Capture-recapture animal model
0000000

Bayesian inference using MCMC
00

The Blue tits as a case study
000000000000

What's next?
0000000

Combining longitudinal data and pedigree information to assess heritability of survival

Julien Papaix, Sarah Cubaynes, Anne Charmantier, Philippe Perret and Olivier Gimenez

Context

- Quantitative genetics for assessing the ability of a trait to respond to natural selection.

- Recently proposed for wild animal and plant populations.

- Applications in evolutionary ecology, management and conservation biology.

- Heritability: Proportion of the total phenotypic variance that can be ascribed to additive genetic variance.

- For (demographic) parameters strongly related to fitness, we expect low heritability.

- But, predictions not so clear in wild populations.

Where are we?

- Animal models: mixed models incorporating genetic, environmental and other factors.

- Capture-recapture models: assess demographic parameters with $p < 1$ and individual variability.

Where are we?

- Animal models: mixed models incorporating genetic, environmental and other factors.

- Capture-recapture models: assess demographic parameters with $p < 1$ and individual variability.

- *The idea of combining Animal and Capture-recapture models is in the air (O'Hara et al. 2008; Cam 2009).*

The idea is in the air (1), O'Hara et al. 2008

The idea is in the air (2), Cam 2009

"Estimating the genetic basis of quantitative traits can be tricky for wild animal populations in natural environments: environmental variation often obscures the underlying evolutionary patterns (...) **[The animal model has] been applied to estimation of heritability in life history traits, either in the rare study populations where detection probability is close to 1, or without considering the probability of detecting animals that are alive and present in the study area** (recapture or resighting probability). Applications of the animal model to demographic parameters (fitness components) (...) where detection probability is < 1 require **trans-disciplinary efforts**"

Where are we?

- Animal models: mixed models incorporating genetic, environmental and other factors.

- Capture-recapture models: assess demographic parameters with $p < 1$ and individual variability.

- The idea of combining Animal and Capture-recapture models is in the air (O'Hara et al. 2008; Cam 2009).

Main objective

A unified model combining pedigree information and capture-recapture data to assess heritability of survival in natural conditions.

Plan

1 Capture-recapture animal model: CRAM

2 Bayesian inference using MCMC

3 The Blue tits as a case study

4 What's next?

Outline

1 Capture-recapture animal model: CRAM

2 Bayesian inference using MCMC

3 The Blue tits as a case study

4 What's next?

State–space model

Definition:

A state–space model describes the evolution of two time series running in parallel, one referred to as the state process and the other as the observation process (Buckland et al. 2004).

State–space model for capture–recapture data

- Define $X_{i,t} = \begin{cases} 1 & \text{if } i \text{ is alive at time } t \\ 0 & \text{if } i \text{ is dead at time } t \end{cases}$

State equations – survival process:

$$X_{i,t+1}|X_{i,t} \sim Bernoulli(X_{i,t} \cdot \phi_{i,t})$$

- Define $Y_{i,t} = \begin{cases} 1 & \text{if } i \text{ captured at time } t \\ 0 & \text{if } i \text{ non-captured at time } t \end{cases}$

Observation equations – detection:

$$Y_{i,t}|X_{i,t} \sim Bernoulli(X_{i,t} \cdot p_{it})$$

- Clark et al. (2005), McCarthy (2006), Gimenez et al. (2007), Royle (2008) (...)

Threshold model to deal with discrete traits

Survival was related to a continuous underlying latent variable $l_{i,t}$, which, given $X_{i,t} = 1$, satisfied:

$$X_{i,t+1} = \begin{cases} 1 & \text{if } l_{i,t} > \kappa, \\ 0 & \text{if } l_{i,t} \leq \kappa. \end{cases}$$

where κ is a threshold value.

Distribution for the so–called liability $l_{i,t}$

We assumed $l_{i,t} \sim N(\mu_{i,t}, \sigma_\epsilon^2)$.

Threshold model to deal with discrete traits

What is the link between survival and liability?

- Recall: $l_{i,t} \sim N(\mu_{i,t}, \sigma_\epsilon^2)$

$$
\begin{aligned}
\phi_{i,t} &= \Pr(X_{i,t+1} = 1 | X_{i,t} = 1) \\
&= \Pr(l_{i,t} > \kappa) \\
&= F\left(\frac{\mu_{i,t} - \kappa}{\sigma_\epsilon}\right)
\end{aligned}
$$

where F is the cumulative function of a $N(0,1)$

- For identifiability reasons, and without loss of generality, we fixed σ_ϵ to 1 and κ to 0.

Relationship between survival and liability

We have: $\phi_{i,t} = F(\mu_{i,t})$

- F^{-1} is usually referred to as the probit function.

Capture-recapture animal model   Bayesian inference using MCMC   The Blue tits as a case study   What's next?

○○○○○●○   ○○   ○○○○○○○○○○○   ○○○○○○○

Towards a capture–recapture animal model (CRAM)

- Recall: $\phi_{i,t} = F(\mu_{i,t})$ where $F \sim N(0,1)$

Plug the animal model in the capture-recapture model:

$$\text{probit}(\phi_{i,t}) = F^{-1}(\phi_{i,t}) = \mu_{i,t} = \eta + b_t + e_i + a_i$$

- η is a constant term for the mean survival on the probit scale,
- b_t is a yearly effect with $b_t \sim N(0, \sigma_t^2)$,
- e_i is an individual non–genetic value with $e_i \sim N(0, \sigma_e^2)$,
- a_i is a genetic value, with $\boldsymbol{a} \sim MN(0, \sigma_a^2 \boldsymbol{A})$, σ_a^2 is the genetic additive variance and $\boldsymbol{A}$ the additive genetic relationship matrix. $\boldsymbol{A}$ is built up from the pedigree (e.g. $A_{individual,itself} = 1$, $A_{offspring,parent} = 0.5$).

Capture-recapture animal model   Bayesian inference using MCMC   The Blue tits as a case study   What's next?

○○○○○○●   ○○   ○○○○○○○○○○○○   ○○○○○○○○

Calculating heritability

- Recall: $\text{probit}(\phi_{i,t}) = \eta + b_t + e_i + a_i$.
- Decomposing components of variance in survival.
- The proportion of variance explained by the additive genetic variance σ_a^2.

> ### The heritability is thus obtained as:
>
> $$h^2 = \frac{\sigma_a^2}{\sigma_t^2 + \sigma_e^2 + \sigma_a^2 + 1}.$$

Outline

1 Capture-recapture animal model: CRAM

2 Bayesian inference using MCMC

3 The Blue tits as a case study

4 What's next?

Estimation via MCMC

- A priori distributions:
  - $p \sim U(0, 1)$,
  - $\eta \sim N(0, 100)$,
  - σ_e, σ_a and $\sigma_b \sim U(0, 10)$.

- Parameters of the MCMC runs:
  - 2 chains,
  - 5000 iterations as a burn–in,
  - 15000 iterations after the burn–in,
  - thinning every 10th iterations,
  - inference using 1000 remaining iterations.

- Checking convergence:
  - visual inspection of mixing and posterior distributions,
  - calculate the BGR statistic ($\hat{R} < 1.2$).

- Implementation:
  - call OpenBUGS from R (block-updating!).

Is heritability "significant", i.e. $\sigma_a^2 \neq 0$?

- Introduce binary variables to have labels on models (Congdon 2005, Royle 2008).

- σ is in the model if $w_\sigma = 1$, and is not if $w_\sigma = 0$.

- For example, $\{w_{\sigma_e} = 0, w_{\sigma_a} = 1\}$ is a model with genetic variance only: $\text{probit}(\phi_{i,t}) = \eta + b_t + a_i$.

- Calculate the posterior probability of a given model using the mean number of iterations spent in this given model.

- All models containing σ_a give $h^2 \neq 0$.

Outline

1. Capture-recapture animal model: CRAM

2. Bayesian inference using MCMC

3. The Blue tits as a case study

4. What's next?

Details on capture-recapture dara

- Blue tits – Study site in Corsica.

- 1979 – 2007 $\Rightarrow$ 29 years of monitoring.

Information on pedigree

- Characteristics of the (social) pedigree:
    - 654 individuals,
    - 218 fathers (sires),
    - 215 mothers (dams),
    - 12 generations.

- Dependence among individuals.

- Complex relationships over several generations.

Capture-recapture animal model
○○○○○○○

Bayesian inference using MCMC
○○

The Blue tits as a case study
○○●○○○○○○○○○

What's next?
○○○○○○○○

Mixing for variance terms

Mixing for mean survival and detection probabilities

Posterior distribution of the detection probability *p*

- Recall: prior is $U(0, 1)$.
- Mean = 0.767,
  Median = 0.769,
  Credible interval =
  $[0.707, 0.822]$.

Posterior distribution of the mean survival $F(\eta)$

- Recall: prior on η is $N(0, 100)$.
- Mean = 0.569, Median = 0.569, Credible interval = $[0.493, 0.650]$.

Posterior distribution of the temporal variance σ_t^2

- Recall: prior is $U(0, 10)$.
- Mean = 0.307,
  Median = 0.301,
  Credible interval =
  $[0.105, 0.558]$.

Posterior distribution of the non–genetic variance σ_e^2

- Recall: prior is $U(0, 10)$.
- Mean = 0.105,
  Median = 0.089,
  Credible interval =
  $[0.005, 0.284]$.

Additive genetic variance σ_a^2 posterior distribution

- Recall: prior is $U(0, 10)$.
- Mean = 0.122,
  Median = 0.110,
  Credible interval =
  [0.006, 0.308].

Posterior distribution of the heritability h^2

- Mean = 0.018,
  Median = 0.011,
  Credible interval =
  $[0.000, 0.077]$.

Model selection 'à la Congdon–Royle' favored...

Model	Posterior probability
0	0.960
σ_e^2	0.019
σ_a^2	0.020
σ_e^2, σ_a^2	0.001

Table: The posterior model probabilities associated with different models for the tits data.

Model selection with DIC

Model	Posterior probability	DIC (pD)
0	0.960	857.6 (437.2)
σ_e^2	0.019	860.2 (441.9)
σ_a^2	0.020	875.6 (456.7)
σ_e^2, σ_a^2	0.001	848.4 (433.7)

Table: The posterior model probabilities and the DIC values associated with different models for the tits data.

Outline

1. Capture-recapture animal model: CRAM

2. Bayesian inference using MCMC

3. The Blue tits as a case study

4. What's next?

Capture-recapture animal model
OOOOOOO

Bayesian inference using MCMC
OO

The Blue tits as a case study
OOOOOOOOOOO

What's next?
●OOOOOOO

Discussion

- Combination of two existing models: animal and capture–recapture models.
- Key role of the liability.
- What if p is assumed to be 1?

Perspectives

- Extension to other parameters (age at first reproduction (Swans) – PhD S. Cubaynes; dispersion (Flycatchers) – post–doc (?) with B. Doligez).

- Simulation study to:
    - check that we are able to detect heritability,
    - explore data requirements (sample size, power),
    - decide which model selection procedure to use, and
    - make sure we have no bugs in the code.

- MCMC or numerical integration (Gimenez and Choquet, submitted; R. Choquet's talk tomorrow morning).

- Goodness-of-fit testing?

Capture-recapture animal model   Bayesian inference using MCMC   The Blue tits as a case study   What's next?

○○○○○○○   ○○   ○○○○○○○○○○○○○   ○○●○○○○○○

Bibliography

Combining pedigree information and mark–recapture data:

- Cam, E. 2009. Contribution of capture-mark-recapture Modeling to studies of evolution by natural selection. Modeling Demographic Processes in Marked Populations. Environmental and Ecological Statistics 3: 83–129.

- O'Hara R.B., Cano, J.M., Ovaskainen, O., Ahlo, J.S., Teplitsky, C. Bayesian Approaches in Evolutionary Quantitative Genetics. Journal of Evolutionary Biology. 21: 949–957.

- Papaix, J., Cubaynes, S., Charmantier, A., Perret, P. and Gimenez, O. (rejected). Combining longitudinal data and pedigree information to assess heritability of demographic parameters in the wild: capture–recapture animal models. Evolution.

State–space modelling and population dynamics:

- Buckland, S.T., Newman, K.B., Thomas, L. and Koesters, N.B. 2004. State-space models for the dynamics of wild animal populations. Ecol. Model. 171: 157-175.

- Clark, J.S. and O. Bjornstad. 2004. Population time series: Process variability, observation errors, missing values, lags, and hidden states. Ecology. 85: 3140–3150.

- Gimenez, O., V. Rossi, R. Choquet, C. Dehais, B. Doris, H. Varella, J.-P. Vila, and R. Pradel. 2007. State-space modelling of data on marked individuals. Ecol. Model., 206: 431–438.

- McCarthy, M. 2007. Bayesian Methods for Ecology. Cambridge University Press, New York.

- Royle, J. A. 2008. Modeling individual effects in the Cormack–Jolly–Seber model: a state–space formulation. Biometrics, 64: 364–370.

Modelling individual histories with state uncertainty

- What? Multievent capture–recapture workshop.

- How? See http://www.cefe.cnrs.fr/biom/Workshops/.

- When? November 16-20, 2009.

- Where? Southern France, Montpellier (good and lots of food too).

- Who? R. Pradel, R. Choquet, J.–D. Lebreton, O. Gimenez, L. Rouan, and....

Paul Doherty: 'I love Montpellier!'

Modelling individual histories with state uncertainty

- What? Multievent capture–recapture workshop.

- How? See http://www.cefe.cnrs.fr/biom/Workshops/.

- When? November 16-20, 2009.

- Where? Southern France, Montpellier (lots of good food too).

- Who? R. Pradel, R. Choquet, J.–D. Lebreton, O. Gimenez, L. Rouan, and P. Doherty

Reparameterizing **a** for improving MCMC

- Henderson (1976): in $\mathbf{a} \sim N(\mathbf{0}, \sigma_a^2 \mathbf{A})$, we can write $\mathbf{A} = \mathbf{TDT}'$ where **T** is a matrix gathering the fraction of genes of animal i inherited by animal j (sparse).

- We can rewrite **a** as $\mathbf{a} = \sigma_a \mathbf{TD}^{\frac{1}{2}} \gamma$ where $\gamma \sim N(\mathbf{0}, \mathbf{Id})$

- Quaas (1989): we use a recursive algorithm to calculate $\mathbf{v} = \mathbf{Ts}$, in which we write: $\mathbf{s} = \mathbf{D}^{\frac{1}{2}} \gamma$ and therefore we deduce: $\mathbf{a} = \sigma_a^2 \mathbf{v}$

- We have: $v_i = D_{ii}^{\frac{1}{2}} \cdot \gamma_i + \dfrac{v_{si} + v_{di}}{2} \quad \forall i = 1 \ldots N$ where the indices si and di are for the sire and dam of individual i.

- See Damgaard (2007) and Waldmann (2009).

Identifiability issues in the threshold model

- Apply constraints to make the threshold model identifiable, namely $\kappa = 0$ and $\sigma_\epsilon^2 = 1$

- Any non–zero κ will be offset by a corresponding shift in the intercept.

- Multiply both sides of $l_{i,t} = \eta + b_t + e_i + a_i + \epsilon_{i,t}$ by $c > 0$. We note that $\theta = (\theta, \sigma_t^2, \sigma_e^2, \sigma_a^2)$ fits the data as well as $\theta^\star = (c\theta, c^2\sigma_t^2, c^2\sigma_e^2, c^2\sigma_a^2)$. Setting σ_ϵ^2 to any positive constant will pin down the latent variable (and hence η, σ_t^2, σ_e^2, and σ_a^2), and the value 1 is chosen for convenience.