Occupancy models in R

Occupancy models

Basic principles

- Repeated visits (K) to a collection of sites (J) where species is detected (y_{jk}=1) or not detected (y_{jk}=0)
- True occupancy state of a site j, z_j, does not change during study
- A site is occupied with probability ψ_j

$$z_j \sim Bernoulli(\psi_j)$$

- Occupancy probability can be modeled as a function of covariates

$$logit(\psi_j) = \beta_0 + \beta_1 * X.\psi_j$$

- Because detection is imperfect (p<1), we may fail to detect the species at a site even though it is there
- Detection probability can vary as a function of occasion and/or site level covariates

$$y_{jk} \sim Bernoulli(p_{jk} \times z_j)$$
$$logit(p_{jk}) = \alpha_0 + \alpha_1 * X.p1_j + \alpha_2 * X.p2_k$$

Occupancy models in R with “unmarked”

- Create a desktop folder called “Lab 6”
- Open Rstudio, set the working directory to “Lab 6”, open a new script
- Install package “unmarked” and load it into workspace
- Open help page for package “unmarked” – which function will we use in this lab?

Willow tit data

- Data from the Swiss survey of common breeding birds
- Sample units are 1 km^2 quadrats that are surveyed 2-3 times each
- Each survey has variable length (the route the observer walks within the quadrat can be quite variable), sampled over the course of 100 days
- Forest and elevation are recorded for each location

Read in willow tit data

- From Canvas, download file “wtmatrix.csv”, save it in Lab 6 folder
- Read file into R; assign its content to object called “dat”
- Look at top rows of dat

- y.1, y.2, y.3: observations (0 or 1) for visit 1, 2, 3
- elev, forest: elevation and forest cover for each site

→ Habitat and elevation can affect whether a species occurs

- day.1, day.2, day.3: Day of the year for each visit

→ Day of year can affect detection

- length: length of survey route

→ Can affect detection

Set up data for occu()

- We need to create an unmarkedFrameOccu object
- First, set up observations in a site-by-occasion matrix
- Then set up site level covariates in a data frame
- Last, set up occasion level covariates in a list of data frames
- Important: Occasion level covariates are different for each occasion

→ multiple columns within a data frame refer to the same covariate

→ That's why we need multiple data frames, combined in a list, if we have multiple occasion level covariates

- Site-level covariates do not vary across occasions → each covariate is only a single column

Exploring the data

Basic summary statistics:

- How many sites were sampled?
- How many occasions?
- Range of covariate values?
- At how many sites was the species detected?

- Use apply() to determine whether species was detected at least once for each of the sites

apply(X, MARGIN, FUN)

- X is a matrix
- MARGIN is the dimension over which you want to perform a calculation
- FUN is the function you want to carry out

Exploring the data

- At how many sites was the species detected?

```
> obs.tot<-apply(obs, 1, sum, na.rm=T)
```

- What is the range of total detections? Why?
- Use obs.tot to calculate number of sites where species was detected at least once

- Raw occupancy: number of sites with detection divided by total number of sites

Analyzing data in unmarked

- Use ?occu to determine function arguments
- occu(formula, data , …)
- Formula: ~detection ~occupancy
- Data: umf

- Run occupancy model without any covariates and look at the output

Interpreting model output

- Results of the “null model” – no covariates on detection or occupancy

```
Occupancy (logit-scale):
 Estimate    SE     z  P(>|z|)
   -0.665 0.139 -4.77 1.82e-06
```

$$z_j \sim Bernoulli(\psi_j)$$
$$\boxed{logit(\psi_j)} = \beta_0$$

```
Detection (logit-scale):
 Estimate    SE    z  P(>|z|)
     1.32 0.174 7.61 2.82e-14
```

$$y_{jk} \sim Bernoulli(p_{jk} \times z_j)$$
$$\boxed{logit(p_{jk})} = \alpha_0$$

Back-transforming model estimates

- Estimates are given on the logit-scale
- We can use the inverse-logit to calculate occupancy and detection probability on their natural/real scale (between 0 and 1) manually

```
> plogis(-0.665)
> plogis(1.32)
```

- Alternatively, we can use the backTransform() function
- Why use backTransform?
  - Because it automatically chooses the right mathematical function
  - Because it also back-transforms standard errors, which are harder to obtain manually
  - Note: this only works for models without covariates
  - For models with covariates, use predict()

Interpreting output form models with covariates

- Run model with covariate “elevation” on occurrence and “length” on detection
- Bring up summary results

Interpreting output form models with covariates

- Run model with covariate “elevation” on occurrence and “length” on detection
- Bring up summary results

```
Occupancy (logit-scale):
            Estimate       SE     z  P(>|z|)
(Intercept)  -3.3813 0.583940 -5.79 7.02e-09
elev          0.0025 0.000554  4.51 6.39e-06
```

$$z_j \sim Bernoulli(\psi_j)$$

$$logit(\psi_j) = \boxed{\beta_0} + \boxed{\beta_1} Elevation$$

Coefficient for
(effect of) elevation

Interpreting output form models with covariates

- Run model with covariate “elevation” on occurrence and “length” on detection
- Bring up summary results

```
DDetection (logit-scale):
            Estimate    SE     z  P(>|z|)
(Intercept)    -1.71 0.719 -2.38 0.017318
length          0.51 0.151  3.38 0.000724
```

$$y_{jk} \sim Bernoulli(p_{jk})$$
$$logit(p_{jk}) = \boxed{\alpha_0} + \boxed{\alpha_1} * Length$$

Coefficient for (effect of) length

Plotting relationships with covariates

- Create a range of possible values for the covariates
- Use predict() to obtain expected value of psi/p
- Plot these against the new covariate values
- Add confidence intervals to the plot

Model selection

Strategy:

- Using elevation and forest as covariates on occupancy, build models with no detection covariate, length, day, and length+day as covariates on detection (4 models total)
- Collect models in a fitlist object and compare with AIC to determine the top model

```
> modList<-fitList(mod1, mod2, …)
> modSel(modList)
```

- Which is the top model?

Interpreting results – one more time

Occupancy component

```
Occupancy (logit-scale):
             Estimate       SE     z  P(>|z|)
(Intercept) -6.06601 0.674387 -8.99 2.37e-19
elev         0.00256 0.000448  5.72 1.06e-08
forest       0.05212 0.007087  7.36 1.91e-13
```

- What is the probability that the willow tit occurs at a site with 0 forest cover and 0 elevation?
- In words, what is the specific relationship of occupancy probability with elevation, forest?

Interpreting results – one more time

Occupancy component

```
Detection (logit-scale):
            Estimate      SE      z P(>|z|)
(Intercept) -0.83568 1.12116 -0.745   0.456
length       0.33824 0.20967  1.613   0.107
day          0.00827 0.00864  0.957   0.338
```

- What is the probability that the willow tit is detected during a visit to a site where it occurs, with 0 survey length at day 0 (January 1)?
- In words, what is the specific relationship of detection probability with survey length, day of the year?