

Computational Aspects of N-Mixture Models

Emily B. Dennis,* Byron J.T. Morgan, and Martin S. Ridout

School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent CT2 7NF, U.K.

*email: ed234@kent.ac.uk

SUMMARY. The N-mixture model is widely used to estimate the abundance of a population in the presence of unknown detection probability from only a set of counts subject to spatial and temporal replication (Royle, 2004, *Biometrics* **60**, 105–115). We explain and exploit the equivalence of N-mixture and multivariate Poisson and negative-binomial models, which provides powerful new approaches for fitting these models. We show that particularly when detection probability and the number of sampling occasions are small, infinite estimates of abundance can arise. We propose a sample covariance as a diagnostic for this event, and demonstrate its good performance in the Poisson case. Infinite estimates may be missed in practice, due to numerical optimization procedures terminating at arbitrarily large values. It is shown that the use of a bound, K , for an infinite summation in the N-mixture likelihood can result in underestimation of abundance, so that default values of K in computer packages should be avoided. Instead we propose a simple automatic way to choose K . The methods are illustrated by analysis of data on Hermann's tortoise *Testudo hermanni*.

KEY WORDS: Abundance estimation; Method of moments; Multivariate negative binomial; Multivariate Poisson; Optimal design; Sampling; Temporal replication.

1. Introduction

Estimating the abundance of a population is an important component of ecological research. N-mixture models can be used to estimate animal abundance from counts with both spatial and temporal replication whilst accounting for imperfect detection (Royle, 2004a). Whereas alternative sampling methods for obtaining estimates of abundance exist, such as capture–recapture, distance, removal and multiple-observer sampling, these may be expensive in effort or cost, or impractical for some species and scenarios. A benefit of the N-mixture model is the reasonably low comparative cost and effort required for data collection which does not require individuals to be identified. This is especially true of many citizen-science based monitoring programs.

Consequently, since development by Royle (2004a), many applications and extensions of the N-mixture model have been made. These include applications to various taxa, including birds (Kéry, Royle, and Schmid, 2005), mammals (Zellweger-Fischer, Kéry, and Pasinelli, 2011), and amphibians (Dodd and Dorazio, 2004; McIntyre et al., 2012). In addition, covariates have often been used to examine spatial patterns in abundance and detection (Kéry, 2008) and hence create maps of spatial abundance (Royle, Nichols, and Kéry, 2005).

Despite the popularity of the N-mixture model, few studies have made comparisons with estimates derived via alternative methods or undertaken simulation studies of performance (Kéry et al., 2005; Hunt, Weckerly, and Ott, 2012; Couturier et al., 2013). A potential issue for fitting the model using classical inference is the need to specify an upper bound, K , to approximate an infinite summation in the likelihood. We found this matter was rarely mentioned in publications. For example, McIntyre et al. (2012) used simulated data to support their amphibian study, highlighting the benefit of more

sampling occasions, particularly when detection probability was low, however the value of K used was not provided. When software such as `unmarked` (Fiske and Chandler, 2011) written in R (R Core Team, 2014) and `PRESENCE` (Hines, 2011) is used for model fitting, it is possible that only default values of the bound are employed. Couturier et al. (2013) suggest bias could be induced by the choice of K for low detection probabilities.

In this article, we investigate computational aspects of fitting N-mixture models, in particular via a simulation study for scenarios where detection probability is low and/or the number of sampling occasions is small. This may be important for the study of cryptic species, and have implications for sample design: many applications to date have made only three visits, whereas in Royle (2004a) simulations were tested for five visits and an application made to data with 10 visits. When only one sampling visit is made, it is well known that the N-mixture model reduces to a thinned Poisson distribution, with only one estimable parameter, the product of mean abundance and detection probability, a feature which underlies aspects of the work which follows.

The N-mixture model is described in Section 2. In Section 3 we explain the equivalence of the Poisson N-mixture model with a multivariate Poisson distribution. We use this formulation to show that infinite estimates of abundance may arise, and provide a simple diagnostic to identify such cases. The multivariate Poisson formulation has the advantage of not requiring a constant K to be set. Section 4 provides the probability function in the bivariate negative-binomial case. In Section 5, we show how the choice of K in the N-mixture model interacts with the occurrence of infinite estimates of abundance, and how incorrect conclusions may arise. An automatic method for choosing K is provided. Section 6 provides moment

estimates and evaluates the use of two diagnostic tests for the negative-binomial case for when infinite estimates of abundance may arise. Section 7 provides an application to real data and the article ends with discussion and recommendations.

2. The N-Mixture Model

Under the study design in Royle (2004a), a set of counts is made during sampling visits $t = 1, 2, \dots, T$ at $i = 1, 2, \dots, R$ locations (sites). The population is assumed to be closed during the period of sampling and each individual is assumed to have the same detection probability p . The counts n_{it} at site i and time t are assumed to be independent binomial random variables,

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

where N_i is the unknown population size at site i . To fit the model using classical inference, we assume the N_i to be independent random variables with probability function $f(N; \theta)$, and then maximize the likelihood

$$L(p, \theta; \{n_{it}\}) = \prod_{i=1}^R \left\{ \sum_{N_i=\kappa_i}^{\infty} \left(\prod_{t=1}^T \text{Bin}(n_{it}; N_i, p) \right) f(N_i; \theta) \right\}, \quad (1)$$

where $\kappa_i = \max_t n_{it}$. As noted by Royle (2004a), numerical maximization of (1) requires the replacement of the infinite summation over N_i by a sum with upper limit K . The value of K may be selected by fitting the model for a succession of increasing values and selecting K when the parameter estimates appear to stabilize (Royle, 2004a). We shall consider both Poisson and negative-binomial mixing distributions.

It is our experience that the N-mixture model can produce unrealistically large estimates of abundance and we explain this feature in the article.

3. Equivalence of the Poisson N-Mixture Model With a Multivariate Poisson Model

The number of individuals observed at a site at time t can be written as the convolution of independent random variables, corresponding to those seen only once, those seen twice, etc. This natural feature of the N-mixture model can be formalized as we now show.

Let \mathcal{S} denote the set of non-empty subsets of $\{1, \dots, T\}$, and let the random variable $X_{i,s}$ ($s \in \mathcal{S}$) denote the number of individuals seen at site i only on occasion s . For example, $X_{i,124}$ denotes the individuals seen at site i on occasions 1, 2, and 4 only. Then, if we let \mathcal{S}_t denote those elements of \mathcal{S} that include t , we can decompose n_{it} as

$$n_{it} = \sum_{s \in \mathcal{S}_t} X_{i,s}.$$

For example, with $T = 3$, we have

$$\begin{aligned} n_{i1} &= X_{i,1} + X_{i,12} + X_{i,13} + X_{i,123} \\ n_{i2} &= X_{i,2} + X_{i,12} + X_{i,23} + X_{i,123} \\ n_{i3} &= X_{i,3} + X_{i,13} + X_{i,23} + X_{i,123}. \end{aligned}$$

Conditional on N_i , the joint distribution of the set of random variables $X_{i,s}$ ($s \in \mathcal{S}$) is multinomial, with index N_i and probabilities $\pi_{i,s} = p^{|s|}(1-p)^{T-|s|}$, where $|s|$ denotes the number of elements in the set s . When $N_i \sim \text{Pois}(\lambda)$, the $X_{i,s}$ ($s \in \mathcal{S}$) are independent Poisson random variables, with

$$E(X_{i,s}) = \lambda p^{|s|}(1-p)^{T-|s|},$$

see Johnson, Kotz, and Balakrishnan (1997, p. 146). The thinned Poisson is the case $T = 1$.

It follows that the joint distribution of (n_{i1}, \dots, n_{iT}) is multivariate Poisson (Johnson et al., 1997, Chapter 37), with

$$E(n_{ij}) = \sum_{s \in \mathcal{S}_j} E(X_{i,s}) = \sum_{s \in \mathcal{S}_j} \lambda p^{|s|}(1-p)^{T-|s|}.$$

There are $\binom{T-1}{k-1}$ subsets $s \in \mathcal{S}_t$ such that $|s| = k$ ($k = 1, \dots, T$). Hence

$$E(n_{ij}) = \sum_{k=1}^T \binom{T-1}{k-1} \lambda p^k (1-p)^{T-k} = \lambda p.$$

Similarly, if we let \mathcal{S}_{tu} denote the elements of \mathcal{S} that include both t and u then

$$\text{cov}(n_{it}, n_{iu}) = \sum_{s \in \mathcal{S}_{tu}} \text{var}(X_{i,s}) = \sum_{s \in \mathcal{S}_{tu}} \lambda p^{|s|}(1-p)^{T-|s|}.$$

There are $\binom{T-2}{k-2}$ subsets $s \in \mathcal{S}_{tu}$ such that $|s| = k$ ($k = 2, \dots, T$). Hence, for $t \neq u$,

$$\text{cov}(n_{it}, n_{iu}) = \sum_{k=2}^T \binom{T-2}{k-2} \lambda p^k (1-p)^{T-k} = \lambda p^2,$$

and $\text{corr}(n_{it}, n_{iu}) = p$ ($t \neq u$).

This result is a special case of Johnson et al. (1997, equation 37.88), which is stated without proof.

Example: T=2, Poisson Case

Cormack (1989) mentions this case in closed-population capture-recapture modeling of data from one site only.

Suppressing site dependence, we have

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

where X_1, X_2, X_{12} are independent with $X_1, X_2 \sim \text{Pois}(\theta_1)$, where $\theta_1 = \lambda p(1-p)$ and $X_{12} \sim \text{Pois}(\theta_0)$, where $\theta_0 = \lambda p^2$. Note that small p would result typically in small values for X_{12} , and as p tends to zero n_1 and n_2 become independent, so that the model reverts to a thinned Poisson.

The counts (n_1, n_2) follow a bivariate Poisson distribution with $\text{corr}(n_1, n_2) = p$, and the bivariate Poisson probability is

$$\begin{aligned} \Pr(n_1, n_2; \lambda, p) &= \sum_{u=0}^{\min(n_1, n_2)} \left[\frac{e^{-\lambda p^2} (\lambda p^2)^u e^{-2\lambda p(1-p)}}{u!(n_1-u)!(n_2-u)!} \right. \\ &\quad \left. \times \{\lambda p(1-p)\}^{n_1+n_2-2u} \right] \\ &= \{p(1-p)\}^{n_1+n_2} \\ &\quad \sum_{u=0}^{\min(n_1, n_2)} \frac{e^{-\lambda(2p-p^2)} \lambda^{n_1+n_2-u}}{(1-p)^{2u} u!(n_1-u)!(n_2-u)!}. \end{aligned} \quad (2)$$

Including site dependence, the likelihood is

$$\begin{aligned} L(p, \lambda; \{n_{it}\}) &= e^{-(2\theta_1 + \theta_0)} \prod_{i=1}^R \\ &\quad \left\{ \frac{\theta_1^{n_{i1}+n_{i2}}}{n_{i1}! n_{i2}!} \sum_{u=0}^{\min(n_{i1}, n_{i2})} \binom{n_{i1}}{u} \binom{n_{i2}}{u} u! \left(\frac{\theta_0}{\theta_1^2} \right)^u \right\}. \end{aligned} \quad (3)$$

For $T = 2$ the expressions of (1) and (3) are identical, but the likelihood of (3) may be maximized without requiring selection of a value K .

3.1. Multivariate Poisson Distribution

For general T , let \mathcal{X}_i denote the set of all possible values $x_{i,s}$ of the random variables $X_{i,s}$, $s \in \mathcal{S}$ such that

$$n_{it} = \sum_{s \in \mathcal{S}_t} x_{i,s}, \quad t = 1, \dots, T.$$

Because the random variables $X_{i,s}$ are independent, the joint probability function of (n_{i1}, \dots, n_{iT}) is

$$\Pr(n_{i1}, \dots, n_{iT}) = \sum_{\mathcal{X}_i} \prod_{s \in \mathcal{S}} \Pr(X_{i,s} = x_{i,s}),$$

and

$$\begin{aligned} &\prod_{s \in \mathcal{S}} \Pr(X_{i,s} = x_{i,s}) \\ &= \prod_{s \in \mathcal{S}} \frac{\exp\{-\lambda p^{|s|} (1-p)^{T-|s|}\} \{\lambda p^{|s|} (1-p)^{T-|s|}\}^{x_{i,s}}}{x_{i,s}!}. \end{aligned}$$

There are $\binom{T}{k}$ elements $s \in \mathcal{S}$ such that $|s| = k$, for $k = 1, \dots, T$. Hence

$$\begin{aligned} \prod_{s \in \mathcal{S}} \exp\{-\lambda p^{|s|} (1-p)^{T-|s|}\} &= \exp\left\{-\lambda \sum_{k=1}^T p^k (1-p)^{T-k}\right\} \\ &= \exp\left[-\lambda \{1 - (1-p)^T\}\right]. \end{aligned}$$

Therefore, we can write

$$\begin{aligned} \Pr(n_{i1}, \dots, n_{iT}; \lambda, p) &= \sum_{\mathcal{X}_i} \prod_{s \in \mathcal{S}} \frac{\{p^{|s|} (1-p)^{T-|s|}\}^{x_{i,s}}}{x_{i,s}!} \\ &\quad \times \exp\left[-\lambda \{1 - (1-p)^T\}\right] \lambda^{\sum_{s \in \mathcal{S}} x_{i,s}}. \end{aligned} \quad (4)$$

The case $T = 2$ is given in (2). The associated R program incorporates efficient construction of \mathcal{X}_i .

3.2. Performance of the Multivariate Poisson Model

For illustration, we investigate performance of the multivariate Poisson model via simulation from the fitted model. We assess output for the cases $T = 2, 3, 4$ based upon 1000 simulations where $\lambda = 2, 5, 10$, $p = 0.1, 0.25$ and $R = 20$. The chosen parameter values were guided by those used in Royle (2004a). The model was fitted using the `optim` function in the R software package (R Core Team, 2014) using the default Nelder–Mead algorithm and a tolerance value of 1×10^{-12} . The results were checked with those from using several other `optim` algorithms, including simulated annealing and quasi-Newton. We observe that estimates for λ were very large in some cases (the maximum estimate from 1000 simulations was 1.36×10^{13} when $\lambda = 5$, $p = 0.25$, and $T = 2$). Figure 1 shows that non-positive values of a covariance diagnostic,

$$\text{cov}^*(n_1, n_2) = \bar{n}_1 \bar{n}_2 - \{(\bar{n}_1 + \bar{n}_2)/2\}^2, \quad (5)$$

can identify the high estimates of λ from fitting the bivariate Poisson. Here $\bar{n}_1 \bar{n}_2$ denotes the mean of the product $n_1 n_2$ over S sites. Note that this (intra-class) estimate is appropriate as $\mathbb{E}[n_1] = \mathbb{E}[n_2]$. A proof that a local maximum of the likelihood occurs at $p = 0$ when $\text{cov}^*(n_1, n_2) \leq 0$ is given in the Appendix; we are working on a general proof for $T > 2$, as well as a proof that there are no other maxima when the diagnostic is satisfied. Hence, in these instances when $\hat{p} = 0$, in order to have finite $\hat{\lambda} p$, $\hat{\lambda}$ is actually infinite and the large range of high estimates of abundance obtained in practice, as in Figure 1, is partly an artefact of the optimization routine stopping prematurely when the likelihood is flat.

For more than two visits ($T > 2$), the appropriate covariance diagnostic can be estimated as

$$\begin{aligned} \text{cov}^*(n_1, \dots, n_T) &= \frac{2}{T(T-1)} (\bar{n}_1 \bar{n}_2 + \dots + \bar{n}_{T-1} \bar{n}_T) \\ &\quad - \left(\frac{\bar{n}_1 + \dots + \bar{n}_T}{T} \right)^2, \end{aligned} \quad (6)$$

where the first term consists of the average of the means of all $T(T-1)/2$ pairwise products. Our conjecture that the diagnostic extends for $T > 2$ is supported by Web Figure 4 which compares the covariance diagnostic (6) with $\hat{\lambda}$ from the multivariate Poisson model for $T = 3$, when $\lambda = 2, 5, 10$.

Performance of the covariance diagnostic is demonstrated further in Table 1, which shows close correspondence between the proportion of simulations where the diagnostic is negative and the proportion where $\hat{\lambda}$ is large ($\hat{\lambda} > 500$). Table 1 also shows the prevalence of infinite estimates of $\hat{\lambda}$, particularly as λ , T , and p decrease. In fact for the case where $\lambda = 2$, $p = 0.1$,

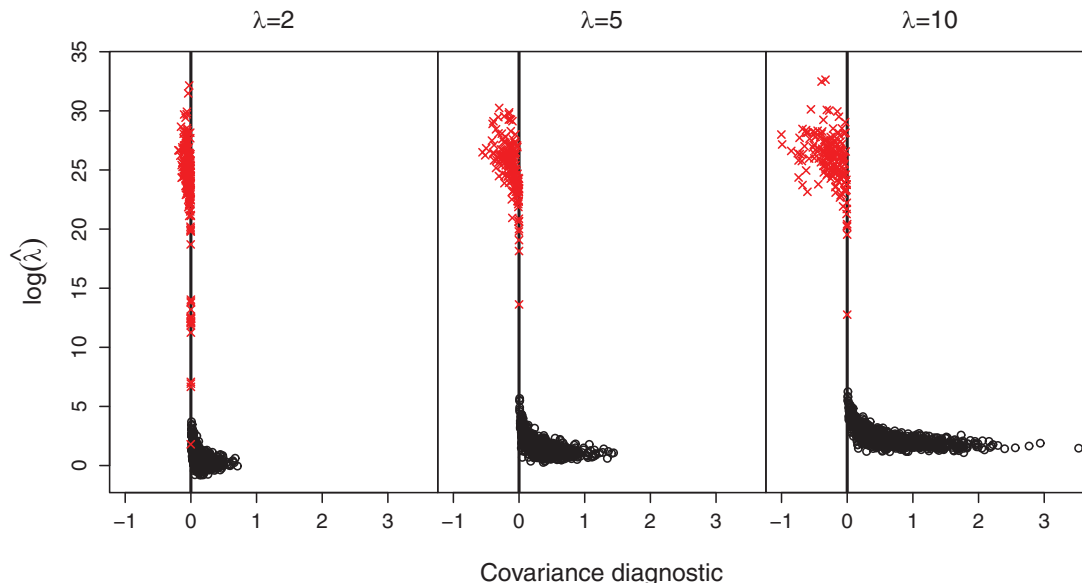


Figure 1. $\log(\hat{\lambda})$ from the bivariate Poisson model plotted against the covariance diagnostic, $\text{cov}^*(n_1, n_2) \leq 0$ from (5), based upon 1000 simulated datasets for $R = 20$, $\lambda = 2, 5, 10$ and $p = 0.25$. Values at which the covariance diagnostic is negative are shown by crosses. This figure appears in color in the electronic version of this article.

and $T = 2$, a finite value of $\hat{\lambda}$ was not achievable in over half of 1000 simulations.

4. Explicit Form for the Bivariate Negative-Binomial Case

The Poisson distribution may be replaced by a mixed-Poisson distribution, for which $\lambda \sim g(\lambda; \theta)$, when the probability of (2) becomes

$$\begin{aligned} &\Pr(n_1, n_2; p, \theta) \\ &= \{p(1-p)\}^{n_1+n_2} \sum_{u=0}^{\min(n_1, n_2)} \frac{1}{(1-p)^{2u} u!(n_1-u)!(n_2-u)!} \\ &\quad \times \int_0^\infty e^{-\lambda(2p-p^2)} \lambda^{n_1+n_2-u} g(\lambda; \theta) d\lambda. \end{aligned}$$

Table 1

Performance of the covariance diagnostic for the multivariate Poisson model, based upon 1000 simulations for various scenarios of λ , p , and T for $R = 20$ sites. EPN is the proportion of simulations when the sample covariance diagnostic was negative. EPD is the proportion of simulations where the estimate of $\hat{\lambda} > 500$.

λ	p	$T = 2$		$T = 3$		$T = 4$	
		EPN	EPD	EPN	EPD	EPN	EPD
2	0.10	0.505	0.505	0.351	0.351	0.276	0.276
2	0.25	0.225	0.224	0.090	0.089	0.033	0.033
5	0.10	0.427	0.427	0.362	0.361	0.219	0.222
5	0.25	0.167	0.167	0.084	0.084	0.017	0.020
10	0.10	0.398	0.398	0.317	0.318	0.251	0.256
10	0.25	0.180	0.181	0.066	0.066	0.038	0.038

For the negative-binomial distribution, the mixing distribution is gamma with parameters $\theta = (\alpha, \beta)$ and

$$g(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \text{ for } \lambda \geq 0, \tag{7}$$

which results in the NB-2 form (Hilbe, 2011, p. 187). In this case

$$\begin{aligned} &\int_0^\infty e^{-\lambda(2p-p^2)} \lambda^{n_1+n_2-u} g(\lambda; \alpha, \beta) d\lambda \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left\{ e^{-\lambda(2p-p^2+\beta)} \times \lambda^{n_1+n_2-u+\alpha-1} \right\} d\lambda \\ &= \frac{\beta^\alpha \Gamma(n_1+n_2-u+\alpha)}{\Gamma(\alpha)(2p-p^2+\beta)^{n_1+n_2-u+\alpha}}. \end{aligned}$$

Therefore the joint probability for the bivariate negative-binomial model is given by

$$\begin{aligned} \Pr(n_1, n_2; p, \alpha, \beta) &= \frac{\beta^\alpha \{p(1-p)\}^{n_1+n_2}}{\Gamma(\alpha)} \\ &\sum_{u=0}^{\min(n_1, n_2)} \frac{\Gamma(n_1+n_2-u+\alpha)}{u!(n_1-u)!(n_2-u)!(1-p)^{2u}(2p-p^2+\beta)^{n_1+n_2-u+\alpha}}. \end{aligned} \tag{8}$$

In the parameterization of (7), the mean and variance of the gamma distribution are α/β and α/β^2 , respectively. If we now write $\lambda = \alpha/\beta$ for the expected value of the Poisson mean, then the variance is λ^2/α and the coefficient of variation of

the Poisson mean is $1/\sqrt{\alpha}$. The Poisson model arises as the limit $\alpha, \beta \rightarrow \infty$, maintaining $\lambda = \alpha/\beta$.

In terms of the parameters α and λ , $\beta = \alpha/\lambda$ and we can write (8) as

$$\Pr(n_1, n_2; \lambda, p, \alpha) = \frac{\alpha^\alpha \{p(1-p)\}^{n_1+n_2}}{\lambda^\alpha \Gamma(\alpha)}$$

$$\sum_{u=0}^{\min(n_1, n_2)} \frac{\Gamma(n_1 + n_2 - u + \alpha)}{u!(n_1 - u)!(n_2 - u)!(1-p)^{2u}} \left\{ \frac{\lambda}{\lambda p(2-p) + \alpha} \right\}^{n_1+n_2-u+\alpha}.$$
(9)

The case for $T > 2$ follows in the same way, by integrating the expression of (4), to give the multivariate negative-binomial probability as

$$\Pr(n_{i1}, \dots, n_{iT}; \lambda, p, \alpha)$$

$$= \frac{\alpha^\alpha}{\lambda^\alpha \Gamma(\alpha)} \sum_{x_i} \prod_{s \in \mathcal{S}} \frac{\{p^{|s|}(1-p)^{T-|s|}\}^{x_{i,s}} \Gamma(\sum_{s \in \mathcal{S}} x_{i,s} + \alpha)}{x_{i,s}! \left\{1 - (1-p)^T + \frac{\alpha}{\lambda}\right\}^{\sum_{s \in \mathcal{S}} x_{i,s} + \alpha}}.$$

The expression $n_{it} = \sum_{s \in \mathcal{S}_t} X_{i,s}$ also applies to the negative binomial case, but the $\{X_{i,s}\}$ are no longer independent.

5. The Effect of the Choice of K on Fitting the N-Mixture Model: Poisson Case

5.1. Incorrect Estimates due to the Choice of K

We now consider how the choice of K for computing the Poisson N-mixture likelihood of (1) interacts with the occurrence of infinite estimates of λ . Output is obtained for 1000 simulations based on the parameter values used in Royle (2004a), where $\lambda = 5$, $p = 0.25$ and $R = 20, 50$, but for number of sampling occasions $T = 2, 3, 4, 5$. The models were again fitted using `optim` in the R software package. The parameters p and λ were constrained to be in range via logit and log link functions, respectively. Each simulated dataset was fitted with $K = 100, 500, 1000$.

We see that large finite estimates of abundance can arise, in particular where the number of sampling occasions T is small (Figure 2). Specifically, a proportion of simulations result in a second peak in the sampling distribution for $\hat{\lambda}$ and the value at which this is found increases with the value of K . Fitting the multivariate Poisson model to simulated data created under comparable scenarios for $T = 2, 3, 4$ also produced a second peak in the sampling distribution for λ , but as described in Section 3.2, the estimates were substantially greater in the absence of the limiting value K in the N-mixture model. An increase in the number of sampling occasions reduces the incidence of high estimates of λ , which become rare for $T > 3$, as more information is available as T increases. For $T = 5$ very few high estimates of λ occurred in the 1000 simulations. An

increase in the number of sites also reduces the proportion of high values (Web Figure 5).

Thus when the N-mixture model is fitted by maximizing the likelihood of (1), when $\hat{\lambda}$ should be infinite, λ is estimated as large as possible for a given value of K , and \hat{p} is restricted to be as close to zero as possible. We discuss this matter further in Web Appendix 1. The occurrence of large finite estimates of λ is similar to analogous findings of Wang and Lindsay (2005) in the context of species richness estimation.

5.2. Automatic Choice of K

For the Poisson case the covariance diagnostic identifies when infinite values of $\hat{\lambda}$ arise. When the diagnostic is not satisfied, K may be selected automatically, for example by ensuring that the Poisson upper tail probability is $< 10^{-10}$, so that the value of K will adapt for successive iterations according to the estimate of λ . This approach was also suggested by Guillera-Aroita et al. (2012). We have found this to be a simple and preferable alternative to fitting the model for successively larger values of K until estimates appear to stabilize.

6. Moment Estimation for a Mixed-Poisson N-Mixture Model

Suppose we have an N-mixture model in which N_i follows a mixed-Poisson distribution, as in Section 4, with

$$E(N_i) = \lambda \quad \text{and} \quad \text{var}(N_i) = \sigma^2, \quad \text{with } \sigma^2 \geq \lambda.$$

Conditional on N_i , the random variables n_{i1}, \dots, n_{iT} are independent binomial variables, with

$$n_{ij}|N_i \sim \text{Bin}(N_i, p).$$

Therefore, conditional on N_i

$$E(n_{ij}|N_i) = N_i p$$

$$E(n_{ij}^2|N_i) = N_i p(1-p) + N_i^2 p^2$$

$$E(n_{ij}, n_{ik}|N_i) = N_i^2 p^2 \quad (j \neq k),$$

and the corresponding unconditional expectations are

$$E(n_{ij}) = \lambda p \tag{10}$$

$$E(n_{ij}^2) = \lambda p(1-p) + (\lambda^2 + \sigma^2) p^2 \tag{11}$$

$$E(n_{ij}, n_{ik}) = (\lambda^2 + \sigma^2) p^2 \quad (j \neq k). \tag{12}$$

It follows that

$$\text{cov}(n_{ij}, n_{ik}) = \sigma^2 p^2 \quad \text{and} \quad \text{corr}(n_{ij}, n_{ik}) = \sigma^2 p / \lambda.$$

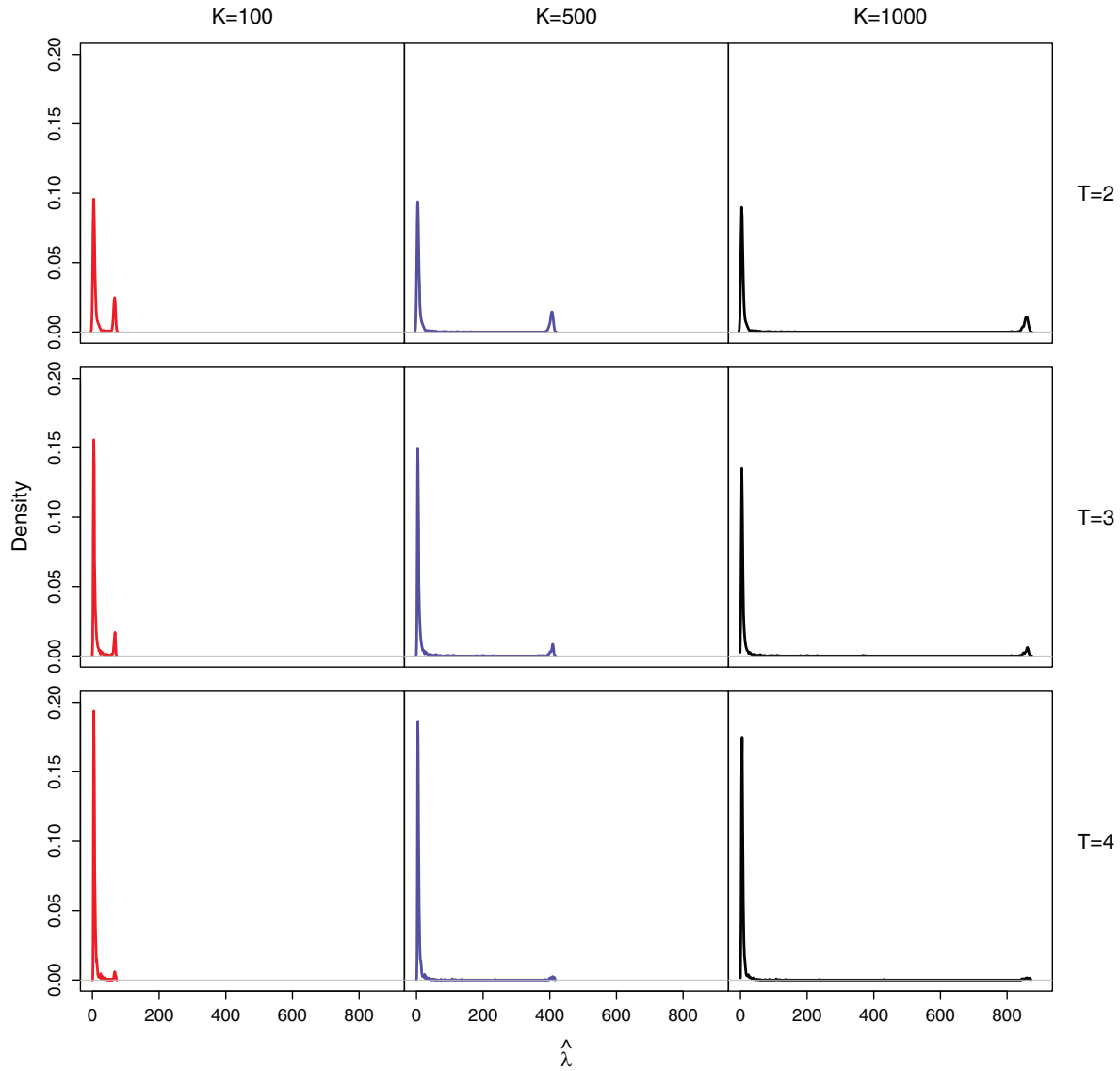


Figure 2. Kernel density estimates of $\hat{\lambda}$ from the Poisson N-mixture model for $R = 20$ sites, $\lambda = 5$ and $p = 0.25$ based upon 1000 simulated datasets for $T = 2, 3, 4$, and $K = 100, 500, 1000$. This figure appears in color in the electronic version of this article.

6.1. *Moment Estimation*

We have the following moment estimates for $E(n_{ij})$, $E(n_{ij}^2)$, and $E(n_{ij}n_{ik})$, respectively:

$$m_1 = \frac{1}{RT} \sum_{t=1}^T \sum_{i=1}^R n_{it}$$

$$m_2 = \frac{1}{RT} \sum_{t=1}^T \sum_{i=1}^R n_{it}^2$$

$$m_{12} = \frac{2}{RT(T-1)} \sum_{t=1}^T \sum_{s=t+1}^T \sum_{i=1}^R n_{it}n_{is}$$

Equating these to the expectations given by (10)–(12) yields the following moment estimators of the parameters λ , p , and σ^2

$$\tilde{p} = (m_1 - m_2 + m_{12})/m_1$$

$$\tilde{\lambda} = m_1/\tilde{p}$$

$$\tilde{\sigma}^2 = (m_{12} - m_1^2)/\tilde{p}^2.$$

Because $\sigma^2 \geq 0$, we require

$$m_{12} - m_1^2 \geq 0, \tag{13}$$

for a valid set of moment estimates. This is the same diagnostic as used previously in (6).

We also require $0 < p \leq 1$. The lower bound yields the new diagnostic

$$m_1 - m_2 + m_{12} > 0, \tag{14}$$

for a finite (moment) estimate of λ . The upper bound yields

$$m_1 - m_2 + m_{12} \leq m_1$$

or

$$m_{12} \leq m_2,$$

which is a consequence of the Cauchy–Schwarz inequality and not a useful diagnostic. The bound $m_1 - m_2 + m_{12} > 0$ given above to ensure $\hat{p} > 0$ and hence $\hat{\lambda}$ finite, gives a new diagnostic.

If we adopt a method-of-moments (MOM) approach for the bivariate Poisson distribution, p is estimated by the sample correlation of the counts, as observed also by Royle (2004b), and λ is estimated by dividing $(n_1 + n_2)/2$ by this estimate of p . For more than two visits ($T > 2$), p can be estimated by the mean of all sample correlations between counts for different sampling occasions. Then $\hat{\lambda}$ is the sample mean of all counts divided by this estimate of p . This generalizes Holgate’s (1964) work, which considered $T = 2$ only. In Web Appendix 2 we assess the performance of MOM estimation as a simple method for parameter estimation compared to maximum likelihood for the N-mixture model.

6.2. Performance of the Multivariate Negative-Binomial Model

Given the proposed diagnostics for the mixed-Poisson case in Section 6.1, here we assess the performance of the multivariate negative-binomial model. Simulated data were fitted as in Section 3.2 but for the negative binomial case, with $\lambda = 2, 5, 10$ and $\alpha = 1.25, 5$. We again assume that $\hat{\lambda} > 500$ equates to infinite $\hat{\lambda}$. If both (13) and (14) are negative, $\hat{\lambda}$ is very likely to be infinite and the mean proportion with $\hat{\lambda} > 500$ from 21 scenarios is 0.921 (Table 2). However performance of the diagnostics when one or more of the two diagnostics is negative is less clear. Additionally, $\hat{\lambda}$ may occasionally be infinite despite both diagnostics being positive and on average $\hat{\lambda} > 500$ for approximately 8.5% of simulations when both diagnostics are positive. Performance for the bivariate cases where $p = 0.25$ and $\alpha = 5$ is illustrated in Figure 3 and for the cases where $T = 3$ and $\alpha = 1.25$ in Web Figures 6–8. We see that neither singly nor in combination do the diagnostics perform as well as the single diagnostic for the Poisson case. We see fewer instances of infinite $\hat{\lambda}$ for large T and p .

7. Application to Hermann’s Tortoise Data

Here we analyze data from a study of the threatened Hermann’s tortoise *Testudo hermanni* in southeastern France. One hundred and eighteen sites were each surveyed three times during a period when the species is most active. Full details are provided in Couturier et al. (2013), and we briefly

Table 2

Performance of the covariance diagnostic for the multivariate negative-binomial model, based upon 1000 simulations for various scenarios of λ , p , α , and T for $R = 20$ sites. EP_1, EP_3 , and EP_5 are the proportion of simulations where both diagnostics are negative, one or more diagnostic is negative, or both diagnostics are positive, respectively. EP_2, EP_4 , and EP_6 are the corresponding proportions of those where $\hat{\lambda} > 500$.

λ	p	α	T	EP_1	EP_2	EP_3	EP_4	EP_5	EP_6
2	0.10	1.25	2	0.192	0.938	0.3	0.853	0.388	0.072
2	0.10	1.25	3	0.093	0.925	0.271	0.841	0.426	0.131
2	0.10	5.00	2	0.199	0.92	0.296	0.804	0.274	0.113
2	0.10	5.00	3	0.104	0.904	0.264	0.822	0.293	0.126
2	0.25	1.25	2	0.046	0.913	0.229	0.777	0.571	0.07
2	0.25	1.25	3	0.002	1	0.138	0.681	0.71	0.048
2	0.25	5.00	2	0.064	0.953	0.184	0.826	0.411	0.097
2	0.25	5.00	3	0.011	1	0.103	0.748	0.473	0.047
5	0.10	1.25	2	0.088	0.966	0.347	0.813	0.472	0.121
5	0.10	1.25	3	0.023	1	0.333	0.757	0.52	0.113
5	0.10	5.00	2	0.139	0.935	0.305	0.803	0.282	0.128
5	0.10	5.00	3	0.064	0.906	0.252	0.829	0.343	0.143
5	0.25	1.25	2	0.006	1	0.217	0.71	0.746	0.068
5	0.25	1.25	3	0	–	0.137	0.533	0.843	0.047
5	0.25	5.00	2	0.038	0.763	0.193	0.741	0.555	0.05
5	0.25	5.00	3	0.002	0.5	0.108	0.694	0.678	0.028
10	0.10	1.25	2	0.032	0.969	0.342	0.813	0.596	0.139
10	0.10	1.25	3	0.005	1	0.325	0.775	0.65	0.097
10	0.10	5.00	2	0.116	0.931	0.322	0.835	0.378	0.108
10	0.10	5.00	3	0.027	0.926	0.302	0.844	0.437	0.105
10	0.25	1.25	2	0	–	0.193	0.674	0.806	0.069
10	0.25	1.25	3	0	–	0.125	0.472	0.87	0.029
10	0.25	5.00	2	0.01	0.9	0.156	0.756	0.726	0.054
10	0.25	5.00	3	0.001	1	0.09	0.656	0.817	0.026

reassess the conclusions drawn in their article and demonstrate the effect of study design on results.

For the tortoise data, optimization of the negative-binomial model confirms that $\hat{\lambda}$ is infinite in the negative binomial model for these data; after 500 iterations, the estimates had reached

$$\hat{\lambda} = 39616973, \quad \hat{p} = 3.322971 \times 10^{-8}, \quad \hat{\alpha} = 1.506465.$$

As noted in Couturier et al. (2013), the fit is much improved compared to the Poisson case, with -maximum log-likelihood 540.34 versus 576.27, but at the expense of $\hat{\lambda}$ becoming infinite. Hence for this dataset a finite estimate of mean abundance can be obtained for the Poisson but not for the negative-binomial. Whilst the first diagnostic (13) is positive, $m_{12} - m_1^2 = 1.05$, so that the Poisson estimate is finite, the additional diagnostic (14) is negative, $m_1 - m_2 + m_{12} = -0.2655$.

The zero-inflated Poisson is an intermediate model between the Poisson and negative-binomial, with -maximum log-likelihood 562.13 for these data. The zero-inflated Poisson therefore provides an improvement upon the Poisson case, but still yields the finite parameter estimate $\hat{\lambda} = 7.58$.

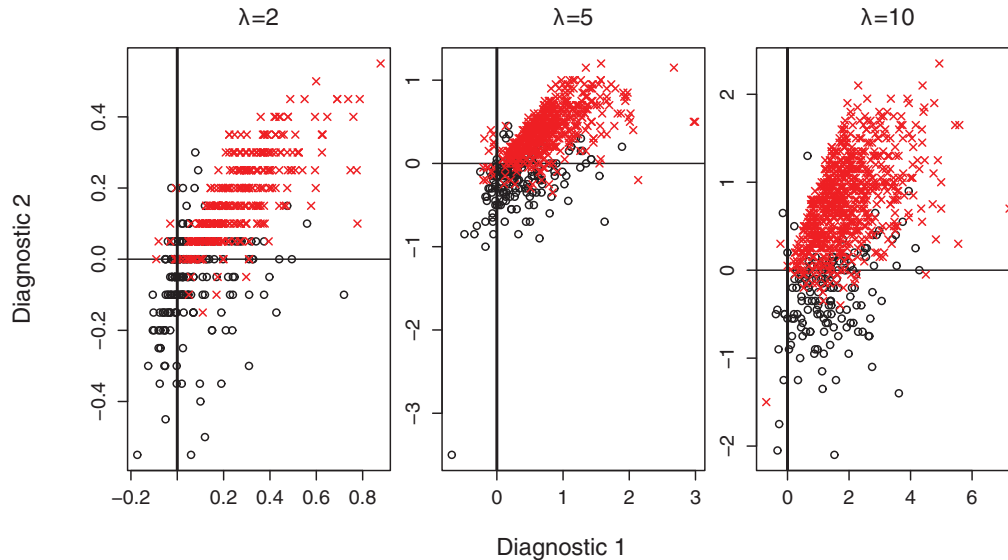


Figure 3. Diagnostic 1 (13) versus diagnostic 2 (14) from the bivariate negative binomial model, based upon 1000 simulated datasets for $R = 20$, $\lambda = 2, 5, 10$, $\alpha = 5$, and $p = 0.25$. Values at which $\lambda > 500$ and $\lambda \leq 500$ are shown by circles and crosses, respectively. This figure appears in color in the electronic version of this article.

To show the potential effect of study design on model performance, we inspect the sample covariance diagnostic (13) for this dataset for the Poisson case 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). The diagnostic based upon all three visits but a random sample of fewer sites, was negative for 1.7% and 0% of 1000 samples, respectively for $R = 20$ and $R = 50$. However for only two visits, the diagnostic was negative for 9.0% and 0.8% of 1000 samples, respectively for $R = 20$ and $R = 50$.

8. Discussion and Recommendations

We have shown that the N-mixture model can produce infinite estimates of abundance, particularly when working with a limited number of sampling occasions and low detection probability. The equivalence of the N-mixture model with the multivariate Poisson has been demonstrated, allowing us to understand and diagnose poor behavior of the N-mixture model.

We believe the equivalence of the Poisson N-mixture model to the multivariate Poisson distribution to be previously largely unknown, especially in statistical ecology. The multivariate Poisson model conveniently avoids the requirement to select an upper bound K . We provide code for fitting the multivariate Poisson and negative-binomial models. Possible alternative techniques for fitting the multivariate distributions include using the EM algorithm (Karlis, 2003), a composite likelihood (Jost, Brcich, and Zoubir, 2006) or a symbolic computation approach (Sontag and Zeilberger, 2010). Consequently this equivalence could also have the alternative purpose of using the N-mixture model to provide simple fitting of the multivariate Poisson and negative-binomial models for particular covariance structures.

A recent extension of the N-mixture model to open populations by including population dynamics parameters offers great potential but also requires an upper bound to be set (Dail and Madsen, 2011). Further exploration of this model via simulation to assess performance is in progress. Kéry et al. (2009) extended the N-mixture model to allow for analysis of data resulting from closed sampling periods connected by open periods and the multivariate formulations also apply in that case. Dorazio, Martin, and Edwards (2013) provide an extension in which p is given a distribution at each visit. The binomial distribution in (1) is then replaced by a beta-binomial. This has also been considered in a Bayesian context by Martin et al. (2011). For the multivariate Poisson case this extension is dealt with by appropriate numerical integration of the probability of (4). An increasing number of studies use a Bayesian approach for parameter estimation (Kéry et al., 2009; Graves et al., 2011). Further simulation study comparing a Bayesian approach with maximum-likelihood estimation could show whether this approach can also produce poor estimates in some scenarios. Some comparisons have been made by Toribio, Gray, and Liang (2012), based upon parameter values from Royle (2004a).

In practice, covariates are frequently used to describe variation in abundance and detection. Further analysis could determine how the inclusion of covariates might affect instances where a finite abundance estimate cannot be obtained for a model with constant abundance and detection, and in particular determine whether parameters may become identifiable.

Good experimental design is vital for occupancy studies; see for example Guillera-Aroita, Ridout, and Morgan, (2010, 2014). The same issues apply for N-mixture work, though with the different perspective of avoiding poor model-fitting behavior. If possible, study design effort should be distributed to ensure more than two visits are made to

each site (in addition to including a reasonable number of sites). Alternatively a study design where more visits are made to a subset of the sites is worth exploring.

For maximum-likelihood estimation, we recommend using MOM estimates to start the iterative search for MLEs. In the Poisson case the covariance diagnostic may be used to determine when infinite estimates of abundance may arise. Infinite estimates of abundance may occur for some model choices but not others, as for the Hermann's Tortoise case study. Hence we advise fitting the model for multiple distribution choices, to identify which may provide finite estimates of abundance. An R program is available in the Supplementary Materials which allows for covariates in the detection and abundance parameters.

9. Supplementary Materials

The Web Appendices referenced in Sections 3.2, 5.1, and 6.2, together with R code, are available with this paper at the *Biometrics* website on Wiley Online Library.

ACKNOWLEDGEMENTS

This work was part-funded by EPSRC grants EP/I000917/1 and EP/P505577/1. We thank Bill Link, Peter Jupp and Marc Kéry for their useful discussion, and Thibaut Couturier for supplying the Hermann's tortoise data. Useful input was provided by two referees and an Associate Editor.

REFERENCES

- Cormack, R. M. (1989). Log-linear models for capture–recapture. *Biometrics* **45**, 395–413.
- 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.
- Dail, D. and Madsen, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics* **67**, 577–587.
- Dodd, C. K. and Dorazio, R. M. (2004). Using counts to simultaneously estimate abundance and detection probabilities in a salamander community. *Herpetologica* **60**, 468–478.
- Dorazio, R. M., Martin, J., and Edwards, H. H. (2013). Estimating abundance while accounting for rarity, correlated behavior, and other sources of variation in counts. *Ecology* **94**, 1472–1478.
- Fiske, I. and Chandler, R. B. (2011). *Unmarked*: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software* **43**, 1–23.
- Graves, T. A., Kendall, K. C., Royle, J. A., Stetz, J. B., and Macleod, A. C. (2011). Linking landscape characteristics to local grizzly bear abundance using multiple detection methods in a hierarchical model. *Animal Conservation* **14**, 652–664.
- Guillera-Arroita, G., Ridout, M. S., and Morgan, B. J. T. (2010). Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution* **1**, 131–139.
- Guillera-Arroita, G., Ridout, M. S., and Morgan, B. J. T. (2014). Two-stage sequential Bayesian study design for species occupancy estimation. *Journal of Agricultural, Biological, and Environmental Statistics* **19**, 278–291.
- Guillera-Arroita, G., Ridout, M. S., Morgan, B. J. T., and Linkie, M. (2012). Models for species-detection data collected along transects in the presence of abundance-induced heterogeneity and clustering in the detection process. *Methods in Ecology and Evolution* **3**, 358–367.
- Hilbe, J. M. (2011). *Negative Binomial Regression*. New York: Cambridge University Press.
- Hines, J. E. (2011). *Program PRESENCE 4.1—Software to estimate patch occupancy and related parameters*. U.S. Geological Survey Patuxent Wildlife Research Center, Maryland.
- Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika* **51**, 241–245.
- Hunt, J. W., Weckerly, F. W., and Ott, J. R. (2012). Reliability of occupancy and binomial mixture models for estimating abundance of golden-cheeked warblers (Setophaga chrysoparia). *The Auk* **129**, 105–114.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. New York: Wiley.
- Jost, T. A., Brcich, R. F., and Zoubir, A. M. (2006). Estimating the parameters of the multivariate Poisson distribution using the composite likelihood concept. In *The Proceedings of the 31st IEEE International Conference on Acoustics, Speech and Signal Processing*. Toulouse, France: IEEE.
- Karlis, D. (2003). An EM algorithm for multivariate Poisson distribution and related models. *Journal of Applied Statistics* **30**, 63–77.
- Kéry, M. (2008). Estimating abundance from bird counts: Binomial mixture models uncover complex covariate relationships. *The Auk* **125**, 336–345.
- Kéry, M., Dorazio, R. M., Soldaat, L., Van Strien, A., Zuiderwijk, A., and Royle, J. A. (2009). Trend estimation in populations with imperfect detection. *Journal of Applied Ecology* **46**, 1163–1172.
- Kéry, M., Royle, J. A., and Schmid, H. (2005). Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications* **15**, 1450–1461.
- Martin, J., Royle, J. A., Mackenzie, D. I., Edwards, H. H., Kéry, M., and Gardner, B. (2011). Accounting for non-independent detection when estimating abundance of organisms with a Bayesian approach. *Methods in Ecology and Evolution* **2**, 595–601.
- McIntyre, A. P., Jones, J. E., Lund, E. M., Waterstrat, F. T., Giovanini, J. N., Duke, S. D., Hayes, M. P., Quinn, T., and Kroll, A. J. (2012). Empirical and simulation evaluations of an abundance estimator using unmarked individuals of cryptic forest-dwelling taxa. *Forest Ecology and Management* **286**, 129–136.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Royle, J. A. (2004a). N-mixture models for estimating population size from spatially replicated counts. *Biometrics* **60**, 108–115.
- Royle, J. A. (2004b). Generalized estimators of avian abundance from count survey data. *Animal Biodiversity and Conservation* **27**, 375–386.
- Royle, J. A., Nichols, J. D., and Kéry, M. (2005). Modelling occurrence and abundance of species when detection is imperfect. *Oikos* **110**, 353–359.
- Sontag, E. D. and Zeilberger, D. (2010). A symbolic computation approach to a problem involving multivariate Poisson distributions. *Advances in Applied Mathematics* **44**, 359–377.
- Toribio, S., Gray, B., and Liang, S. (2012). An evaluation of the Bayesian approach to fitting the N-mixture model for use with pseudo-replicated count data. *Journal of Statistical Computation and Simulation* **82**, 1135–1143.

Wang, J.-P. Z. and Lindsay, B. G. (2005). A penalized nonparametric maximum likelihood approach to species richness estimation. *Journal of the American Statistical Association* **100**, 942–959.

Zellweger-Fischer, J., Kéry, M., and Pasinelli, G. (2011). Population trends of brown hares in Switzerland: The role of land-use and ecological compensation areas. *Biological Conservation* **144**, 1364–1373.

Received August 2013. Revised August 2014.
Accepted August 2014.

APPENDIX

Proof that when $T = 2$ a local maximum of the likelihood occurs at $p = 0$ when $\text{cov}^(n_1, n_2) \leq 0$.*

It is convenient here to set $\theta_0 = \lambda p^2, \theta_2 = \lambda p$. It is simple to prove that $\hat{\theta}_2 = \frac{\sum_i (n_{1i} + n_{2i})}{2R}$, as noted in Holgate (1964). For observation (n_{1i}, n_{2i}) , we write

$$g_i(\theta_0) = \sum_{m=0}^{u_i} \frac{\theta_0^m (\hat{\theta}_2 - \theta_0)^{n_{1i} + n_{2i} - 2m}}{m!(n_{1i} - m)!(n_{2i} - m)!}, \quad \text{where } u_i = \min(n_{1i}, n_{2i}).$$

The profile log-likelihood function for θ_0 is then given by

$$\ell(\theta_0) = R(\theta_0 - 2\hat{\theta}_2) + \sum_{i=1}^R \log(g_i(\theta_0)) \quad \text{and}$$

$$\frac{d\ell}{d\theta_0} = R + \sum_{i=1}^R \frac{g'_i(\theta_0)}{g_i(\theta_0)}.$$

We deduce that

$$\frac{g'_i(0)}{g_i(0)} = \frac{1}{\hat{\theta}_2^2} \{n_{1i}n_{2i} - \hat{\theta}_2(n_{1i} + n_{2i})\}.$$

Thus

$$\left. \frac{d\ell}{d\theta_0} \right|_{\theta_0=0} = R + \frac{1}{\hat{\theta}_2^2} \left\{ \sum_{i=1}^R n_{1i}n_{2i} - \hat{\theta}_2 \sum_{i=1}^R (n_{1i} + n_{2i}) \right\},$$

and

$$\begin{aligned} \left. \frac{d\ell}{d\theta_0} \right|_{\theta_0=0} \leq 0 &\equiv \frac{1}{R} \sum_i n_{1i}n_{2i} \leq \hat{\theta}_2^2 \\ &\equiv \text{cov}^*(n_1, n_2) \leq 0. \end{aligned}$$