INCORPORATING INDIVIDUAL TIME-VARYING COVARIATES WITHIN THE ANALYSIS OF CAPTURE-RECAPTURE DATA*

*Joint work with Roland Langrock

University of St Andrews

Ruth King

Introduction

- Motivating data – Soay sheep.
- Issues associated with continuous time-varying covariate information.
- Previous approaches:
  - Trinomial method;
  - Bayesian data augmentation.
- The use of (approximate) hidden Markov-type model.
- Discussion.

Data - Soay sheep†

- Soay sheep are uniquely marked via ear-tags as lambs.
- Each year observers go into the field and "record" all individuals that they see.

- Individuals can be "captured" or "resighted".
- If an individual is captured, an associated weight may also be recorded.
- We consider data for 1633 females collected from 1986-2010 – there are a total of 6261 captures/resightings, 1230 weights recorded and 1046 recoveries (of dead individuals).

† Many thanks to Tim Coulson for the data

Capture-Recapture-Recovery (MRR)

- This type of data is often referred to as capture-recapture-recovery (or mark-recapture-recovery; MRR) data.
- MRR data are typically summarised in the form of the capture history of each individual observed in the study.
- An example of a capture history is:

$$1\ 1\ 0\ 1\ 0\ 2$$

0 = unobserved;

1 = observed alive (either resighted or recaptured);

2 = recovered dead.

Likelihood

- The corresponding likelihood is the product over each individual observed of the probability of their corresponding capture history, conditional on their initial capture.
- It is a function of survival (ϕ), recapture (***p***) and recovery (λ) probabilities.
- For example, for the previous capture history, the corresponding likelihood contribution is:

  1 1 0 1 0 2: $\phi_1 p_2 \phi_2 (1-p_3) \phi_3 p_4 \phi_4 (1-p_5) (1-\phi_5) \lambda_6$

Individual time-varying covariates

- We are particularly interested in the relationship between weight (as a proxy for condition) of an individual and their associated survival.
- We specify the survival probabilities as a deterministic function of their weight (i.e. a time-varying individual covariate):

$$\text{logit}\ \phi_{it} = \beta_0 + \beta_1\ w_{it}$$

- The corresponding likelihood is no longer available in closed form.
- For example consider the following history:

$$1\ 1\ 0\ 1\ 0\ 2: \quad \phi_1\, p_2\, \phi_2\, (1\text{-}p_3)\, \phi_3\, p_4\, \phi_4\, (1\text{-}p_5)\, (1\text{-}\phi_5)\, \lambda_6$$

Individual time-varying covariates

- We are particularly interested in the relationship between weight (as a proxy for condition) of an individual and their associated survival.
- We specify the survival probabilities as a deterministic function of their weight (i.e. a time-varying individual covariate):

$$\text{logit } \phi_{it} = \beta_0 + \beta_1 \, w_{it}$$

- The corresponding likelihood is no longer available in closed form.
- For example consider the following history:

$$1 \; 1 \; 0 \; 1 \; 0 \; 2: \qquad \phi_1 \, p_2 \, \phi_2 \, (1-p_3) \, \phi_3 \, p_4 \, \phi_4 \, (1-p_5) \, (1-\phi_5) \, \lambda_6$$

Weight unobserved → Survival probabilities "unknown"

Previous approaches

- In the presence of missing values, the likelihood is only expressible in the form of an analytically intractable integral (integrating over all missing covariate values).
- Two approaches are predominantly used in the analysis of time-varying individual covariates:
  - Trinomial (or conditional) approach - the likelihood is constructed by conditioning on ***only*** the observed covariate values, resulting in a closed form (conditional) likelihood.
  - Bayesian data augmentation approach - the missing covariate values are treated as parameters (or auxiliary variables) to be estimated and the joint posterior distribution is defined over the model parameters and auxiliary variables (use MCMC to obtain a sample from distribution).

Trinomial approach

- The "trinomial" (or conditional) approach was proposed by Catchpole, Morgan and Tavecchia (2008).
- The likelihood is constructed by conditioning on ***only*** the observed covariate values, resulting in a closed form (conditional) likelihood.
- Consider the capture history with associated covariate values:

$$1 \longrightarrow 1 \longrightarrow 0 \longrightarrow 1 \longrightarrow 0 \longrightarrow 2$$

$$\phi_1 \, p_2 \quad \phi_2 \, (1\text{-}p_3) \quad \phi_3 \, p_4 \quad \phi_4 \, (1\text{-}p_5) \quad (1\text{-}\phi_5) \, \lambda_6$$

$$0.41 \longrightarrow ? \qquad ? \qquad 0.76 \longrightarrow ? \qquad NA$$

$$\phi_1(0.41) \, p_2 \qquad 1 - (\phi_4(0.76) \, p_5 + (1 - \phi_4(0.76))\lambda_5)$$

Bayesian approach

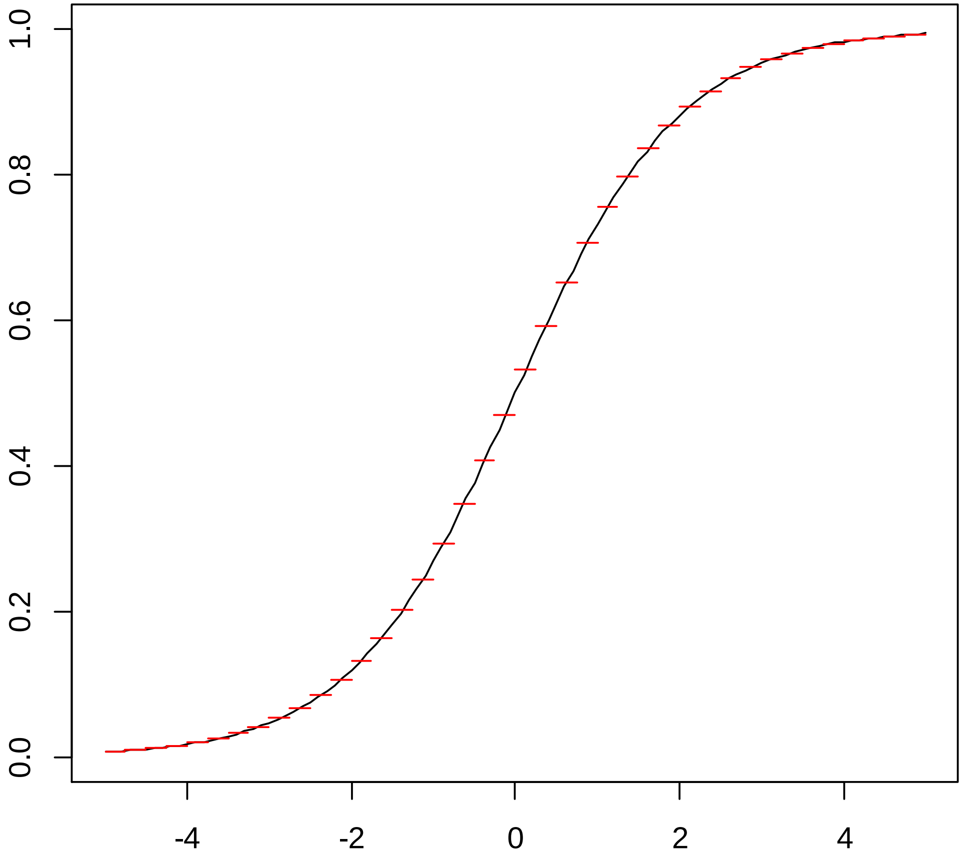
- Data augmentation can be implemented within a Bayesian approach to deal with the missing data.
- The missing covariate values are treated as parameters (or auxiliary variables) and the joint posterior distribution is defined over the model parameters and auxiliary variables.
- MCMC can be used to sample from this posterior distribution and obtain estimates of the posterior summary statistics of interest.

New proposed approach

- The idea is to use a numerical approximation for the integration in the likelihood (the likelihood reduces to a product of 1-dimensional integrals for each missing covariate value).
- Assume first-order Markov process for the covariate process (i.e. weight at time $t+1$ is dependent on only the weight at time t).
- Approximate the integral associated with a missing covariate value by finely discretising the continuous space into a set of finite "bins" (e.g. 40 "bins"); and replace the integral with a summation (survival probabilities within a "bin" are assumed to be constant).
- This allows us to apply the efficient machinery of hidden Markov models for fast computation of the associated likelihood (assuming a Markov structure for covariate model).

Discretisation

Survival probability



Weight (normalised)

Approximate likelihood

- Consider the following capture history:

1	1	0	1	0	2
0.41	?	?	0.76	?	NA

- Assume intervals [0.35,0.45),…,[0.75,0.85),…
- The (approximate) likelihood contribution is: x

$$\sum_{x_2}\sum_{x_3}\sum_{x_5} \phi_1(0.41)p_2\phi_2(x_2)(1\text{-}p_3)\phi_3(x_3)p_4\phi_4(0.76)(1\text{-}p_5)(1-\phi_3(x_5))\lambda_6$$

$$\times\ \psi_1(0.41, x_2)\ \psi_2(x_2, x_3)\ \psi_3(x_3, 0.76)\ \psi_4(0.76, x_5),$$

where each summation is over the midpoints of the intervals.

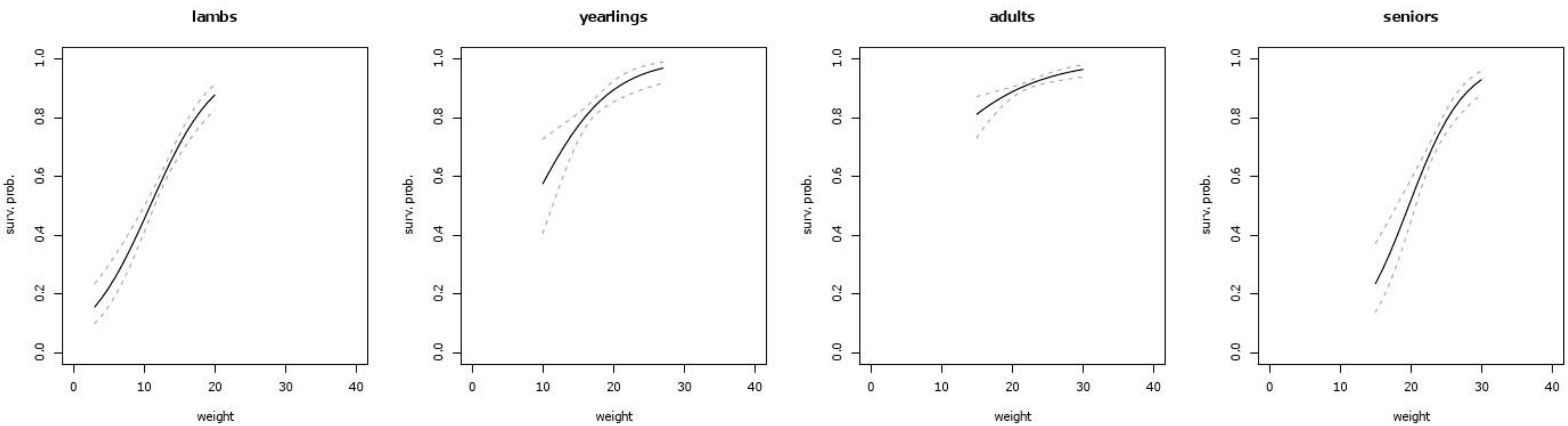
- The transition probabilities ψ are completely defined by the underlying model for the covariate values specified on the continuous scale.

Simulation study (100 datasets)

		Intercept		Slope	
	Model	95% CI width	CI coverage	95% CI width	CI coverage
p=0.95 λ =0.95	Trinomial	1.39	0.96	0.08	0.94
	New approach	1.33	0.94	0.07	0.93
p=0.9 λ =0.3	Trinomial	1.69	0.95	0.12	0.95
	New approach	1.37	0.95	0.08	0.95
p=0.3 λ =0.9	Trinomial	3.08	0.97	0.14	0.97
	New approach	1.46	0.94	0.08	0.95
p=0.3 λ =0.3	Trinomial	3.73	0.98	0.11	0.95
	New approach	1.92	0.94	0.08	0.95

Results – survival

Survival probabilities (as a function of weight) with associated 95% CIs.



	Lambs	Yearlings	Adults	Senior
Intercept	-2.3 (-3.0, -1.7)	-1.5 (-3.2, 0.1)	-0.4 (-1.8, 1.0)	-4.9 (-6.8, -3.1)
Slope	0.21 (0.16, 0.27)	0.18 (0.09, 0.28)	0.12 (0.06, 0.19)	0.25 (0.17, 0.33)

Discussion

- The approximate hidden Markov-type model approach appears to work well.
- Model selection can be performed on both the dependence of the survival/recapture/recovery parameters (e.g. time/age dependence) and the covariate process model: e.g
  - Time dependent p and λ - Δ AIC = 0.
  - Constant p and λ - Δ AIC = 82.86
  - Constant p and time dependent λ - Δ AIC = 31.66.
- Further work includes refining this approach to improve the approximation.