Capture-recapture models in `unmarked`

Richard Chandler

January 29, 2012

Abstract

The "`un`" in `unmarked` is somewhat misleading because the package can be used to analyze data from marked animals. The three most common sampling methods that produce suitable data are removal sampling, double observer sampling, and capture-recapture methods[1]. This document focuses on the analysis of capture-recapture data using a class of models known as multinomial N-mixture models [2, 1], which assume that capture-recapture data have been collected at a collection of sample locations ("sites"). Capture-recapture models can be fitted with constant parameters (M_0), time-specific parameters (M_t), and behavioral responses (M_b). In addition, spatial variation in abundance or capture probability can also be modeled using site-specific covariates. unmarked has two functions for fitting capture-recapture models: multinomPois and gmultmix. Both allow for user-defined functions to describe the capture process, and the latter allows for modeling of temporary emigration.

1 Introduction

In traditional capture-recapture models, n individuals are captured at a site during the course of J sampling occasions. The encounter history for each individual is used as information about capture probability p such that the total population size N can be regarded as the size parameter of a binomial distribution, $n \sim \text{Binomial}(N, p)$. When capture probability is assumed to be constant among individuals and over time, the model is refered to as model M_0. Temporal variation in p is assumed by the so-called model M_t. Likewise, behavioral responses, such as trap-avoidance or "trap-happiness" can be modeled using model M_b.

Although traditional capture-recapture models represent the primary method for estimating population size, the do not allow one to model variation in abundance which is a central focus of much ecological research. **(author?)** [2] developed a framework for modeling variation in both abundance and capture probability when capture-recapture data is collected at a set of R sites. Site-specific abundance ($N_i; i = 1, 2, ..., R$) is regarded as latent variable following a discrete distrubution such as the Poisson or negative binomial. The encounter histories are then tabulated at each site so that they can be regarded as an outcome of a multinomial distribution with cell probabilities determined by capture probability (see next section for details). Assuming a Poisson distribution, the model can be written as

$$N_i \sim \text{Poisson}(\lambda)$$
$$\mathbf{y_i} | N_i \sim \text{Multinomial}(N_i, \pi(p)) \tag{1}$$

In the above, λ is the expected number of individuals at each site. $\mathbf{y_i}$ is a vector containing the number of individuals with encounter history $k; k = 1, 2, ...K$ at site i. The number of observable encounter histories K depends on the sampling protocol. For a capture-recapture study with 2 time periods, K equals 3 because the possibilities are $(11, 10, 01)$. In Equation 1, $\pi(p)$ is a function that that converts capture probability (p) to multinomial cell probabilities, *i.e.*, the proportion of individuals expected to have capture history k. The definition of $\pi(p)$ is also specific to the sampling protocol. For example, the cell probabilities corresponding to the capture histories listed above are

$$\pi(p) = \{p^2, p(1-p), (1-p)p, (1-p)^2\}.$$

Note that the last multinomial cell probability corresponds to the probability of not being captured.

Spatial variation in abundance can be modeled using covariates with a log-link function

$$\exp(\lambda_i) = \beta_0 + \beta_1 x_i$$

---

[1]Sometimes animals are not actually marked when using these methods, but they are treated as though they are

where x_i is some site-specific covariate such as habitat type or elevation. A more general form is written as

$$\exp(\lambda_i) = X_i'\beta$$

where X is a design matrix and β is a vector of coefficients, possibly including an intercept. Capture probability can be modeled in the same way using the logit instead of the log link. For instance, we could have

$$logit(p_{ij}) = \alpha_0 + \alpha_1 v_{ij}$$

where v_{ij} is some covariate specific to the site and capture occasion.

2 Data

As previously mentioned, the data required by unmarked are an $R \times K$ matrix in which each row is the vector of tabulated encounter histories for animals captured at some site. Capture-recapture data, however, is typically recorded in the format shown in Table 1.

Table 1: Capture-recapture data for 3 individuals. There were 3 trapping occasions

Animal ID	Site	Capture history
GB	A	101
YR	A	101
RO	A	111
PP	A	100
GY	B	100
PR	B	010

In the absence of individual covariates, these data can be collapsed and formatted as shown in Table **??**.

Table 2: Capture-recapture data from Table 1 in the format required by unmarked. Note that no animals were captured in sites C or D.

Site	100	010	001	110	011	101	111
A	1	0	0	0	0	2	1
B	1	1	0	0	0	0	0
C	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0

3 Analysis in unmarked

3.1 Closed population capture-recapture models

```
> alfl.capRecap <- read.csv(system.file("csv", "alfl.capRecap.csv",
      package = "unmarked"), row.names = 1)
> names(alfl.capRecap)
 [1] "visit1_001" "visit1_010" "visit1_011" "visit1_100" "visit1_101"
 [6] "visit1_110" "visit1_111" "visit2_001" "visit2_010" "visit2_011"
[11] "visit2_100" "visit2_101" "visit2_110" "visit2_111" "visit3_001"
[16] "visit3_010" "visit3_011" "visit3_100" "visit3_101" "visit3_110"
[21] "visit3_111" "struct"     "woody"      "time.1"     "time.2"
[26] "time.3"     "date.1"     "date.2"     "date.3"

> crPiFun <- function(p) {
      cbind((1 - p[, 1]) * (1 - p[, 2]) * p[, 3], (1 -
          p[, 1]) * p[, 2] * (1 - p[, 3]), (1 - p[, 1]) *
          p[, 2] * p[, 3], p[, 1] * (1 - p[, 2]) * (1 -
          p[, 3]), p[, 1] * (1 - p[, 2]) * p[, 3], p[,
```

```
           1] * p[, 2] * (1 - p[, 3]), p[, 1] * p[, 2] *
           p[, 3])
 }
> p <- matrix(0.4, 2, 3)
> crPiFun(p)
       [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
[1,] 0.144 0.144 0.096 0.144 0.096 0.096 0.064
[2,] 0.144 0.144 0.096 0.144 0.096 0.096 0.064
> rowSums(crPiFun(p))
[1] 0.784 0.784
```

obsToY needs to be a matrix with the number of rows equal to the number of columns for some obsCov, and the number columns equal to the number of columns in y If obsToY[i,j] is 1, then a missing value in obsCov translates to a missing value in y

```
> o2y <- matrix(1, 3, 7)

> visitMat <- matrix(c("V1", "V2", "V3"), 50, 3, byrow = TRUE)
> visitMat[1, 2] <- NA
> visitMat[2, ] <- NA
> head(visitMat)
     [,1] [,2] [,3]
[1,] "V1" NA   "V3"
[2,] NA   NA   NA
[3,] "V1" "V2" "V3"
[4,] "V1" "V2" "V3"
[5,] "V1" "V2" "V3"
[6,] "V1" "V2" "V3"
> library(unmarked)
> umf.cr1 <- unmarkedFrameMPois(y = alfl.capRecap[, 1:7],
      siteCovs = alfl.capRecap[, c("woody", "struct")],
      obsCovs = list(visit = visitMat), obsToY = o2y, piFun = "crPiFun")
> (M0 <- multinomPois(~1 ~ 1, umf.cr1))
Call:
multinomPois(formula = ~1 ~ 1, data = umf.cr1)

Abundance:
 Estimate    SE      z P(>|z|)
  0.00714 0.141 0.0504    0.96

Detection:
 Estimate    SE    z P(>|z|)
     1.43 0.216 6.63 3.41e-11

AIC: 258.2382

> (Mt <- multinomPois(~visit ~ 1, umf.cr1))
Call:
multinomPois(formula = ~visit ~ 1, data = umf.cr1)

Abundance:
 Estimate    SE       z P(>|z|)
  -0.0137 0.146 -0.0938   0.925

Detection:
            Estimate    SE      z  P(>|z|)
(Intercept)    1.402 0.371  3.779 0.000157
visitV2        0.292 0.543  0.538 0.590599
visitV3       -0.248 0.499 -0.496 0.619611

AIC: 250.449
```

4 Capture-recapture models allowing for temporary emigration

5 Individual Heteogeneity in Capture Probability

The model of xxx assumes that variation in capture probability can be explained by site, time, or behavioral factors. Individual heterogeneity is not allowed, although, for example, sex-specific differences can be studied by analyzing the data from the two sexes seperately. Continuous animal-specific covariates, however, cannot be considered in unmarked. The so-called model M_h, which assumes random variation in capture probability among individuals, is also not allowed. Some might find comfort in this given the concerns about M_h raised by Link (2003!!CITE).

6 Spatially-explicit Capture-recapture Models

Another source of individual heterogeneity in capture probability arises from the distance between an individuals area of activity and the trap location. Recently developed spatial capture-recapture models, however, make use of the trap location data to model density and distance-related heterogeneity in capture probability. SCR models thus offer important advantages over traditional methods and should be used when possible if the assumptions are deemed reasonable. See xyz for more information.

References

[1] Ian Fiske and Richard Chandler. **unmarked**: An **R** package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, 43(10):1–23, 2011.

[2] J. A. Royle. Generalized estimators of avian abundance from count survey data. *Animal Biodiversity and Conservation*, 27(1):375–386, 2004.

[3] B K Williams, J D Nichols, and M J Conroy. *Analysis and management of animal populations: modeling, estimation, and decision making.* Academic Pr, 2002.