Diagnostic tests and model assessment for capture-recapture data

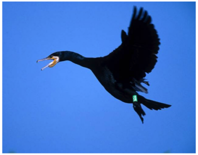
Rachel McCrea
Royal Statistical Society
London, February 2016
@UniKentSEaK www.capturerecapture.co.uk

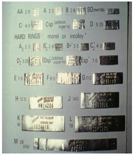
Collaborators

- Anita Jeyam
- Roger Pradel
- Olivier Gimenez
- Byron Morgan

Outline

Individual marking





Capture-recapture data

- 1 0 0 1 0
- 1 1 0 1 1
- 0 0 1 0 1
- ...

Cormack-Jolly-Seber (CJS) model

- ϕ_t: probability an individual survives from time t to $t+1$
- p_t: probability an individual alive at time t is captured at time t

Cormack-Jolly-Seber (CJS) model

- ϕ_t: probability an individual survives from time t to $t+1$
- p_t: probability an individual alive at time t is captured at time t

- `1 0 0 1 0` $\phi_1(1-p_2)\phi_2(1-p_3)\phi_3 p_4 \chi_4$
- `1 1 0 1 1` $\phi_1(p_2)\phi_2(1-p_3)\phi_3 p_4 \phi_4 p_5$
- `0 0 1 0 1` $\phi_3(1-p_4)\phi_4 p_5$
- ...

where $\chi_t = (1-\phi_t) + \phi_t(1-p_{t+1})\chi_{t+1}$ and $\chi_T = 1$.

Sufficient statistics

- R_i: number of individuals released at occasion i
- m_{ij}: number of individuals released at occasion i and next recaptured at occasion j

Sufficient statistics

- R_i: number of individuals released at occasion i
- m_{ij}: number of individuals released at occasion i and next recaptured at occasion j

Number of releases	Recapture occasion 2	3	4	5	Never seen again
R_1	m_{12}	m_{13}	m_{14}	m_{15}	$R_1 - \sum_{j=2}^{5} m_{1j}$
R_2		m_{23}	m_{24}	m_{25}	$R_2 - \sum_{j=3}^{5} m_{2j}$
R_3			m_{34}	m_{35}	$R_3 - \sum_{j=4}^{5} m_{3j}$
R_4				m_{45}	$R_4 - m_{45}$

Sufficient statistics

Number of releases	Recapture occasion 2	3	4	5	Never seen again
R_1	m_{12}	m_{13}	m_{14}	m_{15}	$R_1 - \sum_{j=2}^{5} m_{1j}$
R_2		m_{23}	m_{24}	m_{25}	$R_2 - \sum_{j=3}^{5} m_{2j}$
R_3			m_{34}	m_{35}	$R_3 - \sum_{j=4}^{5} m_{3j}$
R_4				m_{45}	$R_4 - m_{45}$

e.g. 1 1 0 1 1

Cormack-Jolly-Seber (CJS) model

Number of releases	Recapture occasion 2	3	4	5	Never seen again
R_1	$\phi_1 p_2$	$\phi_1 \bar{p}_2 \phi_2 p_3$	$\phi_1 \bar{p}_2 \phi_2 \bar{p}_3 \phi_3 p_4$	$\vdots$	χ_1
R_2		$\phi_2 p_3$	$\phi_2 \bar{p}_3 \phi_3 p_4$	$\vdots$	χ_2
R_3			$\phi_3 p_4$	$\vdots$	χ_3
R_4				$\vdots$	χ_4

where $\bar{p}_t = 1 - p_t$.

Diagnostic goodness-of-fit tests

Factorisation Theorem

$$L = \Pr(\text{data}|s) \times \Pr(s;\theta)$$

s: sufficient statistics, θ: model parameters

- $\Pr(\text{data}|s)$ can be used to check model assumptions
- Pollock et al (1985) showed that the assessment of model adequacy can be decomposed into Test 2 and Test 3.

Diagnostic goodness-of-fit tests

- Test 2: captured and non-captured individuals have an equal chance of being captured at the next capture occasion:
  - 2.CT (trap-dependence)
  - 2.CL
- Test 3: newly and already-marked animals have an equal chance of being seen again:
  - 3.SR (transience)
  - 3.Sm
- All tests are performed by means of a contingency table test of homogeneity.

Test 2.CT(2)

Looks for differences between individuals captured and not captured at occasion 2

Number of releases	Recapture occasion 2	3	4	5	Never seen again
R_1	m_{12}	m_{13}	m_{14}	m_{15}	$R_1 - \sum_{j=2}^{5} m_{1j}$
R_2		m_{23}	m_{24}	m_{25}	$R_2 - \sum_{j=3}^{5} m_{2j}$
R_3			m_{34}	m_{35}	$R_3 - \sum_{j=4}^{5} m_{3j}$
R_4				m_{45}	$R_4 - m_{45}$

m_{13}	$m_{14}+m_{15}$
m_{23}	$m_{24}+m_{25}$

Test 2.CT(3)

Looks for differences between individuals captured and not captured at occasion 3

Number of releases	Recapture occasion 2	3	4	5	Never seen again
R_1	m_{12}	m_{13}	m_{14}	m_{15}	$R_1 - \sum_{j=2}^{5} m_{1j}$
R_2		m_{23}	m_{24}	m_{25}	$R_2 - \sum_{j=3}^{5} m_{2j}$
R_3			m_{34}	m_{35}	$R_3 - \sum_{j=4}^{5} m_{3j}$
R_4				m_{45}	$R_4 - m_{45}$

$m_{14} + m_{24}$	$m_{15} + m_{25}$
m_{34}	m_{35}

Test 2.CT(3)

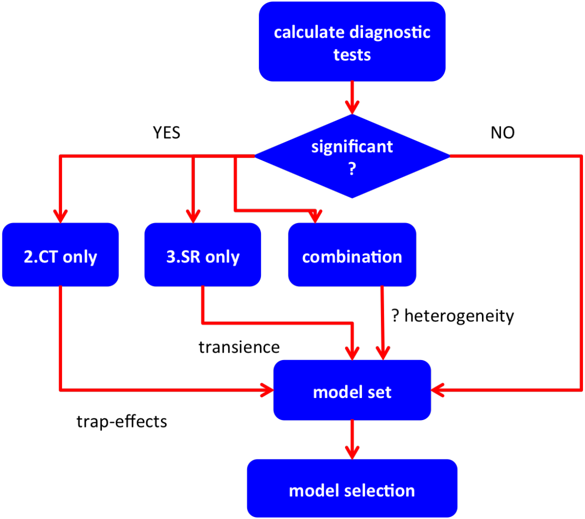
Looks for differences between individuals captured and not captured at occasion 3

Number of releases	Recapture occasion 2	3	4	5	Never seen again
R_1	m_{12}	m_{13}	m_{14}	m_{15}	$R_1 - \sum_{j=2}^{5} m_{1j}$
R_2		m_{23}	m_{24}	m_{25}	$R_2 - \sum_{j=3}^{5} m_{2j}$
R_3			m_{34}	m_{35}	$R_3 - \sum_{j=4}^{5} m_{3j}$
R_4				m_{45}	$R_4 - m_{45}$

$m_{14} + m_{24}$	$m_{15}+m_{25}$
m_{34}	m_{35}

Test 2.CT = Test 2.CT(2) + Test 2.CT(3) + $\cdots$ + Test 2.CT(T-2)

Current approach

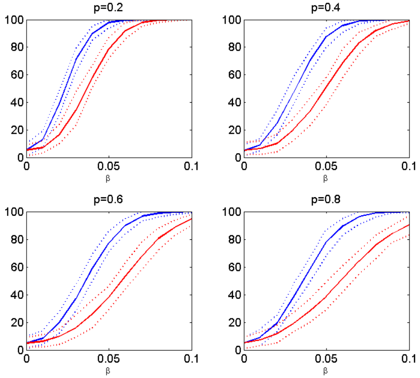


Equivalence to score tests

- Shown that certain diagnostic goodness-of-fit tests are equivalent to score tests (McCrea et al, 2015).
- Test 2.CT
  - p_t^*: probability of capture at occasion t given the individual was captured at occasion $t-1$
  - Score test of $H_0 : p_t = p_t^*$.
- Test 3.SR
  - ϕ_t^*: probability an individual captured at occasion t survives until $t+1$
  - Score test of $H_0 : \phi_t = \phi_t^*$.

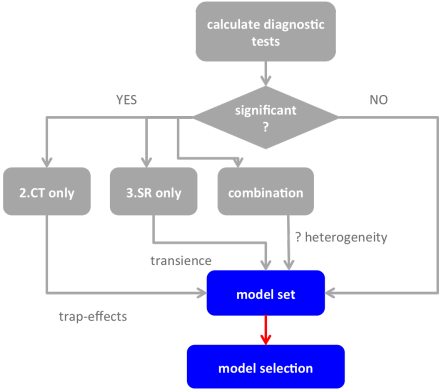
Power to detect trap-dependence

- Suppose $p^* = p + \beta$
- Score test $H_0 : p^* = p$
- Test 2.CT (equivalent to $H_0 : p_t^* = p_t$)



Alternative strategy

- Are diagnostic goodness-of-fit tests necessary?
- Strategy of including alternative models within a model set and performing step-wise model selection procedure using score tests (McCrea and Morgan, 2011)



Humpback whale application

Model Code	Model	k	s	P
A0	$\phi(\cdot), p(\cdot)$	2		
A1	$\phi(t), p(\cdot)$	7	4.78	0.44
A2	$\phi(\mathbf{trans}), \mathbf{p}(\cdot)$	3	5.80	0.02
A3	$\phi(\cdot), p(t)$	7	2.22	0.82
A4	$\phi(\cdot), p(trap)$	3	1.59	0.21
B0	$\phi(trans), p(\cdot)$	3		
B1	$\phi(trans * t), p(\cdot)$	12	12.53	0.19
B2	$\phi(trans), p(t)$	8	3.69	0.59
B3	$\phi(trans), p(trap)$	4	0.81	0.37

Humpback whale application

Model Code	Model	k	s	P
A0	$\phi(\cdot), p(\cdot)$	2		
A1	$\phi(t), p(\cdot)$	7	4.78	0.44
A2	$\phi(\mathbf{trans}), \mathbf{p}(\cdot)$	3	5.80	0.02
A3	$\phi(\cdot), p(t)$	7	2.22	0.82
A4	$\phi(\cdot), p(trap)$	3	1.59	0.21
B0	$\phi(trans), p(\cdot)$	3		
B1	$\phi(trans * t), p(\cdot)$	12	12.53	0.19
B2	$\phi(trans), p(t)$	8	3.69	0.59
B3	$\phi(trans), p(trap)$	4	0.81	0.37

- Transience is an important biological feature in this population

Capture heterogeneity

- Not all individuals in a study will have the same probability of capture
- Failing to account for capture heterogeneity can bias other model parameters

Anthill population size

Ayre (1962) estimated an anthill population to be 109 when there were known to be more than 3,000 ants in it.

Capture heterogeneity

- Not all individuals in a study will have the same probability of capture
- Failing to account for capture heterogeneity can bias other model parameters

Anthill population size

Ayre (1962) estimated an anthill population to be 109 when there were known to be more than 3,000 ants in it.

- Existing diagnostic goodness-of-fit tests lack the power to detect capture heterogeneity
- Aim: develop more powerful method to detect capture heterogeneity without model fitting.

Test of positive association

- If some animals have a higher capture probability than others, they will be seen more often.
- At a given capture occasion, animals with a high number of previous encounters will probably have a high number of future encounters.

Test of positive association

ID99	1	0	0	1	1	1	1	1	0	0
ID100	0	0	1	0	0	0	1	0	1	0

Test of positive association

ID99	1	0	0	1	1	1	1	1	0	0
ID100	0	0	1	0	0	0	1	0	1	0

- Condition on first release and known to be alive

Test of positive association

⇓

ID99	1	0	0	1	1	1	1	1	0	0
ID100	0	0	1	0	0	0	1	0	1	0

- Calculate the proportion of previous and future captures from occasion $i(=5)$

Test of positive association

					⇓					
ID99	1	0	0	1	1	1	1	1	0	0
ID100	0	0	1	0	0	0	1	0	1	0

- Calculate the proportion of previous and future captures from occasion $i(=5)$
- Previous encounters for ID99: $2/4$
- Future encounters for ID99: $2/2$

Test of positive association

⇓

ID99	1	0	0	1	1	1	1	1	0	0
ID100	0	0	1	0	0	0	1	0	1	0

- Calculate the proportion of previous and future captures from occasion $i(=5)$
- Previous encounters for ID99: $2/4$
- Future encounters for ID99: $2/2$
- Previous encounters for ID100: $0/2$
- Future encounters for ID100: $1/3$

Test of positive association

⇓

ID99	1	0	0	1	1	1	1	1	0	0
ID100	0	0	1	0	0	0	1	0	1	0

- Rank the proportions:

	Previous		Future	
ID99	2/4	2	2/2	2
ID100	0/2	1	1/3	1

Goodman-Kruskal's gamma

- This statistic is based on the pairs of discordant and concordant observations: $\gamma = \frac{C-D}{C+D}$.
- A pair of individuals is concordant if the observation ranking higher (lower) for the previous encounters, also ranks higher (lower) for the future encounters.
- In the example animal ID99 is ranked higher than animal ID100 for the previous and future encounters, and therefore form a concordant pair.

Goodman-Kruskal's gamma

- This statistic is based on the pairs of discordant and concordant observations: $\gamma = \frac{C-D}{C+D}$.
- A pair of individuals is concordant if the observation ranking higher (lower) for the previous encounters, also ranks higher (lower) for the future encounters.
- In the example animal ID99 is ranked higher than animal ID100 for the previous and future encounters, and therefore form a concordant pair.
- In the case of heterogeneity in capture, we expect a high number of concordant pairs.
- One-sided test for $\gamma > 0$.

Power to detect capture heterogeneity

Table: Percentage of significant tests (250 simulation runs) for 2000 individuals

Simulation scenario	Occasion 3	4	5	6	7
No heterogeneity	0.8	0	0	0	0.4
Binomial mixture	86	99	100	100	98.4
Survival heterogeneity	1.2	0.4	0.4	0.4	0.4
Trap-happiness	3.2	1.6	4	5.2	8

Power to detect capture heterogeneity

Table: Percentage of significant tests (250 simulation runs) for 500 individuals

Simulation scenario	Occasion 3	4	5	6	7
No heterogeneity	1.2	1.2	0.8	0.4	0.8
Binomial mixture	36	48	66	59	44
Survival heterogeneity	0	2.82	1.2	0.4	1.2
Trap-happiness	1.5	0.4	0.8	3.2	1.2

Power to detect capture heterogeneity

Table: Percentage of significant tests (250 simulation runs) for 500 individuals

Simulation scenario	Occasion 3	4	5	6	7	Global test
No heterogeneity	1.2	1.2	0.8	0.4	0.8	0.8
Binomial mixture	36	48	66	59	44	86.4
Survival heterogeneity	0	2.82	1.2	0.4	1.2	2.4
Trap-happiness	1.5	0.4	0.8	3.2	1.2	2.4

- Cannot pool the tests as the tests are not independent
- If little temporal variation is expected, one may consider a global test using only the most informative occasion for each animal
- This would be the middle occasion between first and last capture

Great cormorant application

- Capture heterogeneity is suspected in this data set due to the nesting habits of the birds and sampling protocol of the study.

- Existing diagnostic goodness-of-fit tests have not been conclusive.

Table: Test of positive association

Capture occasions	P
3	-
4	0.226
5	0.194
6	0.108
7	**0.022**
8	0.186
Global	**0.030**

Multi-state capture-recapture model

The CJS model can be easily extended for multi-state capture-recapture data:

- `A 0 0 B 0`
- `B B 0 B A`
- `0 0 A 0 B`
- ...

Multi-state capture-recapture model

The CJS model can be easily extended for multi-state capture-recapture data:

- `A 0 0 B 0`
- `B B 0 B A`
- `0 0 A 0 B`
- ...

The parameters for the model are now:

- $\phi_t(r)$: probability an individual alive and in state r at occasion t survives until occasion $t+1$
- $\psi_t(r,s)$: probability an individual alive and in state r at occasion t moves to state s by occasion $t+1$
- $p_t(s)$: probability an individual in state s at occasion t is captured

Mover-stayer scenario

- Exploring whether adapted versions of the test of positive association can be used to detect heterogeneity in transition
- Calculate the number of transitions made rather than the number of captures, conditioning on the number of times an individual is captured

Mover-stayer scenario

- Exploring whether adapted versions of the test of positive association can be used to detect heterogeneity in transition
- Calculate the number of transitions made rather than the number of captures, conditioning on the number of times an individual is captured

- Simulation results look promising so far
- Application to Canada goose data set gives a significant results (P=0.0008)

Conclusions

- Some diagnostic goodness-of-fit tests are in fact score tests
- An alternative approach could be taken which directly incorporates the departures from model assumptions within the model selection step

Conclusions

- Some diagnostic goodness-of-fit tests are in fact score tests
- An alternative approach could be taken which directly incorporates the departures from model assumptions within the model selection step

- Existing diagnostic tests lack the power to detect capture heterogeneity
- Proposed a new approach using a test of positive association
- Test works well on simulated and real data
- A global test can be implemented

Conclusions

- Some diagnostic goodness-of-fit tests are in fact score tests
- An alternative approach could be taken which directly incorporates the departures from model assumptions within the model selection step

- Existing diagnostic tests lack the power to detect capture heterogeneity
- Proposed a new approach using a test of positive association
- Test works well on simulated and real data
- A global test can be implemented

- Possible to extend the test to detect mover-stayer behaviour in multi-state capture-recapture data

Further work

- Heterogeneity in survival difficult to detect as time of death is not known - possible development with ring-recovery data
- Mixture models are a special case of a more general class of models, multievent models
- Develop diagnostic tests for multievent models and propose sensible model selection strategy

References

Jeyam, A., McCrea, R. S. and Pradel, R. (2016) A test of positive association for detecting heterogeneity in capture for capture-recapture data. In submission.

McCrea, R. S., Morgan, B. J. T. and Gimenez, O. (2015) A new strategy for diagnostic model assessment in capture-recapture. In submission

McCrea, R. S. and Morgan, B. J. T. (2014) *Analysis of capture-recapture data*. Chapman and Hall/CRC Press. Boca Raton.

McCrea, R. S. and Morgan, B. J. T. (2011) Multi-site mark-recapture model selection using score tests. *Biometrics*, **67**, 234–241.

Pollock, K., Nichols, J. and Hines, J. (1985) Goodness-of-fit tests for open capture-recapture models. *Biometrics*, **41**, 399-410.