

A Technical Guide for Monitoring Wildlife Populations__draft

Part 4: Estimating abundance

Lukmann Haqem

24 June, 2022

Contents

Estimating abundance	1
Example	1
Fitting a simple N-mixture model	3
Fitting advanced N-mixture models	7
Model assumptions	7
Additional resources	8
References	8

Estimating abundance

The conventional way to estimate species abundance is the mark-recapture framework. This framework can be a costly endeavor and requires technical knowledge and specialized equipment to begin with. Moreover, mark-recapture routines such as live-trapping and marking individuals could be detrimental to sensitive species if not performed correctly. On the other hand, identifying individuals of the target species can be challenging in most cases. This could be for several reasons, such as the animal's body pattern, or it could be just an error on the part of the observer. What it comes to is that animals typically do not want to be seen by humans, so our observed abundance and the actual abundance at the site are probably going to be far off.

Given these limitations, the N-mixture analysis framework was developed as an alternative. The technique is very similar to estimating abundance from mark-recapture data but does not require any marking of animals. The sampling method under the N-mixture framework is also much simpler; The number of individuals detected at a site is counted using standard monitoring techniques (e.g., point counts, plots, transects, etc.). Each site is generally surveyed multiple times. The repeated counts at a site are then used to estimate the detection probability and abundance jointly.

The N-mixture analysis framework could provide a reliable abundance estimate comparable to the conventional approaches. It offers additional advantages such as a smaller sampling effort and no need to trap or manipulate animals, which reduces the risk of harming the observers or the animals and disease transmission. Because they are relatively inexpensive and easy to carry out in the field, such a framework enables researchers to collect abundance information over large areas for the broad-scale assessment of populations. Also, the nice thing about estimating abundance under this N-mixture framework that we will develop here is that occupancy (covered later) comes for free.

Example

In this example, the data set you will use is the Rhinoceros hornbill (*Buceros rhinoceros*) count collected from the avian survey. This study was conducted at a coupe before the timber harvest. Experienced observers sample irregular transects in 10 1kmx1km quadrats across a coupe three times in a year (April, May & June), during which the number of birds seen by observers is then recorded. The quadrats are placed 2km apart, about the average home range size of the bird (Leighton 1982).



Figure 1: An example of the Rhinoceros hornbill and the sampling design that is used to collect the data. (Images - www.wikimedia.org)

You can load the data set from the file named “hornbill1.csv.” This is the kind of data you expect to get if you are doing N-mixture modeling, but it does not matter which order but may be structured like this. Typically in N-mixture models, each data set row is a sample unit. Here, each row represents one quadrat.

The first two columns of the data set are the sites' coordinates (optional), the following two columns are the count data from each sampling occasion, and the columns after that are the covariates. You also have two types of covariates; *elevation* and *forest* describe the sites, and *dur*- and *day*- are sampling covariates describing changes in detection during each session.

```
#load the data set
hornbill <- read.csv("hornbill1.csv", header = TRUE)
head(hornbill)
```

```
##   northing easting occ1 occ2 occ3 elevation forest dur1 dur2 dur3 day1 day2
## 1    65276 1007942    6    6    5      1830     82  320  295  310   60   72
## 2    192276 987942    0    0    0       650      0  240  210  170   22   45
## 3    221276 1163942    2    1    0      1140     50  165  165  180   25   46
## 4    112276 1256942    0    0    0       550     67  300  300  300   29   38
## 5     71276  928942    0    0    0       510      2  210  225  235   28   56
## 6    150276 1250942    0    0    0       830     42  300  315  295   21   56
##   day3
## 1    74
## 2    56
## 3    70
## 4    50
## 5    73
## 6    72
```

You can have a quick look of the data set. The data set have 10 rows in total.

```
nrow(hornbill)
```

```
## [1] 10
```

Before embarking on modeling, start by performing initial investigations on data to discover any patterns and spot anomalies. This step will also give you a rough indication of the abundance. You can look at the detection of the quadrats, and you can see a relatively small variation over time. For instance, at the third quadrat, only the first and second occasions had detected birds but not the last occasion. So this means that detection is not perfect, and we do not have a 100% chance of seeing birds each time we go to survey.

```
y <- hornbill[, 3:5]
y
```

```
##   occ1 occ2 occ3
## 1     6     6     5
## 2     0     0     0
## 3     2     1     0
## 4     0     0     0
## 5     0     0     0
## 6     0     0     0
## 7     1     0     0
## 8     0     0     0
## 9     0     0     0
## 10    2     4     1
```

You can have a look at the number of detection on different occasions. You can see on the first occasion (April), that you detected 11 birds, another 11 birds on the second occasion (May), and 6 birds on the third occasion (June). You can also look at the spatial variation in detection by summing the rows. Depending on the quadrat, you have a relatively strong variation where in some quadrats, you detected 0 birds, while in some quadrats, you detected up to 17 birds in total.

```
colSums(y)
```

```
## occ1 occ2 occ3  
##    11    11     6
```

```
rowSums(y)
```

```
## [1] 17  0  3  0  0  0  1  0  0  7
```

Fitting a simple N-mixture model

To fit a N-mixture model, you will use the **unmarked** package in R. This package is not part of the standard installation of R. You will have to install it first and then load it into your R library before you can start building models with it.

```
#load R package  
install.packages("unmarked")  
library(unmarked)
```

N-mixture models (Royle 2004) allow you to estimate the abundance of a species while taking into account imperfect detection. A simple N-mixture model estimates two parameters:

- *abundance*, N : average abundance at a site for the species. This is also called the “state” process.
- *detection probability*, p : probability that an individual of the species is detected if present at a site during sampling. This is also called the “detection” process.

First, your data set needs to be re-organised using the `unmarkedFramePCount()` function so that it is readable by **unmarked**. The **unmarkedFrame**’s are a special class of objects used in **unmarked** for organising data. The benefit of this is a formal definition and validity checking to prevent accidentally breaking it, but the consequence is a slightly different syntax for examining it. For an N-mixture analysis, the input data fall into three categories:

- *count data*, c : A data frame of count records. Rows are sites and columns are repeat surveys.
- *site covariates data*, $siteCovs$: A data frame of the site-level covariates. These are things that do not change between surveys like elevation, transect length, etc.
- *observation covariates data*, $obsCovs$: A list of data frames for the observation-level covariates. Each covariate is its own data frame with rows as sites and columns as repeat surveys, and these are things that can change between surveys like sampling duration and date, etc.

```
#re-organise input data
umf <- unmarkedFramePCount(
  y = hornbill[, c("occ1", "occ2", "occ3")],
  siteCovs = data.frame(elevation = hornbill$elevation, forest = hornbill$forest),
  obsCovs = list(day = hornbill[, c("day1", "day2", "day3")],
                 duration = hornbill[, c("dur1", "dur2", "dur3")]))
summary(umf)
```

```
## unmarkedFrame Object
##
## 10 sites
## Maximum number of observations per site: 3
## Mean number of observations per site: 3
## Sites with at least one detection: 4
##
## Tabulation of y observations:
##  0  1  2  4  5  6
## 21  3  2  1  1  2
##
## Site-level covariates:
##      elevation      forest
##  Min.   : 510    Min.   : 0.00
##  1st Qu.: 575    1st Qu.:20.25
##  Median : 985    Median :45.00
##  Mean   : 995    Mean   :42.40
##  3rd Qu.:1272    3rd Qu.:62.75
##  Max.   :1830    Max.   :84.00
##
## Observation-level covariates:
##      day      duration
##  Min.   :18.00    Min.   :135.0
##  1st Qu.:30.25    1st Qu.:210.0
##  Median :52.50    Median :250.0
##  Mean   :50.90    Mean   :244.3
##  3rd Qu.:67.75    3rd Qu.:300.0
##  Max.   :92.00    Max.   :320.0
```

Now you can fit a model. Doing an N-mixture model uses the `pcount()` function in `unmarked`, and it requires a double right-hand side formula describing covariates for detection and abundance, in that order. Essentially *~detection formula ~abundance formula*. The argument `K` of the function is simply a “tuning parameter” and should be set high enough so that it does not affect the parameter estimation. Let us fit the simplest model, `m1`. This model assumes neither the detection nor abundance change during the study (`~1` means no covariate effects).

```
#fitting input data to a model
m1 <- pcount(~1 ~1, K = 100, data = umf)
```

You can check the summary of the model you fitted. The `pcount()` function returns the estimates for the two components of the N-mixture model: average detection probability and abundance — the estimates of the detection probability and abundance as logit and log scales, respectively. If you want to have these values on their natural scale, you need to use the `plogis()` and `exp()` functions to convert the estimates back to their real values (Figure 2). After conversion, you should get an estimate of the detection probability of an individual about 0.632, and the average abundance at a site of approximately 1.477 individuals.

```
#check the model summary
summary(m1)
```

```
##
## Call:
## pcount(formula = ~1 ~ 1, data = umf, K = 100)
##
## Abundance (log-scale):
##   Estimate    SE    z P(>|z|)
##      0.39 0.288 1.35  0.176
##
## Detection (logit-scale):
##   Estimate    SE    z P(>|z|)
##      0.541 0.459 1.18  0.239
##
## AIC: 76.52631
## Number of sites: 10
## optim convergence code: 0
## optim iterations: 12
## Bootstrap iterations: 0
```

```
#calculate p on "natural" scale (probability)
p <- plogis(coef(m1[2]))
p
```

```
##   p(Int)
## 0.6319672
```

```
#calculate N on "natural" scale (count)
N <- exp(coef(m1[1]))
N
```

```
## lam(Int)
## 1.47686
```

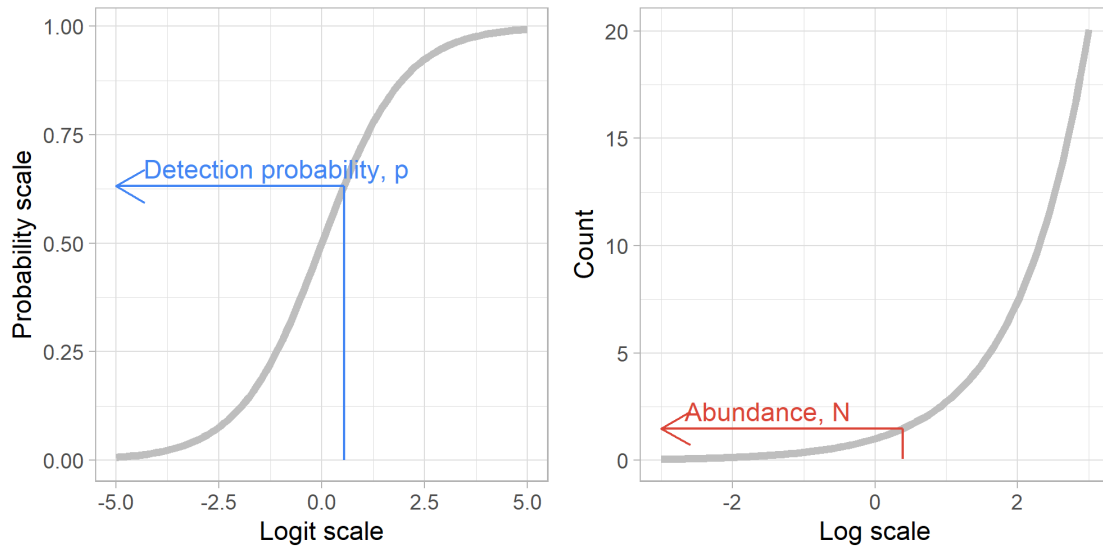


Figure 2: Scale conversion with the logit and log links.

You can also calculate the 95% confidence intervals for each estimate. The 95% confidence intervals are intervals constructed to have a 95% chance of including the true value i.e., the uncertainty is approximately a 95% confidence interval.

```
#calculate 95% confidence intervals for p
p.ci <- plogis(confint(m1, type = "det"))
p.ci
```

```
##           0.025    0.975
## p(Int) 0.4113578 0.808406
```

```
#calculate 95% confidence intervals for N
N.ci <- exp(confint(m1, type = "state"))
N.ci
```

```
##           0.025    0.975
## lam(Int) 0.8394862 2.598154
```

So, you have done a simple analysis with the N-mixture model. The final step is communicating your findings. Many readers will only look at your report without reading the main text. Therefore, ensure your results stand out from the text and display your most significant results. Figures and tables are often the quickest way to share large amounts of complex information that would be complicated to explain in text. Below are basic examples to describe your survey findings to the readers. You should include error bars if you have measurement error information.

As words:

...based on the N-mixture model, the average detection probability of the Rhinoceros hornbill when present at a site was 63.2% (95% CI: 41.1–80.8), while the species' average abundance at a site was 1.477 (95% CI: 0.839-2.598) individuals...

As table:

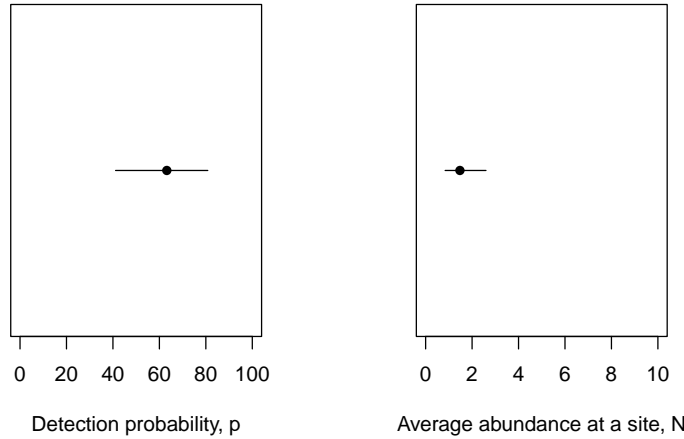
Species	Detection probability, p	Average abundance at a site, N
Rhinoceros hornbill	63.2% (95%CI: 41.1–80.8)	1.477 (95%CI: 0.839–2.598)

As graph:

```
#create one row of plots (side-by-side)
par(mfrow = c(1, 2))

#plot p
plot(p*100, 1, pch = 16, xlim = c(0, 100), yaxt = "none", las = 1,
     xlab = "Detection probability, p",
     ylab = " ")
segments(p.ci[1]*100, 1, p.ci[2]*100, 1)

#plot psi
plot(N, 1, pch = 16, xlim = c(0, 10), yaxt = "none", las = 1,
     xlab = "Average abundance at a site, N",
     ylab = " ")
segments(N.ci[1], 1, N.ci[2], 1)
```



Fitting advanced N-mixture models

This section will walk you through some extensions of the N-mixture analysis. Because this section is very long and involved, it will not be included in this book. If you are interested to learn about it, please use the link here (<https://github.com/lukmannhaqem/A-Technical-Guide-for-Monitoring-Wildlife-Populations>). Click on “Code” button and then “Download ZIP” to download the files.

Model assumptions

The five main assumptions of the N-mixture model are:

- **Population is “closed” to changes in abundance.** Abundance at site does not change between successive occasions (i.e. no births, deaths, immigration, and emigration), but it can change between sampling seasons.
- **Sites are independent.** Abundance at a site is independent of abundance at other sites. This might be a problem if your sites are closely spaced, allowing individuals to move among sites and be detected at multiple sites.
- **No double-counting.** Observers counted individuals only once per survey.
- **Individuals detection is independent of each other.** All individuals have the same detection probability.
- **Abundance and detection at a site are Poisson and binomial distributed.** Poisson and binomial distributions are true descriptions of state/detection processes.

Additional resources

- Vignette by the authors (<https://cran.r-project.org/web/packages/unmarked/vignettes/unmarked.pdf>)
- An article about the unmarked package (<https://core.ac.uk/download/pdf/26859391.pdf>)
- unmarked Google group/forum (<https://groups.google.com/g/unmarked>)
- An overview of N-mixture modeling by Nathan Rudd (https://www.lcrmscp.gov/crtr/presentations/2017/crtr17_19.pdf)
- An example of N-mixture modeling by Javier Fernandez-Lopez (<https://jabiologo.github.io/web/tutorials/nmixture.html>)

References

- Leighton, Mark. 1982. “Fruit Resources and Patterns of Feeding, Spacing and Grouping Among Sympatric Bornean Hornbills (Bucerotidae).” PhD thesis, University of California, Davis.
- Royle, J Andrew. 2004. “N-Mixture Models for Estimating Population Size from Spatially Replicated Counts.” *Biometrics* 60 (1): 108–15.