

# Capture-recapture models in `unmarked`

Richard Chandler

January 27, 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<sup>1</sup>. This document focuses on the analysis of capture-recapture data using a class of models known as multinomial  $N$ -mixture models [2, 1], which can be fit using the functions `multinomPois` and `gmultmix`. 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. `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

Capture-recapture models have a long history in ecology and are used to estimate population size while controlling for variation in capture probability [3]. Traditional capture-recapture models, however, do not allow one to model spatial variation in abundance—a central objective of much ecological research. (author?) [2] proposed a simple hierarchical model to overcome this limitation. The model requires that  $R$  “sites” are surveyed using capture-recapture methods, and that site-specific abundance ( $N_i, i = 1, 2, \dots, R$ ) can be modeled as a discrete random variable. For example, if assert that that abundance is Poisson distributed, we can describe the model as

$$\begin{aligned} N_i &\sim \text{Poisson}(\lambda) \\ \mathbf{y}_i | N_i &\sim \text{Multinomial}(N_i, \pi(p)) \end{aligned} \tag{1}$$

In the above,  $\lambda$  is the expected number of individuals at each site.  $\mathbf{y}_i$  is a vector containing the number of individuals with encounter history  $k, k = 1, 2, \dots, K$  at site  $i$ . The number of possible encounter histories  $K$  depends on the sampling protocol.  $\pi(p)$  is a function 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, in removal sampling<sup>2</sup>, if there are 3 removal passes, the function is

$$\pi(p) = \{p, (1-p)p, (1-p)^2p, (1-p)^3\}.$$

These 4 multinomial cell probabilities correspond to the 4 possible encounter histories:  $H = (100, 010, 001, 000)$ . The probability in  $\pi(p)$  corresponds to the probability of not being captured. This form of  $\pi(p)$  can be specified in `unmarked` by setting the argument `piFun` to ‘removal’ when formatting the data with the `unmarkedFrame` function. Another canned option is for double observer sampling, `piFun`=‘double’. At some point in the near future, we may offer a capture-recapture `piFun`. However, the purpose of this document is to describe how users can do so in a transparent way so that they can provide custom functions to suit their needs.

Before we describe how to format data, it is important to mention that, as with all the abundance models in `unmarked`, spatial variation in abundance can be modeled using covariates with a log-link function

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

---

<sup>1</sup>Sometimes animals are not actually marked when using these methods, but they are treated as though they are

<sup>2</sup>a simple type of capture-recapture sampling

where  $x_i$  is some site-specific covariate such as habitat type or elevation say. Note that multiple covariates, continuous or categorical, can be considered as covariates of  $\lambda$ . Similarly, if we had a single covariate of  $p$ , we could model it using a logit link

$$\text{logit}(p_{ij}) = \alpha_0 + \beta_1 v_{ij}$$

where  $v_{ij}$  is some covariate that may be specific to the site and observation.

## 2 Data

Earlier we mentioned that the data  $\mathbf{y}$  must be a  $R \times K$  matrix in which each row is the vector of tabulated encounter histories for animals captured at some site. A site can be defined in many different ways. For example, a site could be a wetland where an array of live traps are placed to capture wood turtles. Or a site could be a point count plot where we collect capture-recapture data on birds by dividing the survey into time intervals and recording which intervals each bird is detected in.

The raw data that one collects in a capture-recapture study is typically of the form shown in Table 1. Each row of the matrix corresponds to a capture of some individual. Associated with this capture we need to know the capture occasion, the site at which it was captured, and perhaps some covariates. Note that if we have multiple traps at a site, as is often the case, traditional capture-recapture, and the models considered here, models ignore this information. 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.

Simple capture-recapture data is shown in Table.

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

SHOW TYPICAL MULTINOMIAL DATA

## 3 Sampling methods

Although these methods require that individual animals can be assigned to distinct categories, these methods are often used when individuals are unmarked. For example, a common study design divides a 10-minute point count into two or more time intervals, and observers note the time periods during which each bird was detected. Thus, keeping track of individuals during a small time frame is similar to monitoring marked individuals over longer durations.

## 4 Capture-recapture Models

### 4.1 Model $M_0$

### 4.2 Model $M_t$

### 4.3 Model $M_b$

### 4.4 Model $M_t$

### 4.5 Model $M_h$

Cite Royle and Dorazio pg 173 “In the absence of individual effects on  $p$ , the individual encounter histories can be pooled into groups of unique encounter histories, indexed by  $h$ , a unique combination of zeros and ones.”.

**unmarked** does not, however, allow for the modeling of individual heterogeneity in capture probability although data cou

## 5 Analysis in unmarked

## 6 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) { # p should have 3 columns
  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")
> (MO <- 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

```

## 7 Capture-recapture models allowing for temporary emigration

### 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.