Basic overview/features of CMR for paleobiology

Lee Hsiang Liow

Natural History Museum, University of Oslo, Norway

1.9.2022







Lee Hsiang Liow

THE ESSENCE OF CAPTURE RECAPTURE APPROACHES

Day 1: caught 10 rats, put tags on them

Day 2: caught rats in the same place. 2 had your tags, but 8 didn't

What is the capture probability? 2/10=0.2

How many rats are there in that "place?"

50



"Careful—it might be a trap!"

 $\frac{marked\ Day\ 2}{total\ for\ Day\ 2} = \frac{marked\ Day\ 1}{Estimated\ Total}$

The Lincoln–Petersen method (Petersen–Lincoln index)

THE DATA

Time intervals

Taxon	1	2	3	4	5	6	7	8
Α	0	0	1	1	0	0	1	0
В	0	1	0	0	0	0	0	0
С	0	0	0	0	0	1	1	1
D	1	0	1	1	0	1	0	0
Е	0	0	0	1	0	0	1	1
F	0	1	0	0	1	1	0	1
G	0	0	0	1	0	1	1	0
Н	0	0	0	0	0	1	1	0
1	0	0	1	0	1	1	0	1
J	1	0	0	0	0	0	0	0
K	0	0	0	0	1	0	1	1
L	0	0	1	1	0	0	0	0
M	0	1	0	1	1	0	0	0
N	1	1	0	0	1	0	1	0
0	0	0	0	1	0	1	1	1
Р	0	0	1	0	0	0	0	0
Q	0	0	0	0	0	0	1	0
R	0	0	1	0	1	1	0	1
S	0	0	0	0	1	0	0	1
Т	0	0	0	. 1	0	. 0	0 .	0

Liow L.H. & Nichols, J.D. **2010** Estimating rates and probabilities of origination and extinction using taxonomic occurrence data: Capture-recapture approaches. In Quantitative Paleobiology Short Course. Eds. Alroy J. & Hunt G. Paleontological Society pp. 81-94

DECTECTION HISTORIES

Time interval	1	2	3	4	5	6	7	8
L	0	1	1	0	0	0	0	0
M	0	1	0	1	1	0	0	0

Encounter/detection histories

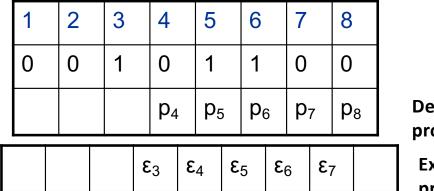
- Series of ones and zeros
- Ones are taken as true presences
- Two types of zeros
 - Not sampled
 - Not sampled or truly absent

		Taxon	1	2	3	4	5	6	7	8
		Α	0	0	1	1	0	0	1	0
		В	0	1	0	0	0	0	0	0
		С	0	0	0	0	0	1	1	1
		D	1	0	1	1	0	1	0	0
Dete	ection	Е	0	0	0	1	0	0	1	1
	ability	F	0	1	0	0	1	1	0	1
•	,			· ·	^				1	0
	^		- /	10		\mathbf{O}	_	1	1	0
	D_{c}	=) //			()			0	1
	P_{6}	•	•						•	0
	- 0							0	0	<u> </u>
	1 0	K	0	0	U	0		0	1	1
	1 0	K L	0	0	1	1	0			
	1 0						0 1	0	1	1
	10	L	0	0	1	1		0	0	0
	10	L M	0	0	1	1	1	0 0	0 0	0 0
	10	L M	0 0 1	0 1 1	1 0 0	1 1 0	1	0 0 0	1 0 0	1 0 0
	10	L M N	0 0 1 0	0 1 1 0	1 0 0 0	1 1 0	1 1 0	0 0 0 0	1 0 0 1	1 0 0 0
	10	L M N O	0 0 1 0	0 1 1 0 0	1 0 0 0	1 1 0 1 0	1 1 0 0	0 0 0 0 1	1 0 0 1 1 0	1 0 0 0 1
		L M N O P	0 0 1 0 0	0 1 1 0 0	1 0 0 0 1	1 1 0 1 0	1 1 0 0 0	0 0 0 0 1	1 0 0 1 1 0	1 0 0 0 1 0
		L M N O P Q R	0 0 1 0 0 0	0 1 1 0 0 0	1 0 0 0 1 0	1 0 1 0 0 0	1 1 0 0 0	0 0 0 0 1 0	1 0 0 1 1 0 1	1 0 0 0 1 0 0

								1	
	Taxon	1	2	3	4	5	6	7	8
	Α	0	0	1	1	0	0	1	0
	В	0	1	0	0	0	0	0	0
	С	0	0	0	0	0	1	1	1
Estimated	D	1	0	1	1	0	1	0	0
no. taxa	Е	0	0	0	1	0	0	1	1
^		_		_			1	0	1
= 2.	s_6 / j	\hat{g}_{c}	= (8)	/ ()	5 =	16L	1	1	0
D_6	$^{D}6'F$	6	- 0	/ O. .			1	1	0
	1	0	0	1	0	1	<u>(1)</u>	0	1
	J	1	0	0	0	0	0	0	0
	K	0	0	0	0	1	0	1	1
	L	0	0	1	1	0	0	0	0
	M	0	1	0	1	1	0	0	0
	N	1	1	0	0	1	0	1	0
	0	0	0	0	1	0	1	1	1
	P	0	0	1	0	0	0	0	0
	Q	0	0	0	0	0	0	1	0
	R	0	0	1	0	1	1	0	1
	S	0	0	0	0	1	0	0	1
	Т	0	0	0	1	0	0	0	0

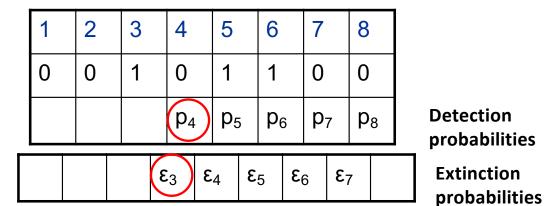
		Taxon	1	2	3	4	5	6	7	8
		A	0	0	1	1	0	0	1	0
		В	0	1	0	0	0	0	0	0
Extinction	'	С	0	0	0	0	0	1	1	1
probabili	ty	D	1	0	1	1	0	1	0	0
	^	1 (1	^r' /		0	1	0	0	1	1
8	$\varepsilon_5 =$	1-(N)	I_6 / I_6	$S_5)$	0	0	1	1	0	1
					0	1	0	1	1	0
					0	0	0	1	1	0
					1	0	1	1	0	1
					0	0	0	0	0	0
					0	0	1	0	1	1
					1	1	0	0	0	0
					0	1	1	0	0	0
					0	0	1	0	1	0
L			· ·		0	1	0	1	1	1
		Р	0	0	1	0	0	0	0	0
		Q	0	0	0	0	0	0	1	0
		R	0	0	1	0	1	1	0	1
		S	0	0	0	0	1	0	0	1
		Т	0	0	0	1	0	0	0	0

	Taxon	1	2	3	4	5	6	7	8	
	Α	0	0	1	1	0	0	1	0	
	В	0	1	0	0	0	0	0	0	
Extinction	С	0	0	0	0	0	1	1	1	
probability	D	1	0	1	1	0	1	0	0	
^	1 (1	^-'		0	1	0	0	1	1	
\mathcal{E}_5 =	$=1-(\Lambda$	M_6 /	(s_5)	0	0	1	1	0	1	
				0	1	0	1	1	0	
	100	/ n		0	0	0	1	1	0	
=1.	$-(\frac{m_6}{2})^{\prime}$	p_6	.)	1	0	1	1	0	1	
1	S	· _	' [0	0	0	0	0	0	
	~	5		0	0	<u>(1)</u>	0	1	1	
	2/6	\ _		1	1	0	0	0	0	
_ 1	3/().5 _\		0	1	1	0	0	0	
-1	16	<u> </u>		0	0	1	0	1	0	
				0	1	0	1	1	1	
	Р	0	0	1	0	0	0	0	0	
	Q	0	0	0	0	0	0	1	0	
	R	0	0	1	0	1	1	0	1	
	S	0	0	0	0	1	0	0	1	
	Т	0	0	0	1	0	0	0	0	

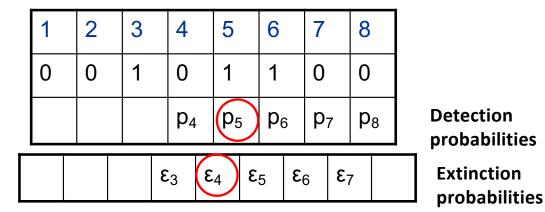


Detection probabilities

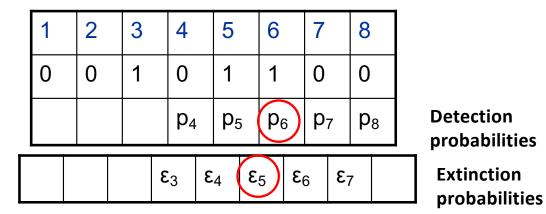
Extinction probabilities



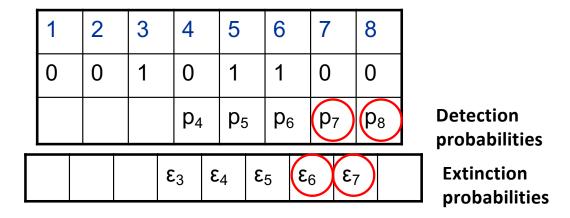
eh = 0 0 1 0 1 1 0 0 Pr(eh = 0 0 1 0 1 1 0 0 | initial encounter in interval 3) = $(1-\epsilon_3)(1-p_4)[1-\epsilon_4]p_5[1-\epsilon_5]p_6[\epsilon_6 + (1-\epsilon_6)(1-p_7)\{\epsilon_7 + (1-\epsilon_7)(1-p_8)\}]$



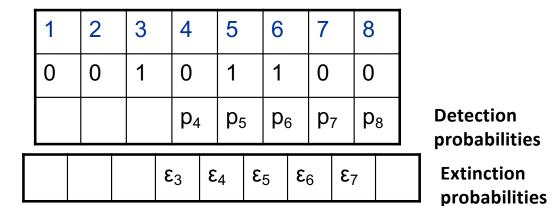
eh = 0 0 1 0 1 1 0 0 Pr(eh = 0 0 1 0 1 1 0 0 | initial encounter in interval 3) = $(1-\epsilon_3)(1-p_4) \ \ (1-\epsilon_4)p_5 \ \ (1-\epsilon_5)p_6 \ \ [\epsilon_6+(1-\epsilon_6)(1-p_7)\{\epsilon_7+(1-\epsilon_7)(1-p_8)\}]$



eh = 0 0 1 0 1 1 0 0 Pr(eh = 0 0 1 0 1 1 0 0 | initial encounter in interval 3) = $(1-\epsilon_3)(1-p_4) \; (1-\epsilon_4)p_5 \; (1-\epsilon_5)p_6 \; [\epsilon_5] + (1-\epsilon_6)(1-p_7)\{\epsilon_7 + (1-\epsilon_7)(1-p_8)\}]$



eh = 0 0 1 0 1 1 0 0 Pr(eh = 0 0 1 0 1 1 0 0 | initial encounter in interval 3) = $+ (1-\epsilon_6)(1-p_7)\{\epsilon_7 + (1-\epsilon_7)(1-p_8)\}]$



eh = 0 0 1 0 1 1 0 0 Pr(eh = 0 0 1 0 1 1 0 0 | initial encounter in interval 3) = $(1-\epsilon_3)(1-p_4) \; (1-\epsilon_4)p_5 \; (1-\epsilon_5)p_6 \; [\epsilon_6 + (1-\epsilon_6)(1-p_7)\{\epsilon_7 + (1-\epsilon_7)(1-p_8)\}]$

"sampling" and vital parameters are both explicit!

We have detection histories – now what?

```
Pr(eh = 0 0 1 0 1 1 0 0) = (1-\epsilon 3)(1-p4)(1-\epsilon 4)p5(1-\epsilon 5)p6[\epsilon 6 + (1-\epsilon 6)(1-p7)\{\epsilon 7 + (1-\epsilon 7)(1-p8)\}] = H1
```

Likelihood (parameters | data) = $H_1^{No. cases} H_2^{No. cases} H_x^{No. cases}$

0001010010	10
0010001101	3
04.04.04.04	4
0101010100	18
1001000000	2
0001010111	1
0011101000	3
0011101000	No. cases

Likelihood of Detection histories

- 1. Estimate parameters (by maximizing the likelihood)
- 2. Estimate uncertainty in parameters
- 3. Compare models
 - e.g. same or different p's or ε's
 - e.g. with or without covariates (important factors that you think might influence p and ϵ)
 - i. Akaike Information Criteria, AIC
 - ii. classical hypothesis testing
 - iii. extendable to Bayesian approaches
- 4. Good statistical properties

- 1. After initial encounters, detection/encounter probabilities are equal for all taxa in the data/group of interest
- 2. After initial encounters, extinction probabilities for all taxa are equal
- 3. Sampling intervals are short relative to the time over which extinction is to be estimated
- 4. The fate of each taxon (with respect to extinction and encounter) is independent of the fate of every other taxon

- 1. After initial encounters, detection/encounter probabilities are equal for all taxa in the data/group of interest
 - Taxon specific covariates
- 2. After initial encounters, extinction probabilities for all taxa are equal
- 3. Sampling intervals are short relative to the time over which extinction is to be estimated
- 4. The fate of each taxon (with respect to extinction and encounter) is independent of the fate of every other taxon

- 1. After initial encounters, detection/encounter probabilities are equal for all taxa in the data/group of interest
- 2. After initial encounters, extinction probabilities for all taxa are equal
 - Taxon specific covariates
- 3. Sampling intervals are short relative to the time over which extinction is to be estimated
- 4. The fate of each taxon (with respect to extinction and encounter) is independent of the fate of every other taxon

- 1. After initial encounters, detection/encounter probabilities are equal for all taxa in the data/group of interest
- 2. After initial encounters, extinction probabilities for all taxa are equal
- 3. Sampling intervals are short relative to the time over which extinction is to be estimated
 - Simulations show that this is not a big problem; other models (e.g. robust design models) tackle this head-on
- 4. The fate of each taxon (with respect to extinction and encounter) is independent of the fate of every other taxon

- 1. After initial encounters, detection/encounter probabilities are equal for all taxa in the data/group of interest
- 2. After initial encounters, extinction probabilities for all taxa are equal
- 3. Sampling intervals are short relative to the time over which extinction is to be estimated
- 4. The fate of each taxon (with respect to extinction and encounter) is independent of the fate of every other taxon
 - Corrections for over-dispersion
 - Co-occurrence analyses

Covariate modeling

- A way to include factors or variables that may be important in explaining variation in the parameters (e.g. extinction, sampling) you are interested in
- Allows us to compare models with different [or no] covariates (Model comparison and selection)
- models to compare
 - ε(constant)p(time-varying)
 - ε(time-varying)p(sea-levels)

Covariate modeling via link functions

$$\operatorname{logit}(\boldsymbol{\varepsilon}_{i,t}) = \operatorname{log}(\frac{\boldsymbol{\varepsilon}_{i,t}}{1 - \boldsymbol{\varepsilon}_{i,t}}) = \beta_0 + \beta(x_i) + \beta_2 y_t$$
Taxon specific covariates

Taxon specific covariates

- •size
- minerology
- taxonomic group

Covariate modeling via link functions

$$\operatorname{logit}(\varepsilon_{i,t}) = \operatorname{log}(\frac{\varepsilon_{i,t}}{1 - \varepsilon_{i,t}}) = \beta_0 + \beta_1 x_i + \beta_2 y_t$$
Time specific covariates

Time specific covariates

- Duration of bin
- •Sea-level
- Temperature

Covariate modeling via link functions

$$logit(\varepsilon_{i,t}) = log(\frac{\varepsilon_{i,t}}{1 - \varepsilon_{i,t}}) = \beta_0 + \beta_1 x_i + \beta_2 y_t$$

$$\varepsilon_{i,t} = \frac{e^{\beta_0 + \beta_1 x_i + \beta_2 y_t}}{1 + e^{\beta_0 + \beta_1 x_i + \beta_2 y_t}}$$

Why Capture-Mark-Recapture (CMR)?

- Detection probability
- Separating between
 - probability of detection (given presence)
 - probability of the parameters in question

(e.g. survivorship, origination, occupancy, immigration) and derived parameters such as species richness/diversity

The probability of detection or sampling is sometimes only a nuisance but sometimes interesting in itself.

- Covariates can be EASILY included in models for both vital parameters and sampling/detection estimates.
- Covariates can be modeled at a variety of levels (e.g. group factors, individual traits, temporal characteristics)