N-mixture models

Emily Dennis, Byron Morgan and Martin Ridout

26 June 2014

Basic Data

Point count data for American redstart

	Sample									
Site	1	2	3	4	5	6	7	8	9	10
1	0	0	0	1	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0
4	0	1	1	3	1	2	2	1	0	1
5	2	0	1	1	0	0	1	0	0	0

Results in an estimated expected number of 2.81, when 50 sites are sampled. Note how small this value is.

The probability distributions used

These are simple, standard discrete distributions:

1. Binomial, Bin(n,p)

$$Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

2. Poisson, Pois (λ)

$$Pr(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

.

3. Negative binomial
   This adds over dispersion to the Poisson.

What the N-mixture model does

- N-mixture models can estimate animal abundance from a set of counts with both spatial and temporal replication whilst accounting for imperfect detection.
- A benefit of the N-mixture model is the reasonably low cost and effort required for data collection compared to alternative sampling methods.
  - Applies for many citizen-science based monitoring programs.
  - Highly cited paper.
- Many extensions exist, including
  - The use of covariates to examine spatial patterns in abundance and detection.
  - The creation of maps of spatial abundance.

Forming the likelihood

A set of counts is made during $t = 1, \ldots, T$ sampling occasions at $i = 1, \ldots, R$ sites of a closed population.

- Each individual has the same detection probability, p.
- The counts n_{it} at site i and time t are

$$n_{it} \sim \text{Bin}(N_i, p),$$

  where N_i is the unknown population size at site i.
- Assuming N_i to be independent random variables with probability function $f(N; \theta)$, such as Poisson or negative binomial, the likelihood is

$$L(p, \theta; \{n_{it}\}) = \prod_{i=1}^{R} \left\{ \sum_{N=\max_t n_{it}}^{\infty} \left(\prod_{t=1}^{T} \mathrm{Bin}(n_{it}; N, p) \right) f(N; \theta) \right\}$$

To infinity and beyond

A set of counts is made during $t = 1, \ldots, T$ sampling occasions at $i = 1, \ldots, R$ sites of a closed population.

- Each individual has the same detection probability, p.
- The counts n_{it} at site i and time t are

$$n_{it} \sim \text{Bin}(N_i, p),$$

  where N_i is the unknown population size at site i.
- Assuming N_i to be independent random variables with probability function $f(N;\theta)$, such as Poisson or negative binomial, the likelihood is

$$L(p, \theta; \{n_{it}\}) = \prod_{i=1}^{R} \left\{ \sum_{N=\max_t n_{it}}^{K} \left(\prod_{t=1}^{T} \mathrm{Bin}(n_{it}; N, p) \right) f(N;\theta) \right\}$$

- Requires selection of a value, K, for ∞.

Computer packages

- PRESENCE
- `unmarked`

Both of these have a default value for K.

As we shall see, this can be dangerous.

It would be nice to avoid having to choose K, and we now show how this can be done.

Equivalence with the multivariate Poisson model

- The N-mixture model is equivalent to the multivariate Poisson with a particular covariance structure.
- The equivalence can be illustrated via comparison of probability generating functions.
- Here we illustrate the bivariate Poisson model ($T = 2$).

The bivariate Poisson

- Consider counts n_1 and n_2 for a particular site and time as

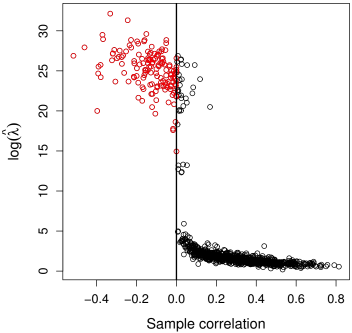
$$n_1 = X_1 + X_{12} \quad \text{and} \quad n_2 = X_2 + X_{12}$$

- $X_1, X_2 \sim \mathrm{Pois}(\theta_1)$, where $\theta_1 = \lambda p(1-p)$, and $X_{12} \sim \mathrm{Pois}(\theta_0)$, where $\theta_0 = \lambda p^2$.
- All values are independent.
- Then (n_1, n_2) follow a bivariate Poisson distribution, with $\mathrm{corr}(n_1, n_2) = p$, and the likelihood is

$$L(p, \lambda; \{n_{it}\}) = e^{-(2\theta_1+\theta_0)} \prod_{i=1}^{R} \left\{ \frac{\theta_1^{n_{i1}+n_{i2}}}{n_{i1}!n_{i2}!} \sum_{m=0}^{\min(n_{i1},n_{i2})} \binom{n_{i1}}{m} \binom{n_{i2}}{m} m! \left(\frac{\theta_0}{\theta_1^2}\right)^m \right\}$$

Performance of the bivariate Poisson model

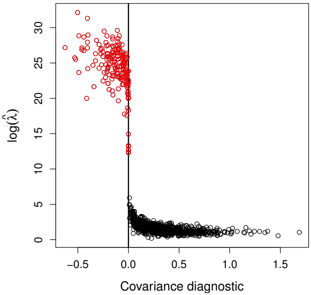
- Investigate model performance via simulation.
  - 1000 simulations for $\lambda = 5$, $p = 0.25$ and $R = 20$.
- In some cases estimates for λ were very large.
  - Associated with small or negative values of the product moment sample correlation.



A covariance diagnostic for $T = 2$

- Negative values of an estimate of the covariance diagnose high estimates of λ.

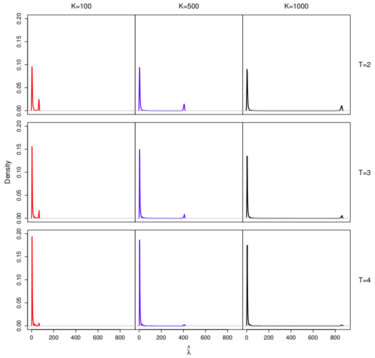
$$\text{cov}^*(n_1, n_2) = \text{mean}(n_1 * n_2) - \{\text{mean}(n_1 + n_2)/2\}^2$$



- Hence in these instances $\hat{\lambda}$ is actually infinite (and $\hat{p} = 0$) and high estimates are obtained as an artefact of the optimisation routine stopping prematurely when the likelihood is flat.
- Simulations suggest that the diagnostic extends for $T > 2$.

Chasing K

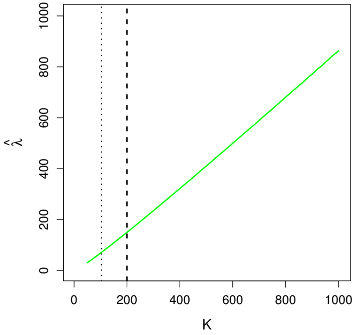
- Investigate model performance via simulation.
    - 1000 simulations for $\lambda = 5,\ p = 0.25$, $T = 2, 3, 4, 5; R = 20$. $K =$100, 200, 500, 1000.



- Estimates from different K overlap for finite $\hat{\lambda}$, but differ when $\hat{\lambda}$ should be infinite and instead approach K.

The effect of the choice of K on fitting the N-mixture model

- The infinite values of $\hat{\lambda}$ for the bivariate Poisson are limited by the value of K in the N-mixture model.
- `unmarked` $K = \max(\text{count}) + 100$ (dotted)
- PRESENCE $K = 200$ (dashed)
- The model should always be fitted for multiple values of K.



covariance diagnostic ≤ 0

Application to Hermann's tortoise data

- Analyse data from a study of the threatened Hermann's tortoise *Testudo hermanni* in southeastern France.
- R=118 sites were each surveyed T=3 times.
- Sample covariance diagnostic suggests stable estimates from the N-mixture model (1.052).
- From the N-mixture model with Poisson mixing distribution $\hat{\lambda} = 4.70$ and $\hat{p} = 0.28$, and were stable for $K \geq 30$.
- But the negative binomial model is better, and estimates using the negative binomial distribution do not stabilise for increasing values of K.
  - Hence results presented for the negative binomial will vary with K.

Hermann's tortoise

Application to Hermann's tortoise data

- Inspect the sample covariance diagnostic for this dataset for a reduced number of sites and/or visits.
  - Taking two of the three visits made at all sites, the diagnostic was always positive (0.97-1.17).
  - Using all visits but a sample of sites, the diagnostic negative for 1.7% and 0% of 1000 samples for $R = 20$ and $R = 50$.
  - For fewer sites and two visits, the diagnostic was negative for 9% and 0.8% of 1000 samples for $R = 20$ and $R = 50$.

Discussion

- The N-mixture model can produce infinite estimates of abundance.
  - Particularly when the number of sampling occasions and/or sites is limited and detection probability low.
- The equivalence with the multivariate Poisson was used to understand and diagnose this behaviour.
  - Avoids the requirement to select an upper bound K.
- Is there a suitable diagnostic for the Negative Binomial distribution?

Recommendations

- We recommend caution when applying this model, especially since detection probability is likely to be low in ecological datasets.
- Good experimental design is important to reduce poor model-fitting behaviour.
  - Distribute effort to maximise the number of visits made to each site.
- We suggest a general strategy for applying the N-mixture model:
  - Always calculate the sample covariance diagnostic to identify datasets where only λp is estimable.
  - if $T \leq 5$ we suggest fitting the multivariate Poisson model. Use of `R` package.
  - if $T > 5$ fit the N-mixture model for increasing values of K.

References

- Couturier, T., Cheylan, M., Bertolero, A., Astruc, G. and Besnard, A. (2013) Estimating abundance and population trends when detection is low and highly variable: a comparison of three methods for the Hermann's tortoise. *The Journal of Wildlife Management*, **77**, 454-462.
- Dennis, E. B., Morgan, B.J.T. and Ridout, M.S. (2014) Computational aspects of N-mixture models. *Biometrics* Invited revision.
- Royle, J.A. (2004) N-mixture models for estimating population size from spatially replicated counts, *Biometrics*, **60**, 108-115.