E-iNET Workshop, University of Kent, December 2014

Introduction to capture-mark-recapture models

Rachel McCrea

NERC SCIENCE OF THE ENVIRONMENT

EPSRC Engineering and Physical Sciences Research Council

ncse National Centre for Statistical Ecology

University of Kent

Overview

- Introduction
- Lincoln-Petersen estimate
- Maximum-likelihood theory*
- Capture-mark-recapture
- Cormack-Jolly-Seber model
- Developments

* Technical aside

Historical Context

- Estimation of the population of France (1802)

- Human demography
  - What are we interested in?
- Marking of birds started ~1899 in Denmark

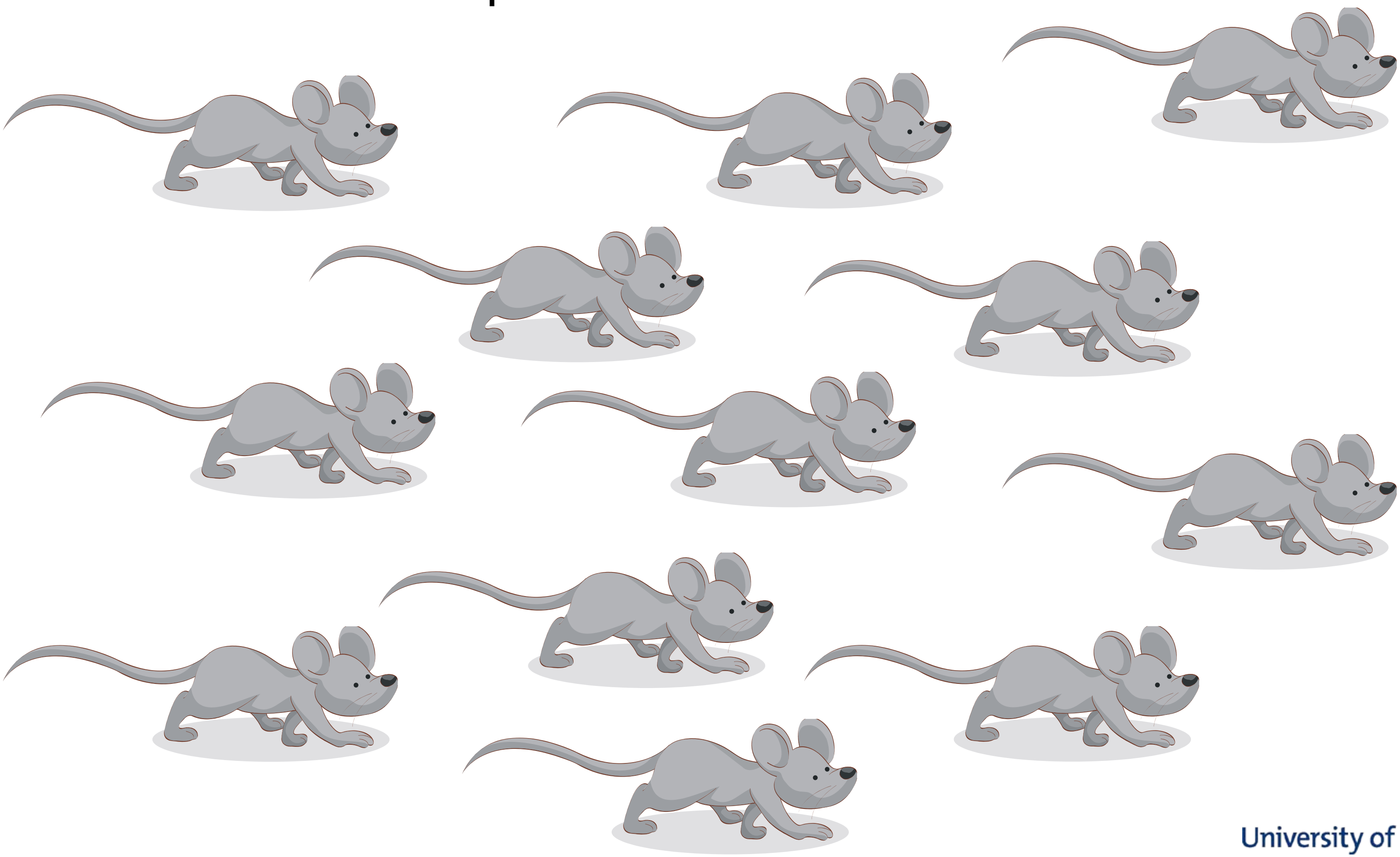
- Capture-recapture techniques widely applied:
  - Medical
  - Epidemiological
  - Sociological
  - Ecological

Lincoln-Petersen Estimate

- Two sample experiment
- Interested in estimating population size, **N**

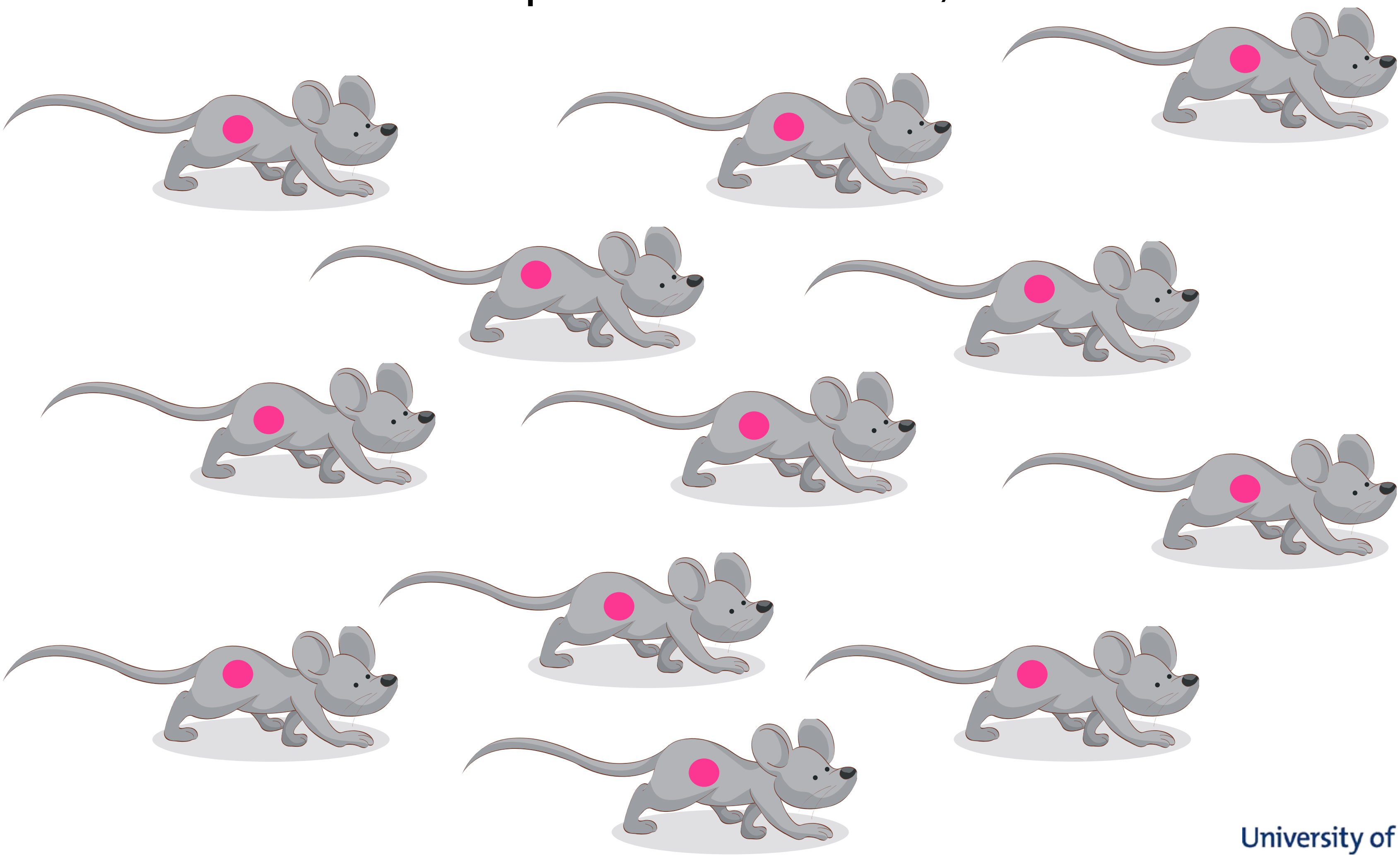
Lincoln-Petersen Estimate

- First sample



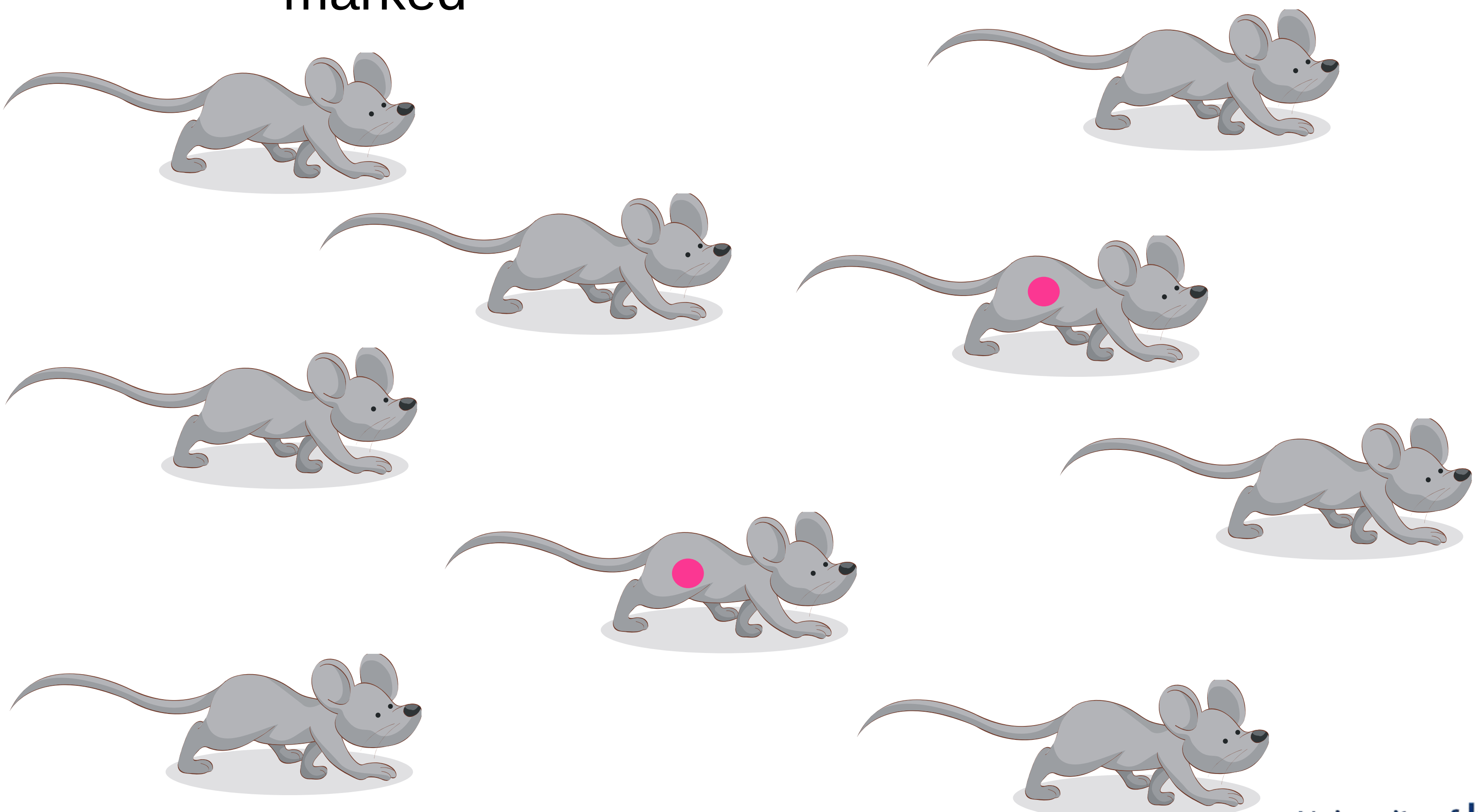
Lincoln-Petersen Estimate

- Mark all captured individuals, K



Lincoln-Petersen Estimate

- Second sample, n animals captured, k already marked



Lincoln-Petersen Estimate

- The proportion of marked individuals caught (k/K) should equal the proportion of the total population caught (n/N)

$$\hat{N} = \frac{Kn}{k}$$

Chapman Estimator

- Biased for small sample sizes, adapted Chapman estimator:

$$\hat{N} = \frac{(K+1)(n+1)}{k+1} - 1$$

- Associated estimated variance:

$$Var(\hat{N}) = \frac{(K+1)(n+1)(K-k)(n-k)}{(k+1)^2(k+2)}$$

Lincoln-Petersen Estimate: Model Assumptions

- Population is closed
- Marks are permanent and do not affect catchability
- All animals equally likely to be captured in each sample
- Sampling time is short in relation to total time
- Samples are taken randomly

Maximum-likelihood theory

- Binomial example
- Consider flipping a coin N times, what is the probability of getting y heads?

$$f(y|N,p) = \binom{N}{y} p^{y}(1-p)^{N-y}$$

- Here we assume we know:
  - N: the number of times that we flip the coin
  - p: the probability of getting a head in a single flip

Maximum-likelihood theory

- However, what if we know N and y, can we work out what p is?

- We want to find the value of p which *maximises the likelihood* that we would observe the data that we did.

$$L(p|N,y) = \binom{N}{y} p^{y} (1-p)^{N-y}$$

Maximum-likelihood theory

- Multinomial example
- The binomial example had two possible outcomes, so what if we have more outcomes? For example, throwing a die with six sides?
- Suppose there are T possible outcomes, and each outcome has probability p_i of occuring.
- N: number of times the experiment is run.
- y_i: number of times outcome i is observed.

Maximum-likelihood theory

- The probability of observing outcomes y_i

$$f(y_1 \dots y_T | N, p_i) = \binom{N}{y_1 \dots y_T} {p_1}^{y_1} {p_2}^{y_2} \dots {p_T}^{y_T}$$

Maximum-likelihood theory

- The probability of observing outcomes y_i

$$f(y_1 \dots y_T | N, p_i) = \binom{N}{y_1 \dots y_T} {p_1}^{y_1} {p_2}^{y_2} \dots {p_T}^{y_T}$$

- The sum of the probabilities must equal 1:

$$f\left(y_1 \dots {y_T}_i \middle| N, p_i\right)$$
$$= \binom{N}{y_1 \dots y_T} {p_1}^{y_1} {p_2}^{y_2} \dots \left(1 - \sum_{i=1}^{T-1} p_i\right)^{\left(N - \sum_{i=1}^{T-1} y_i\right)}$$

Maximum-likelihood theory

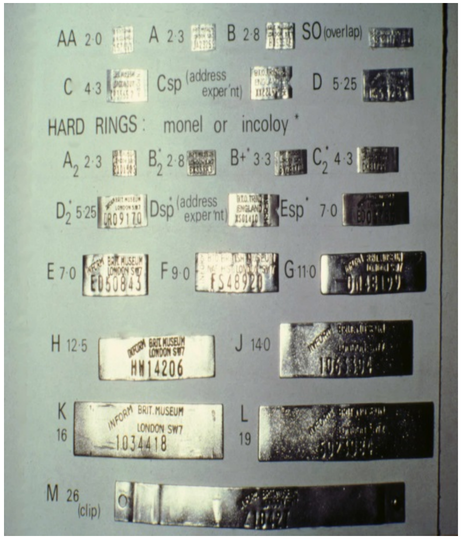
- The multinomial likelihood is then given by

$$L(p_i|N, y_1 \dots y_T)$$

$$= \binom{N}{y_1 \dots y_T} {p_1}^{y_1} {p_2}^{y_2} \dots \left(1 - \sum_{i=1}^{T-1} p_i\right)^{\left(N - \sum_{i=1}^{T-1} y_i\right)}$$

- For more details see Chapter 1 of the Program Mark manual, http://www.phidot.org/software/mark/docs/book/

Capture-mark-recapture



- Capture-recapture data
  - 1001
  - 1101
  - 0011
  - …

- Retain individual and temporal information
- Open population – demographic parameters

Cormack-Jolly-Seber model

- Population is open – births and death can occur within the study period
- Population size is no longer the parameter of interest
- ϕ: survival probability
- Condition on first capture

Cormack-Jolly-Seber model

- Capture-recapture data and probabilities
  - 1001
  - 1101
  - 0011
  - …
- ϕ_t: probability an individual survives from occasion t to t+1
- p_t: probability an individual is captured at occasion t

Cormack-Jolly-Seber model

- Capture-recapture data and probabilities
  - 1001 $\phi_1(1-p_2)\phi_2(1-p_3)\phi_3 p_4$
  - 1101 $\phi_1 p_2 \phi_2 (1-p_3)\phi_3 p_4$
  - 0011 $\phi_3 p_4$
  - …
- ϕ_t: probability an individual survives from occasion t to t+1
- p_t: probability an individual is captured at occasion t

Capture-recapture m-array

Number Released	Recaptured at 2	Recaptured at 3	Recaptured at 4	Never Recaptured
R_1	m_{12}	m_{13}	m_{14}	$R_1-m_{12}-m_{13}-m_{14}$
R_2		m_{23}	m_{24}	$R_2-m_{23}-m_{24}$
R_3			m_{34}	R_3-m_{34}

- R_i: Number of individuals released in year i
- m_{ij}: Number of individuals released in year i and next recaptured in year j

Encounter histories to m-arrays

Number Released	Recaptured at 2	Recaptured at 3	Recaptured at 4	Never Recaptured
R_1	m_{12}	m_{13}	m_{14}	$R_1-m_{12}-m_{13}-m_{14}$
R_2		m_{23}	m_{24}	$R_2-m_{23}-m_{24}$
R_3			m_{34}	R_3-m_{34}

- **1011**
- R_i: newly marked individuals AND re-released individuals

M-array probabilities

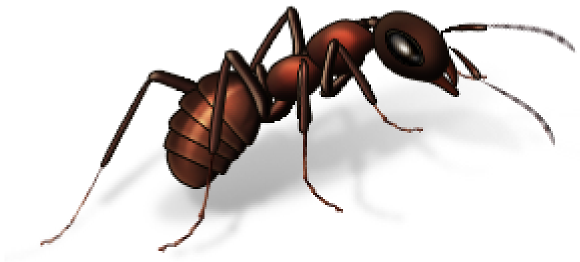
Number Released	Recaptured at 2	Recaptured at 3	Recaptured at 4	Never Recaptured
R_1	$\phi_1 p_2$	$\phi_1(1-p_2)\phi_2 p_3$	$\phi_1(1-p_2)\phi_2(1-p_3)\phi_3 p_4$	χ_1
R_2		$\phi_2 p_3$	$\phi_2(1-p_3)\phi_3 p_4$	χ_2
R_3			$\phi_3 p_4$	χ_3

- ϕ_t: probability an animal alive at time t survives to time t+1
- p_{t+1}: probability an animal alive at time t+1 is recaptured at time t+1
- Each row of data follows a multinomial distribution

Development 1

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

Ayre, L.G. 1962. Problems in using the Lincoln index for estimating the size of ant colonies (*Hymenopter formicidae*). *J.N.Y. Ent. Soc.* **70**: 159-166.



Lincoln-Petersen Estimate: Model Assumptions

- Population is closed
- Marks are permanent and do not affect catchability
- All animals equally likely to be captured in each sample
- Sampling time is short in relation to total time
- Samples are taken randomly

Development 1: Heterogeneity

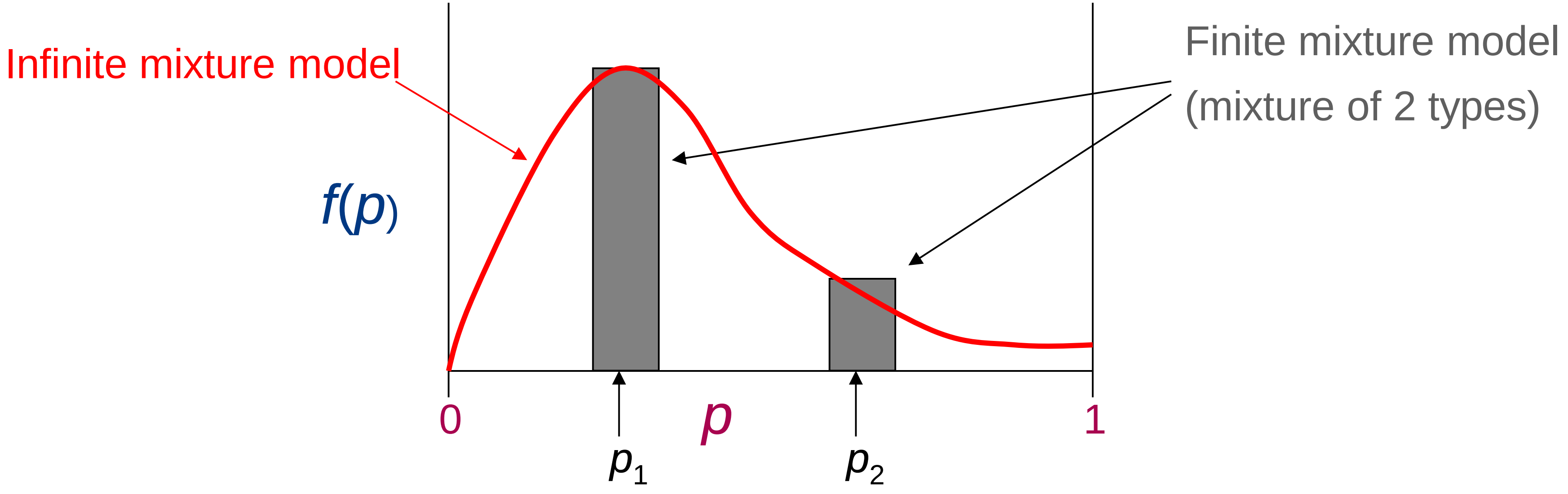
- Not all animals have the same capture probability
- Finite Mixture models[1]: approximate the distribution of p using a mixture of a few ps (conceptually from a few sub-populations).
- Continuous (or infinite mixture) models[2]: Model the distribution of p using some flexible continuous distribution (e.g. Beta)
- Combinations of both[3].

[1] Pledger, (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* **56**: 434-442.

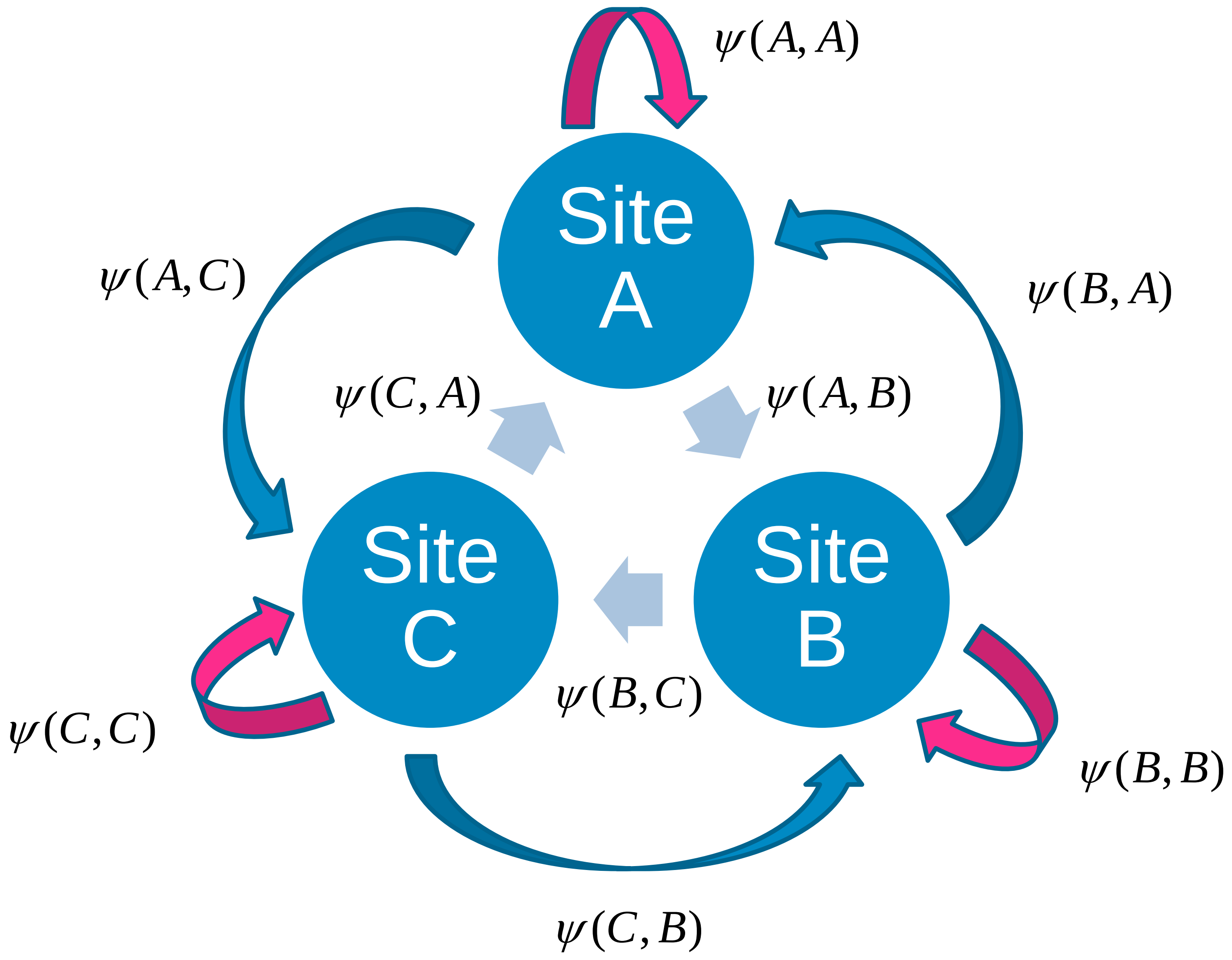
[2]. Dorazio, and Royle, (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **59**: 352-364.

[3] Morgan and Ridout (2008) A new mixture model for capture heterogeneity. *JRSS-C* **57**, 433-446.

Development 1: Heterogeneity



Development 2: Modelling movement



Development 2: Modelling movement

- $\phi_t(r)$: probability of an animal alive at time t in site r, survives until time t+1

- $p_{t+1}(s)$: probability of an animal alive in site s at time t+1 being recaptured

- $\psi_t(r,s)$: probability of an animal alive in site r at time t moving to site s by time t+1

- Lebreton, Nichols, Barker, Pradel and Spendelow (2009) Modeling individual animal histories with multistate capture-recapture models. *Advances in Ecological Research*. **41**, 87-173.

Development 3: Time-dependent covariates

- Temporal variation may be due to measurable factors:
  - Sampling effort (capture probability)
  - Weather conditions/climate indices
  - Population density

- Logistic regression to include covariates in model

$$logit(\phi_t) = \alpha + \beta w_t$$

- North and Morgan (1979) Modelling heron survival using weather data. *Biometrics.* **35**, 667-681.

Development 4: Individual covariates

- Individual time-varying values:
  - Body mass, body condition, body length

- Missing data problem: when individuals are not captured, you do not know what value the covariate takes

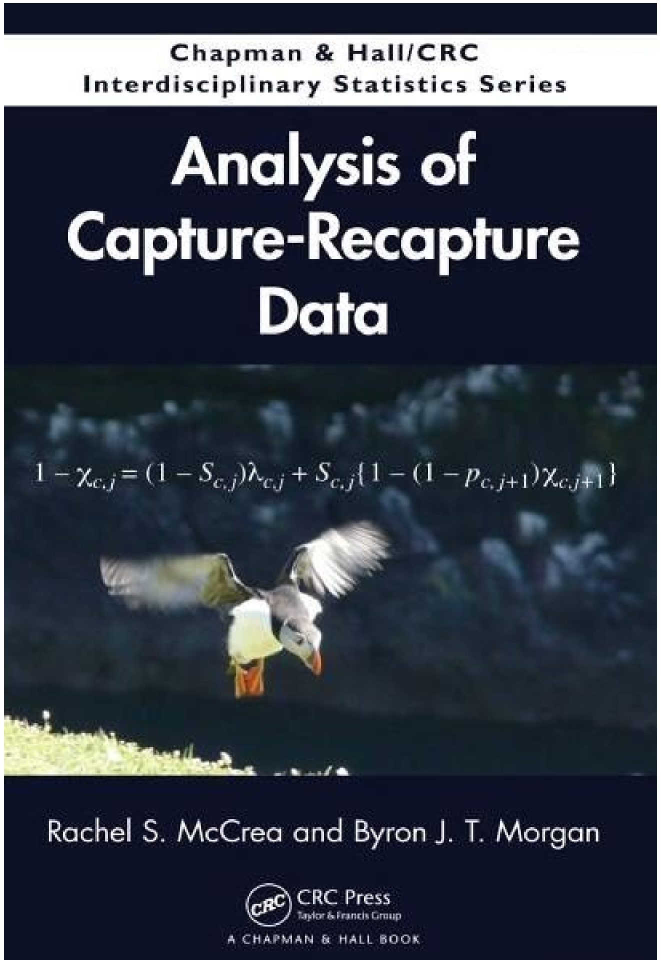
- Current approaches:
  - Trinomial method
  - Bayesian approaches
  - Hidden Markov models

- Section 7.4 of McCrea and Morgan (2014) *Analysis of capture-recapture data.* Chapman and Hall, CRC Press.

Useful References

- Williams, Nichols and Conroy (2002) *Analysis and management of animal populations.* Academic Press.
- Lebreton, Burnham and Clobert (1992) Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological monographs.* **62**, 67-118.

Book, website and e-mail



- www.capturerecapture.co.uk
- R.S.McCrea@kent.ac.uk