Estimating abundance from occupancy data: the Royle-Nichols model

Martin Ridout

School of Mathematics, Statistics and Actuarial Science

Basic idea

Detection probability should depend on abundance.

Since (repeated) occupancy surveys allow us to estimate detection probability, this should also provide information about abundance.

Basic idea

Detection probability should depend on abundance.

Since (repeated) occupancy surveys allow us to estimate detection probability, this should also provide information about abundance.

WARNING

Results may be sensitive to assumptions that are difficult or impossible to verify.

How does detection probability depend on abundance?

$p =$ detection probability

$N =$number of individuals present at a site

How does detection probability depend on abundance?

$p =$ detection probability

$N =$number of individuals present at a site

Assumptions

1. There is a constant probability r of detecting each individual
2. Detections of individuals are independent

How does detection probability depend on abundance?

$p =$ detection probability

$N =$number of individuals present at a site

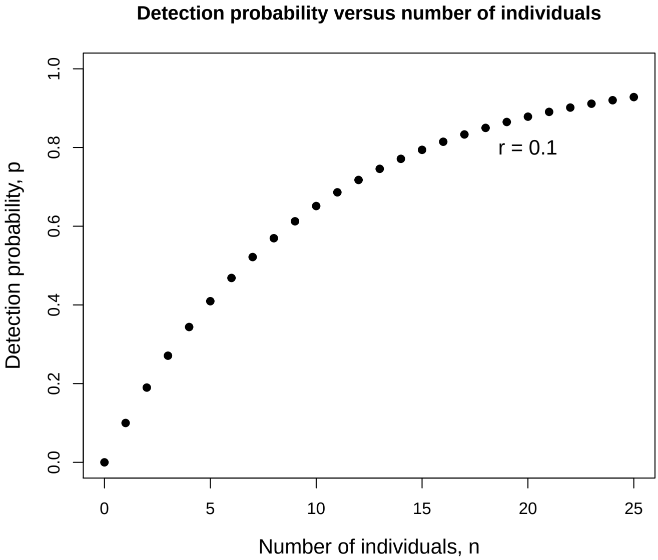
Assumptions

1. There is a constant probability r of detecting each individual
2. Detections of individuals are independent

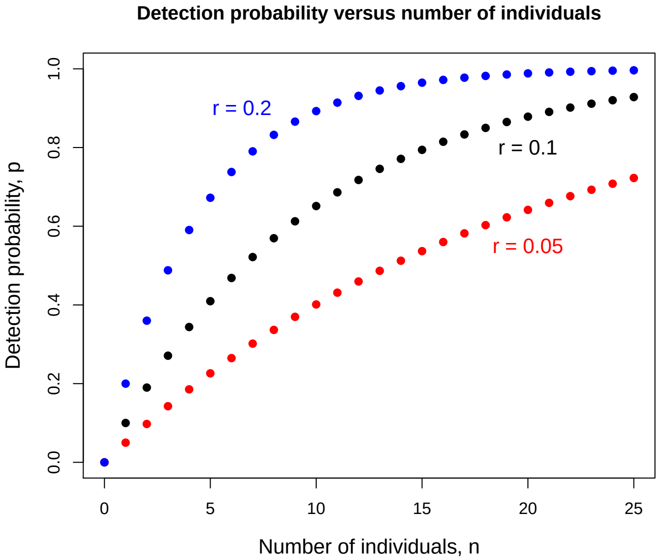
Then it follows that

$$p = 1 - (1 - r)^N.$$

How does detection probability depend on abundance?



How does detection probability depend on abundance?



Number of detections

K = number of replicated surveys

W = number of surveys that detect the species

Number of detections

K = number of replicated surveys

W = number of surveys that detect the species

$$\Pr(W = w) = \binom{K}{w} p^w (1-p)^{K-w}$$

Number of detections

$K =$ number of replicated surveys

$W =$ number of surveys that detect the species

$$\Pr(W = w) = \binom{K}{w} p^w (1-p)^{K-w}$$

Given the value of $N = n$ we can substitute

$$p = 1 - (1 - r)^n$$

to get a conditional probability

$$\Pr(W = w | N = n) = \text{function of } K,\ w,\ r \text{ and } n.$$

Key idea of Royle-Nichols model

$$\Pr(W = w | N = n) = \text{function of } K,\ w,\ r \text{ and } n$$

Key idea of Royle-Nichols model

$$\Pr(W = w | N = n) = \text{function of } K,\ w,\ r \text{ and } n$$

Assume that N varies between sites according to some probability distribution.

Key idea of Royle-Nichols model

$$\Pr(W = w | N = n) = \text{function of } K,\ w,\ r \text{ and } n$$

Assume that N varies between sites according to some probability distribution.

Then we can work out the unconditional probability $\Pr(W = w)$ using the law of total probability

$$\Pr(W = w) = \sum_{n=0}^{\infty} \Pr(W = w | N = n) \times \Pr(N = n).$$

Key idea of Royle-Nichols model

$$\Pr(W = w|N = n) = \text{function of } K, w, r \text{ and } n$$

Key idea of Royle-Nichols model

$\Pr(W = w | N = n) =$ function of K, w, r and n

$\Pr(W = w) =$ function of K, w, r + additional parameters defining the distribution of N

Key idea of Royle-Nichols model

$$\Pr(W = w | N = n) = \text{function of } K,\ w,\ r \text{ and } n$$

$$\Pr(W = w) = \text{function of } K,\ w,\ r + \text{additional parameters defining the distribution of } N$$

So the parameters in the model that we need to estimate are r and the parameters defining the distribution on N.

Key idea of Royle-Nichols model

$$\Pr(W = w|N = n) = \text{function of } K,\ w,\ r \text{ and } n$$

$$\Pr(W = w) = \text{function of } K,\ w,\ r + \text{additional parameters defining the distribution of } N$$

So the parameters in the model that we need to estimate are r and the parameters defining the distribution on N.

EXAMPLE: If N has Poisson distribution with mean μ

$$\Pr(N = n) = \frac{e^{-\mu}\mu^n}{n!}.$$

Key idea of Royle-Nichols model

$$\Pr(W = w | N = n) = \text{function of } K,\ w,\ r \text{ and } n$$

$$\Pr(W = w) = \text{function of } K,\ w,\ r + \text{additional parameters defining the distribution of } N$$

So the parameters in the model that we need to estimate are r and the parameters defining the distribution on N.

EXAMPLE: If N has Poisson distribution with mean μ

$$\Pr(N = n) = \frac{e^{-\mu}\mu^n}{n!}.$$

and the parameters to estimate are r and μ.

What happened to occupancy parameter?

What happened to occupancy parameter?

Question: Where has ψ gone?

What happened to occupancy parameter?

Question: Where has ψ gone?

Answer: Still there, given that we have a probability distribution for N ...

$$\psi = 1 - \Pr(N = 0)$$

E.g. If N has Poisson distribution with mean μ

$$\psi = 1 - \exp(-\mu)$$

Example: Wood thrush (from R package `unmarked`)

Data for wood thrush (*Hylocichla mustelina*) from North American Breeding Bird survey (1991)

Point counts at 50 locations along a survey route in New Hampshire collected over a 30 day period

11 sampling occasions (same observer)

Count	0	1	2	3	4
Frequency	344	162	40	3	1

Counts not always reliable, so reduce data to detection/no detection

Fitting the basic Royle-Nichols model

```
Model with constant abundance (mu) and
           constant detection probability (r)
AIC: 633.9534
```

Estimate of $\log(\mu)$

```
Abundance:
 Estimate    SE    z P(>|z|)
    0.792 0.158 5.03   5e-07
```

Estimate of $\text{logit}(r)$

```
Detection:
 Estimate   SE     z  P(>|z|)
    -1.21 0.17 -7.14 9.41e-13
```

Fitting the basic Royle-Nichols model

```
Model with constant abundance (mu) and
           constant detection probability (r)
AIC: 633.9534
```

Estimate of $\log(\mu)$ $[\hat{\mu} = \exp(0.792) = 2.21]$

```
Abundance:
 Estimate    SE    z P(>|z|)
    0.792 0.158 5.03   5e-07
```

Estimate of $\text{logit}(r)$ $[\hat{r} = 1/(1 + \exp(1.21)) = 0.23]$

```
Detection:
 Estimate   SE     z  P(>|z|)
    -1.21 0.17 -7.14 9.41e-13
```

Fitting the basic Royle-Nichols model - implications

Number of individuals at a site assumed to follow a Poisson distribution with mean estimated as

$$\hat{\mu} = 2.21$$

In particular, proportion of occupied sites (ψ) estimated as

$$\Pr(N > 0) = 1 - \exp(-2.21) = 0.89$$

(Observed proportion of sites where species detected $= 45/50 = 0.9$)

Fitting the basic Royle-Nichols model - implications

Probability of detecting an individual estimated as

$$\hat{r} = 0.23$$

E.g. With 2 individuals per site, detection probability $= 1 - (1 - r)^2 = 0.41$.

Probability of failing to detect on any of 11 visits to a site with 2 individuals $= 1 - (1 - 0.41)^{11} = 0.003$.

Almost certain to detect species on at least one occasion if present

Empirical Bayes estimates of numbers at each site

Previously we had the conditional probability

$$\Pr(W = w | N = n) = \text{function of } K,\ w,\ r \text{ and } n.$$

Empirical Bayes estimates of numbers at each site

Previously we had the conditional probability

$$\Pr(W = w | N = n) = \text{function of } K,\ w,\ r \text{ and } n.$$

We can use Bayes' Theorem to reverse the conditioning:

$$\Pr(N = n | W = w) = \frac{\Pr(W = w | N = n) \times \Pr(N = n)}{\sum_{k=0}^{\infty} \Pr(W = w | N = k) \times \Pr(N = k)}$$

Empirical Bayes estimates of numbers at each site

Previously we had the conditional probability

$$\Pr(W = w|N = n) = \text{function of } K, w, r \text{ and } n.$$

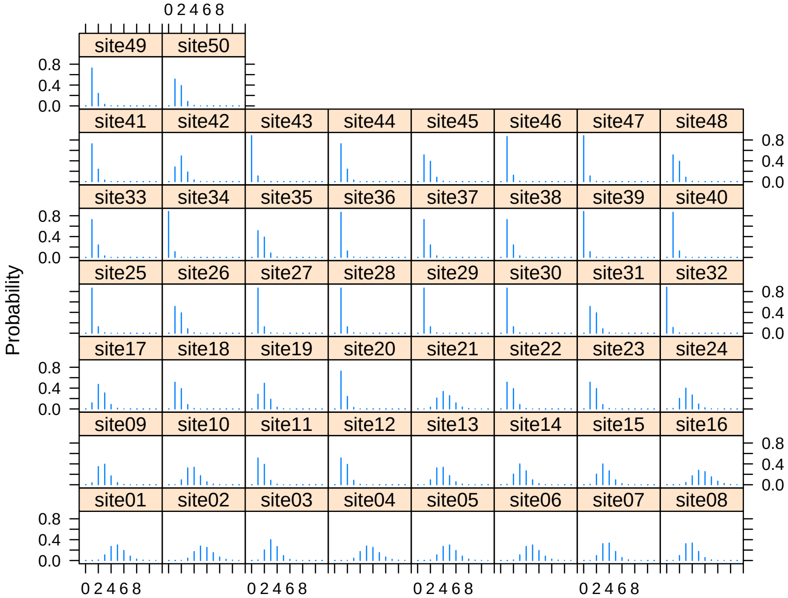
We can use Bayes' Theorem to reverse the conditioning:

$$\Pr(N = n|W = w) = \frac{\Pr(W = w|N = n) \times \Pr(N = n)}{\sum_{k=0}^{\infty} \Pr(W = w|N = k) \times \Pr(N = k)}$$

Once we have fitted the model, we can estimate these probabilities for each site.

The resulting estimators of $\Pr(N = n|W = w)$ are called empirical Bayes estimators.

Empirical Bayes estimates at each site for Woodthrush data



Aside: Back to the point counts

If N varies across sites as Poisson distribution with mean 2.21 and individuals are detected with probability 0.23, then expected distribution of point counts is Poisson with mean $2.21 \times 0.23 = 0.51$.

Aside: Back to the point counts

If N varies across sites as Poisson distribution with mean 2.21 and individuals are detected with probability 0.23, then expected distribution of point counts is Poisson with mean $2.21 \times 0.23 = 0.51$.

Frequency distribution of point counts

Count	0	1	2	3	4+
Observed frequency	344	162	40	3	1
Expected frequency	331.2	168.0	42.6	7.2	1.0

Aside: Back to the point counts

If N varies across sites as Poisson distribution with mean 2.21 and individuals are detected with probability 0.23, then expected distribution of point counts is Poisson with mean $2.21 \times 0.23 = 0.51$.

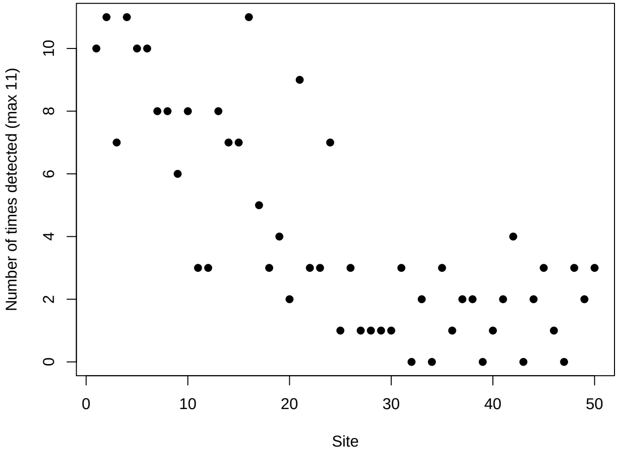
Frequency distribution of point counts

Count	0	1	2	3	4+
Observed frequency	344	162	40	3	1
Expected frequency	331.2	168.0	42.6	7.2	1.0

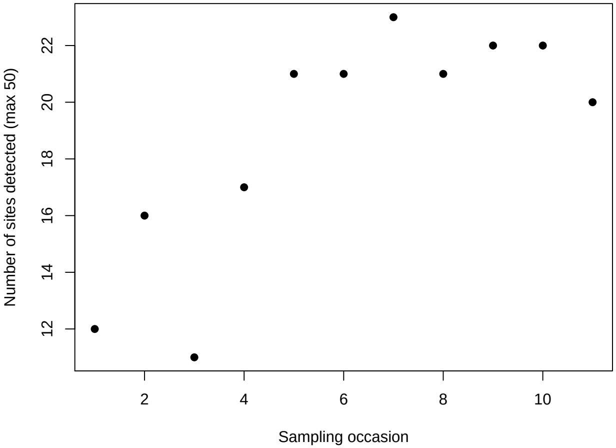
Quite a good fit (chi-squared = 3.32, 3 d.f.)

But . . .

Abundance related to site?



Detection related to sampling visit?



Fitting Royle-Nichols model with covariates

```
Model with abundance (mu) as a function of site number
           detection probability (r) as a function of
           observation number

AIC: 562.0886 (constant model AIC: 633.9534)
```

Abundance:

	Estimate	SE	z	P(>|z|)
(Intercept)	2.9672	0.21465	13.82	1.85e-43
Site number	-0.0676	0.00778	-8.69	3.76e-18

Detection:

	Estimate	SE	z	P(>|z|)
(Intercept)	-2.5607	0.2641	-9.70	3.15e-22
Obs number	0.0791	0.0254	3.11	1.84e-03

Limitations

Assumptions about detectability

1. There is a constant probability r of detecting each individual
2. Detections of individuals are independent

Limitations

Assumptions about detectability

1. There is a constant probability r of detecting each individual
2. Detections of individuals are independent

Assumption of a Poisson distribution for N over sites

Other distributions possible in principle (e.g. negative binomial) but may be hard to fit in practice.

Limitations

Assumptions about detectability

1. There is a constant probability r of detecting each individual
2. Detections of individuals are independent

Assumption of a Poisson distribution for N over sites

Other distributions possible in principle (e.g. negative binomial) but may be hard to fit in practice.

Goodness of fit?