Basic overview/features of CMR for paleobiology

Lee Hsiang Liow

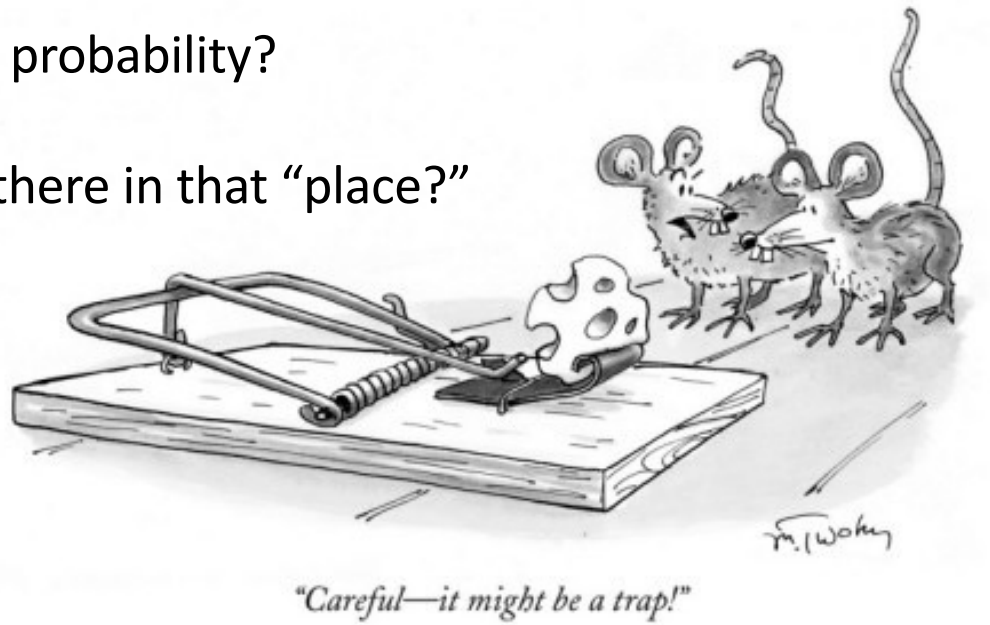
Natural History Museum and Centre for Planetary Habitability, University of Oslo, Norway

15.8.2024 Thursday

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!"

$$\frac{\boldsymbol{marked\ Day\ 2}}{\boldsymbol{total\ for\ Day2}} = \frac{\boldsymbol{marked\ Day1}}{\boldsymbol{Estimated\ Total}}$$

The **Lincoln–Petersen method (**Petersen–Lincoln index)

THE DATA

	Time intervals							
Taxon	1	2	3	4	5	6	7	8
A	0	0	1	1	0	0	1	0
B	0	1	0	0	0	0	0	0
C	0	0	0	0	0	1	1	1
D	1	0	1	1	0	1	0	0
E	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
H	0	0	0	0	0	1	1	0
I	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
O	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
T	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**

Time intervals

Taxon	1	2	3	4	5	6	7	8
A	0	0	1	1	0	0	1	0
B	0	1	0	0	0	0	0	0
C	0	0	0	0	0	1	1	1
D	1	0	1	1	0	1	0	0
E	0	0	0	1	0	0	1	1
F	0	1	0	0	1	1	0	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	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
O	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
T	0	0	0	1	0	0	0	0

Detection probability

$$\hat{p}_6 = 5/10 = 0.5$$

Time intervals

Taxon	1	2	3	4	5	6	7	8
A	0	0	1	1	0	0	1	0
B	0	1	0	0	0	0	0	0
C	0	0	0	0	0	1	1	1
D	1	0	1	1	0	1	0	0
E	0	0	0	1	0	0	1	1
F						1	0	1
G						1	1	0
H						1	1	0
I	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
O	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
T	0	0	0	1	0	0	0	0

Estimated no. taxa

$$\hat{S}_6 = s_6 / \hat{p}_6 = 8 / 0.5 = 16$$

Extinction probability

$$\hat{\varepsilon}_5 = 1 - (\hat{M}_6^{'} / s_5)$$

	Time intervals							
Taxon	1	2	3	4	5	6	7	8
A	0	0	1	1	0	0	1	0
B	0	1	0	0	0	0	0	0
C	0	0	0	0	0	1	1	1
D	1	0	1	1	0	1	0	0
				0	1	0	0	1	1
				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
				0	1	0	1	1	1
P	0	0	1	0	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
T	0	0	0	1	0	0	0	0

Extinction probability

$$\hat{\varepsilon}_5 = 1 - (\hat{M}_6^{'} / s_5)$$

$$= 1 - (\frac{m_6^{'} / \hat{p}_6}{s_5})$$

$$= 1 - (\frac{3 / 0.5}{16})$$

	Time intervals							
Taxon	1	2	3	4	5	6	7	8
A	0	0	1	1	0	0	1	0
B	0	1	0	0	0	0	0	0
C	0	0	0	0	0	1	1	1
D	1	0	1	1	0	1	0	0
				0	1	0	0	1	1
				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
				0	1	0	1	1	1
P	0	0	1	0	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
T	0	0	0	1	0	0	0	0

Time intervals

Taxon	1	2	3	4	5	6	7	8
A	0	0	1	1	0	0	1	0
B	0	1	0	0	0	0	0	0
C	0	0	0	0	0	1	1	1
D	1	0	1	1	0	1	0	0
				0	1	0	0	1	1
				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
				0	1	0	1	1	1
P	0	0	1	0	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	
T	0	0	0	1	0	0	0	0	

Extinction probability

$$\hat{\varepsilon}_5 = 1 - (\hat{M}_6' / s_5)$$

$$= 1 - \left(\frac{m_6' / \hat{p}_6}{s_5}\right)$$

$$= 1 - \left(\frac{3 / 0.5}{16}\right)$$

Alive in T 5 and 6

Alive in T5

Survival Probability From T5 to T6

Representing detection/encounter histories

1	2	3	4	5	6	7	8
0	0	1	0	1	1	0	0
			p_4	p_5	p_6	p_7	p_8

Detection probabilities

			ε_3	ε_4	ε_5	ε_6	ε_7	

Extinction probabilities

eh = 0 0 1 0 1 1 0 0

CJS model

Pr(eh = 0 0 1 0 1 1 0 0 | initial encounter in interval 3) =

Representing detection/encounter histories

1	2	3	4	5	6	7	8
0	0	1	0	1	1	0	0
			p_4	p_5	p_6	p_7	p_8

Detection probabilities

			ε_3	ε_4	ε_5	ε_6	ε_7	

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) =

$(1-\varepsilon_3)(1-p_4)(1-\varepsilon_4)p_5(1-\varepsilon_5)p_6\,[\varepsilon_6 + (1-\varepsilon_6)(1-p_7)\{\varepsilon_7+(1-\varepsilon_7)(1-p_8)\}]$

Representing detection/encounter histories

1	2	3	4	5	6	7	8
0	0	1	0	1	1	0	0
			p_4	p_5	p_6	p_7	p_8

Detection probabilities

			ε_3	ε_4	ε_5	ε_6	ε_7	

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) =

$(1-\varepsilon_3)(1-p_4)(1-\varepsilon_4)p_5(1-\varepsilon_5)p_6\ [\varepsilon_6 + (1-\varepsilon_6)(1-p_7)\{\varepsilon_7+(1-\varepsilon_7)(1-p_8)\}]$

Representing detection/encounter histories

1	2	3	4	5	6	7	8
0	0	1	0	1	1	0	0
			p_4	p_5	p_6	p_7	p_8

Detection probabilities

			ε_3	ε_4	ε_5	ε_6	ε_7	

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) =

$(1-\varepsilon_3)(1-p_4)\ (1-\varepsilon_4)p_5\ (1-\varepsilon_5)p_6\ [\varepsilon_6 + (1-\varepsilon_6)(1-p_7)\{\varepsilon_7+(1-\varepsilon_7)(1-p_8)\}]$

Representing detection/encounter histories

1	2	3	4	5	6	7	8
0	0	1	0	1	1	0	0
			p_4	p_5	p_6	p_7	p_8

Detection probabilities

			ε_3	ε_4	ε_5	ε_6	ε_7	

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) =

$$+ (1-\varepsilon_6)(1-p_7)\{\varepsilon_7+(1-\varepsilon_7)(1-p_8)\}]$$

Representing detection/encounter histories

1	2	3	4	5	6	7	8
0	0	1	0	1	1	0	0
			p_4	p_5	p_6	p_7	p_8

Detection probabilities

			ε_3	ε_4	ε_5	ε_6	ε_7	

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) =

$(1-\varepsilon_3)(1-p_4)\ (1-\varepsilon_4)p_5\,(1-\varepsilon_5)p_6\ [\varepsilon_6 + (1-\varepsilon_6)(1-p_7)\{\varepsilon_7+(1-\varepsilon_7)(1-p_8)\}]$

"sampling" and vital parameters are both explicit! (closer to generating process of the data)

We have detection histories – now what?

Pr(eh = 0 0 1 0 1 1 0 0) =

(1-ε3)(1-p4) (1-ε4)p5 (1-ε5)p6 [ε6 + (1-ε6)(1-p7){ε7+(1- ε7)(1- p8)}]= H1

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

0001010010	10
0010001101	3
0101010100	4
1001000000	18
0001010111	2
0011101000	1
	3
	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**

Assumptions of the CJS model

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**

Assumptions of the CJS model

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**

Assumptions of the CJS model

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

Assumptions of the CJS model

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, never applied in paleo) 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

Assumptions of the CJS model

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

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

Taxon specific covariates

- size
- minerology
- taxonomic group

Covariate modeling via link functions

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

Time specific covariates

- Duration of bin
- Sea-level
- Temperature

Covariate modeling via link functions

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

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

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)

Chapters in MARK book most relevant (if not using MARK)

- Chapter 1 (introduction)
- Chapter 4 (dipper example, but skip the MARK specific bits)
- Chapter 5 (goodness of fit –not covered in lectures but important)
- Chapter 6 (more on covariates and link functions)
- Chapter 11 (individual covariates)
- Chapter 12 (Pradel) and 13 (JS models in general)

http://www.phidot.org/software/mark/docs/book/