Autoregressive Models for Capture-Recapture Data: A Bayesian Approach

Jennifer Hoeting

Devin Johnson

Colorado State University

Paper and software available at
www.stat.colostate.edu/~jah

Research supported by

- Environmental Protection Agency cooperative agreement R-82909501
- National Science Foundation grant DEB-0091961.

What factors affect animal survival?

Historically, survival probabilities were modeled as as fixed constants.

More recently Burnham and others have begun to consider survival probabilities as realizations of a random process.

We model survival as a function of environmental factors and allow for a time series correlation structure between survival probabilities.

We develop methodology for

- Open population mark-recapture models (Cormack-Jolly-Seber Model)
- Band recovery models

Band Recovery Model

Animals are captured, banded and released.
Hunters report when bands are recovered.

Some notation:

$R_i =$ # banded animals released at time t_i

$m_{ij} =$ # animals recovered at time t_j out of the R_i animals released at time t_i

$I =$ # capture occasions when banding is performed

$J =$ # occasions when bands are recovered

$$(m_{i1}, \ldots, m_{iJ}) \sim \text{Multinomial}\,(R_i, p = f(\text{survival,recovery}))$$

Likelihood for Band Recovery Model

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\lambda}; \boldsymbol{R}, \boldsymbol{m}) = \prod_{i=1}^{I} \binom{R_i}{m_{ii}, \ldots, m_{iJ}} \xi_i^{v_i} \prod_{j=i}^{J} \left\{ \lambda_j \prod_{k=i}^{j-1} \phi_k \right\}^{m_{ij}}$$

where

- ξ_i is the probability that an animal is never recovered after release at t_i
- v_i is the number of animals captured at t_i and never subsequently recovered
- λ_j is the probability that a marked animal, alive at t_j, is harvested between time t_j and t_{j+1} and reported to the banding agency.
- ϕ_k is the probability that an animal survives from time t_k to t_{k+1} given that it is alive at time t_k

A Random Effects Model for Survival Probabilities

We consider a generalized linear model for the probability that an animal survives from time t_j to time t_{j+1} of the form

$$g(\phi_j) = \mathbf{X}'_j \boldsymbol{\beta} + \epsilon_j, \qquad j = 1, \ldots, J,$$

where

- g is an appropriate link function to constrain survival between 0 and 1
- $\mathbf{X}_j$ is a $P \times 1$ matrix of environmental covariates for capture occasion j
- $\boldsymbol{\beta}$ is a $P \times 1$ vector of regression coefficients
- $(\epsilon_1, \ldots, \epsilon_J)' \sim N(\mathbf{0}, \boldsymbol{\Sigma})$

Covariance Matrix for the Survival Model

The covariance matrix, $\mathbf{\Sigma}$, can be any general form.

Here we consider an AR(m) model which implies that the ϵ_j error terms are realizations from the stochastic process

$$\epsilon_j = \sum_{k=1}^{m} \rho_k \epsilon_{j-k} + z_j, \qquad j = 1, \ldots, J,$$

where $z_j \sim$ i.i.d. $N(0, \sigma^2)$ and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_m)$ is a set of parameters.

The stationary AR(m) model

- allows for positive or negative correlation between survival probabilities that decreases with an increasing separation in time
- imposes constraints on $\boldsymbol{\rho}$ (see paper for details)

Parameter Estimation

Parameter estimation via maximum likelihood or quasi-likelihood is challenging in this context. We adopt a Bayesian approach.

For the Bayesian approach, we assume that the parameters $\boldsymbol{\beta}$, σ^2, $\boldsymbol{\rho}$, and $\boldsymbol{\lambda}$ are independent *a priori*.

The posterior distribution of the parameters and random effects is then given by

$$\begin{aligned}\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\rho}, \boldsymbol{\epsilon}, \boldsymbol{\lambda} | \mathbf{m}, \mathbf{R}, \mathbf{X}) \;\; \propto \;\; & \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\epsilon}, \boldsymbol{\lambda}; \mathbf{m}, \mathbf{R}, \mathbf{X}) \\ & \times \; \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\boldsymbol{\rho})\pi(\boldsymbol{\lambda}) \\ & \times |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{\boldsymbol{\epsilon}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}}{2} \right\}\end{aligned}$$

A modified Gibbs Sampler

We show that conditional distributions of the parameters are given by

$$
\begin{aligned}
f(\beta_l|\boldsymbol{\beta}_{-l}, \sigma^2, \boldsymbol{\rho}, \boldsymbol{\epsilon}, \boldsymbol{\lambda}, D) &\propto \prod_{i=1}^{I} \xi_i^{v_i} \prod_{j=i}^{J} \left\{ \prod_{k=i+1}^{j-1} \phi_k \right\}^{m_{ij}} \pi(\beta_l) \\
f(\lambda_l|\boldsymbol{\lambda}_{-l}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, \sigma^2, \boldsymbol{\rho}, D) &\propto \prod_{i=1}^{I} \xi_i^{v_i} \prod_{j=i+1}^{J} \lambda_j^{m_{ij}} \pi(\lambda_l) \\
f(\epsilon_l|\boldsymbol{\epsilon}_{-l}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\rho}, \boldsymbol{\lambda}, D) &\propto \prod_{i=1}^{I} \xi_i^{v_i} \prod_{j=i}^{J} \left\{ \prod_{k=i+1}^{j-1} \phi_k \right\}^{m_{ij}} N\left(\mu_l/\eta_l,\ \sigma^2/\eta_l\right),
\end{aligned}
$$

where for an AR(2) error process, for example,

$$
\mu_l = \begin{cases} \rho_1\epsilon_2 + \rho_2\epsilon_3 & l = 1 \\ \rho_1(\epsilon_1 + \epsilon_3) + \rho_2(\epsilon_4 - \rho_1\epsilon_3) & l = 2 \\ \rho_1(1-\rho_2)(\epsilon_{l-1} + \epsilon_{l+1}) + \rho_2(\epsilon_{l-2} + \epsilon_{l+2}) & l = 3, \ldots, J-2 \\ \rho_1(\epsilon_J + \epsilon_{J-2}) + \rho_2(\epsilon_{J-3} - \rho_1\epsilon_J) & l = J-1 \\ \rho_1\epsilon_{J-1} + \rho_2\epsilon_{J-2} & l = J \end{cases}
$$

$$
\eta_l = \begin{cases} 1 & l = 1 \text{ and } J \\ 1 + \rho_1^2 & l = 2 \text{ and } J-1 \\ 1 + \rho_1^2 + \rho_2^2 & l = 3, \ldots, J-2 \end{cases}
$$

$$
\begin{aligned}
f(\boldsymbol{\rho}|\boldsymbol{\beta}, \boldsymbol{\epsilon}, \sigma^2, \boldsymbol{\lambda}, D) &\propto \left(\prod_{j=1}^{m} K_j \right)^{-1} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{J} (\epsilon_j - \nu_j)^2/K_j \right\} \pi(\boldsymbol{\rho}) \\
f(\sigma^2|\boldsymbol{\beta}, \boldsymbol{\epsilon}, \boldsymbol{\rho}, \boldsymbol{\lambda}, D) &\propto \sigma^{-J} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{J} (\epsilon_j - \nu_j)^2/K_j \right\} \pi(\sigma^2)
\end{aligned}
$$

where K_j is a function of $\boldsymbol{\rho}$.

Northern Pintails

Study of female Northern Pintail ducks in California for banding years 1955 – 1983.

Birds were banded in January of each banding year.

No yearly environmental covariates were available, but we considered the possibility of a trend in survival probabilities over time via the model

$$\text{logit } \phi_j = \beta_0 + \beta_1(j - 14) + \epsilon_j \quad j = 1, \ldots, 27$$

So the covariate vector $\mathbf{X}_j = (1, j - 14)$ for $j = 1, \ldots, 27$.

We also consider the logit model without a slope parameter.

Choosing the order for the AR process

We chose an AR(2) model for these data, so the error terms follow the stochastic process

$$\epsilon_j = \rho_1 \epsilon_{j-1} + \rho_2 \epsilon_{j-2} + z_j, \qquad j = 1, \ldots, 27$$

The AR order was chosen based on a correlogram of the maximum likelihood estimates of yearly survival probabilities from the program MARK.

Prior distributions for the parameters

$$\begin{aligned}
(\beta_0, \beta_1)^T &\sim N\left(\mathbf{0}, \frac{1}{0.01}\mathbf{I}\right) \\
\sigma^{-2} &\sim \Gamma(0.001,\ 0.001) \\
\rho_2 &\sim U(-1,\ 1) \\
\rho_1 | \rho_2 &\sim U(-(1-\rho_2), 1-\rho_2) \\
\lambda_j &\sim \text{i.i.d. } U(0,1) \quad j = 1, \ldots, 28
\end{aligned}$$

MCMC sampling was carried out using winBUGS.

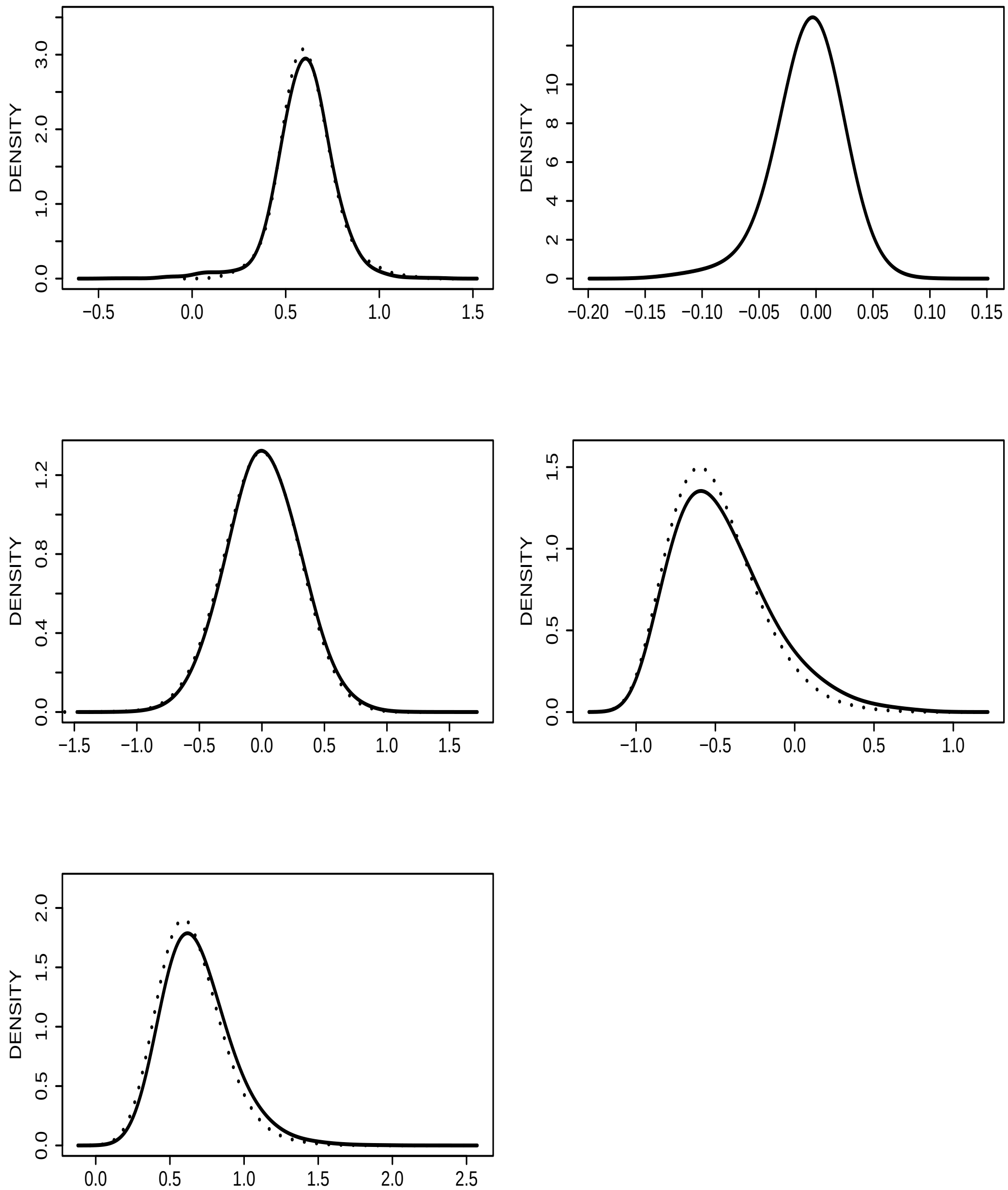

Figure 1. Marginal Posterior Densities.
Solid line = linear time trend model
Dotted line = model with no linear time trend

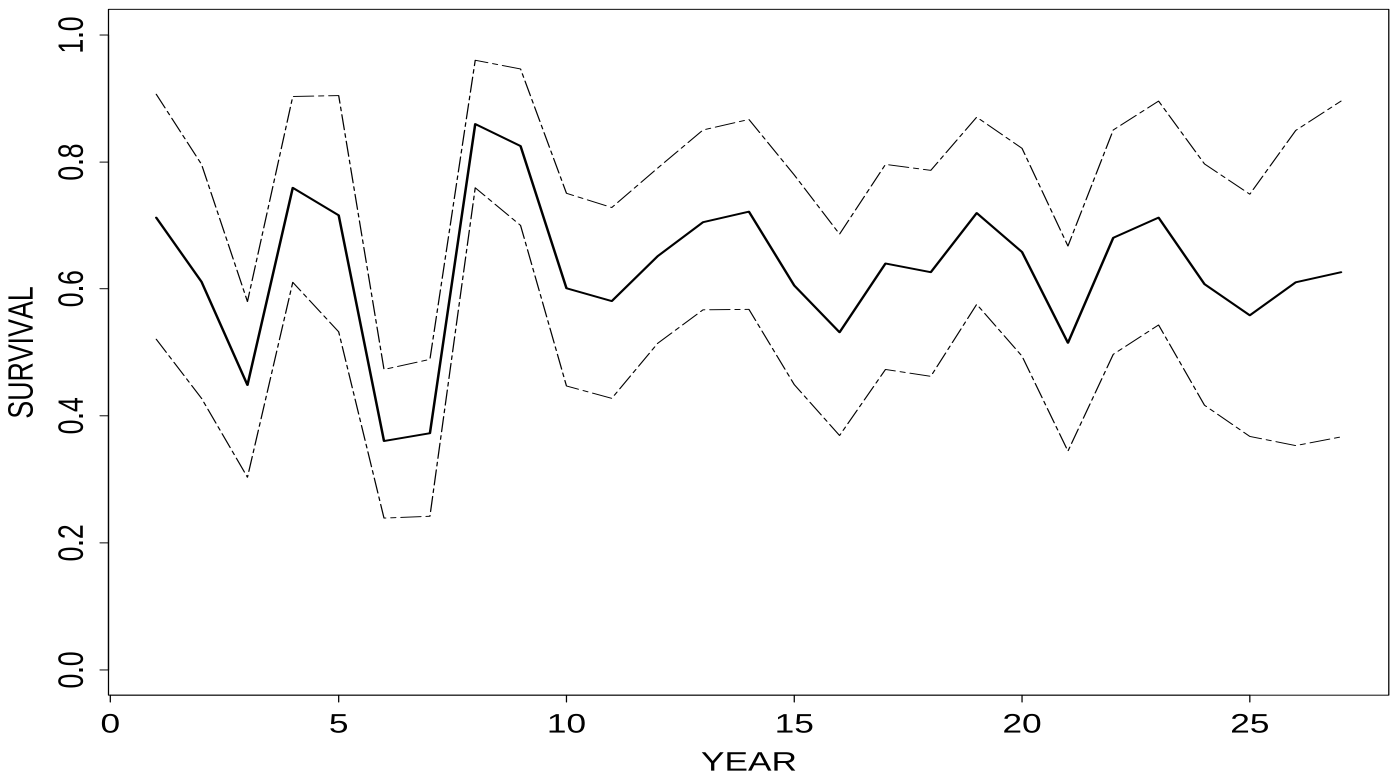

Figure 2. Yearly survival estimates for Northern Pintails (model with no linear time trend). The solid line is the estimated posterior mean survival and the dashed lines represent a 90% HPD interval.

Conclusions

The time series modeling of capture-recapture data can provide additional insights to the survival process.

Extensions and related models include

- Random effects models for recovery parameters
- Multivariate AR process

Other issues include model selection and model uncertainty:

- Which covariates?
- Which link function?
- Which order for the AR process?

Paper and software available at
www.stat.colostate.edu/~jah