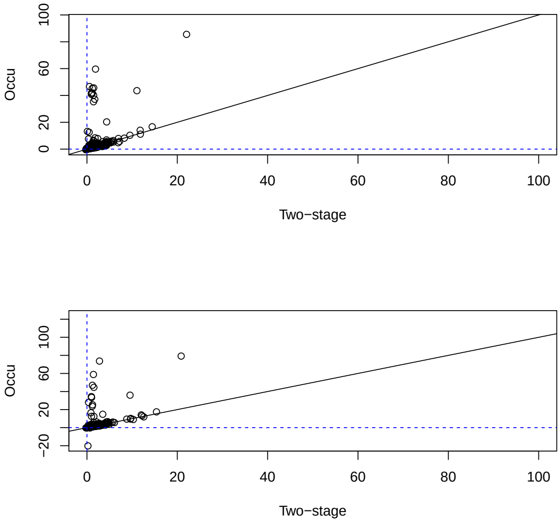


Figure: Comparison of estimated occupancy parameters ($\hat{\boldsymbol{\alpha}}$) between occu and two-stage (IWLS) for 1000 simulations with $\boldsymbol{\alpha} = (1, 1)$, $\boldsymbol{\beta} = (-1.5, -0.5, -0.5)$. Top figure shows intercept estimates and bottom figure estimates for the slope parameter.

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

Outline

GAMS

$$p_i : \quad \kappa_i = u_i^T \alpha + f_1(v_{i1}) + \cdots + f_K(v_{iK}) \quad \begin{cases} u_i &= \text{parametric,} \\ v_i &= \text{nonparametric} \end{cases}$$

$$\psi_i : \quad \nu_i = x_i^T \alpha + g_1(r_{i1}) + \cdots + g_J(r_{iJ}) \quad \begin{cases} x_i &= \text{parametric,} \\ g_i &= \text{nonparametric} \end{cases}$$

Stage 1: $\hat{p}_i$

- GAMS to redetections r_i

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

- GAMS to presence–absences w_i via iterative method with offset

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

GAMs – variances

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

gives

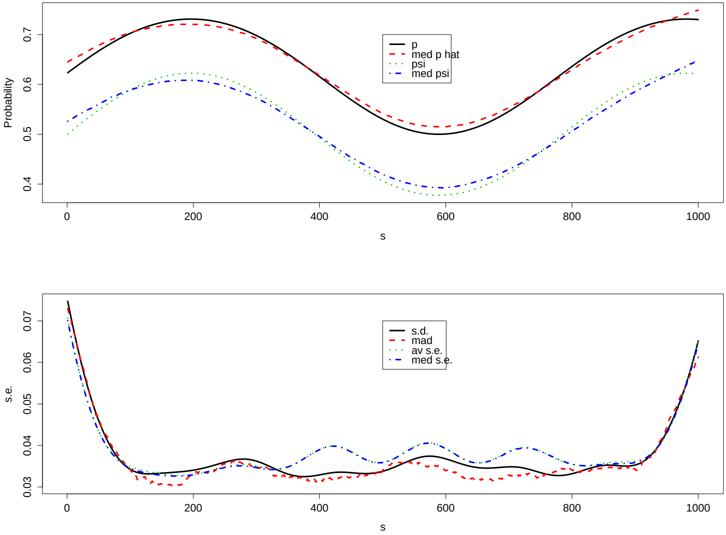
$$\widehat{\text{Var}}(\widehat{\psi}_s^{(\lambda)}) = \{\widehat{\psi}_s(1-\widehat{\psi}_s)\}^2 x_s^T \widehat{\text{Var}}\{\widehat{\alpha}_\lambda(\widehat{\beta})\} x_s$$

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
- $\mathcal{P}^*$ penalty matrix

Outline

GAMs – simulations



Results - GAMs

		beta	se.beta	t.beta	alpha	se.alpha	t.alpha
Iter.	(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
occu	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

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)

twostage R package

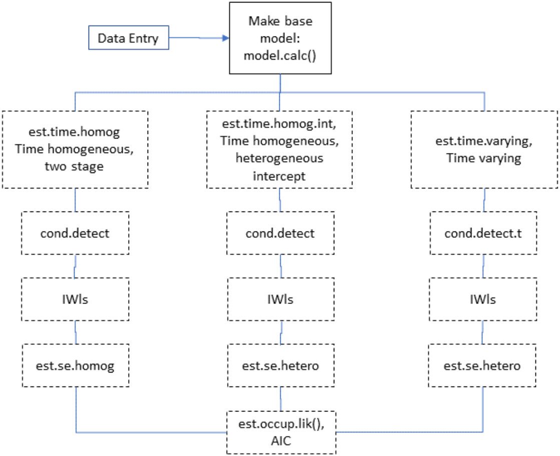


Figure: Flowchart of twostage algorithm

Thank you

About me and contact info at

https://natalie-karavarsamis.github.io

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.

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.