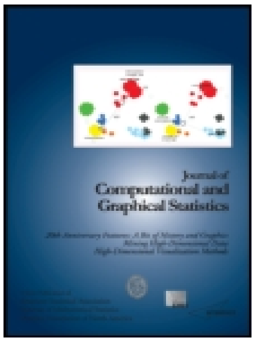

Journal of Computational and Graphical Statistics

ISSN: 1061-8600 (Print) 1537-2715 (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

Analysis of Multinomial Models With Unknown Index Using Data Augmentation

J. Andrew Royle, Robert M Dorazio & William A Link

Analysis of Multinomial Models With Unknown Index Using Data Augmentation

J. Andrew Royle, Robert M. Dorazio, and William A. Link

Multinomial models with unknown index ("sample size") arise in many practical settings. In practice, Bayesian analysis of such models has proved difficult because the dimension of the parameter space is not fixed, being in some cases a function of the unknown index. We describe a data augmentation approach to the analysis of this class of models that provides for a generic and efficient Bayesian implementation. Under this approach, the data are augmented with all-zero detection histories. The resulting augmented dataset is modeled as a zero-inflated version of the complete-data model where an estimable zero-inflation parameter takes the place of the unknown multinomial index. Interestingly, data augmentation can be justified as being equivalent to imposing a discrete uniform prior on the multinomial index. We provide three examples involving estimating the size of an animal population, estimating the number of diabetes cases in a population using the Rasch model, and the motivating example of estimating the number of species in an animal community with latent probabilities of species occurrence and detection.

Key Words: Capture–recapture; Multiple lists; Rasch model; Population size estimation; Site-occupancy; Species richness; Zero-inflation.

1. INTRODUCTION

An important problem in ecology is estimating the size of a multinomial population. The most common applications include estimating the size of a demographically closed animal population using capture–recapture sampling (Otis, Burnham, White, and Anderson 1978) and estimating the number of species in an animal community (Dorazio and Royle 2003). Analogous problems exist in many other settings in which sampling is based on multiple lists. These include surveys of people associated with a particular medical or social condition (Yip et al. 1995) and pre- or post-enumeration surveys designed to estimate census undercount (Wolter 1986; Darroch, Fienberg, Glonek, and Junker 1993). Estimating the number of errors in computer code (Basu and Ebrahimi 2001) is yet another example.

J. Andrew Royle is Research Statistician, U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel MD (E-mail: *aroyle@usgs.gov*). Robert M. Dorazio is Research Statistician and Courtesy Associate Professor, U.S. Geological Survey, Florida Integrated Science Center, Department of Statistics, University of Florida, Gainesville, FL (E-mail: *bdorazio@usgs.gov*). William A. Link is Research Statistician, U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, MD (E-mail: *wlink@usgs.gov*).

Journal of Computational and Graphical Statistics, Volume 16, Number 1, Pages 67–85
DOI: 10.1198/106186007X181425

Consider a population of size N subjected to T samples yielding a sample of n unique individuals. Each of the n individuals may be assigned a binary sequence, say $\mathbf{x} = (x_1, \ldots, x_T)$, whose elements indicate whether the individual is detected ($x_j = 1$) or not ($x_j = 0$) in the jth sample. Such sequences are often referred to as "capture histories." The objective of the estimation problem is to estimate the number of individuals in the population that are not detected in the survey and thus have an all-zero capture history ($\mathbf{x} = \mathbf{0}$).

There are two classical approaches for estimating N in such problems (Sanathanan 1972). The first is based on the so-called "conditional likelihood," in which the multinomial likelihood of the capture histories is formulated by conditioning on the observed sample size n. In this approach, N does not appear in the likelihood; it is computed as a function of the estimated model parameters and the data. However, in some cases it is advantageous to retain N in the likelihood; moreover, removing the parameter of interest from the likelihood in order to estimate nuisance parameters is philosophically unappealing. The second classical approach is to view the number of unobserved individuals in the population, say n_0, as a model parameter and to specify the complete-data likelihood directly in terms of the multinomial index $N = n + n_0$. In simple models either approach is relatively straightforward to implement, and both estimators are asymptotically equivalent (Sanathanan 1972). However, in more complex models that impose a latent structure on the nuisance parameters (Coull and Agresti 1999; Dorazio and Royle 2003, 2005a), considerably more computational effort can be required to integrate the nuisance parameters from the likelihood.

As an example, consider the class of models wherein detection probability is assumed to vary among individuals in the population but not among sampling attempts. This class of models is designated "Model M_h" in the capture–recapture literature (Otis et al. 1978). Let y_i ($i = 1, \ldots, n$) denote the number of times that individual i is observed in T sampling attempts. We assume that y_i is a binomial outcome with index T and success probability p_i and that p_i is a realization of a random variable with probability density $g(p|\theta)$. Classical solutions for estimating N in this model specify the likelihood function in terms of marginal probabilities, $\pi(y_i|\theta) = \int_0^1 c(y_i)\, p_i^{y_i}(1 - p_i)^{T-y_i} g(p_i|\theta)\,\mathrm{d}p_i$, wherein the nuisance parameter p_i is removed by integration and $c(y_i) = T!/y_i!(T - y_i)!$.

This situation, in which the model is parameterized in terms of a high-dimensional nuisance parameter for which the data are only partially observed, would seem to be ideally suited to a Bayesian analysis. Moreover, various conditioning arguments (e.g., when the data are conditioned on N) yield simplifications that should be amenable to Gibbs sampling and contemporary Markov chain Monte Carlo (MCMC) methods. Although Bayesian analyses of this model have been considered previously (Fienberg, Johnson, and Junker 1999; Basu and Ebrahimi 2001; Tardella 2002), the development of Gibbs samplers for models that condition on N has proved to be challenging because the dimension of the nuisance parameter $\mathbf{p} = (p_1, \ldots, p_N)$ changes as the parameter N is updated. To avoid these difficulties, Dorazio and Royle (2005a) developed a Bayesian model based on the conditional likelihood (described earlier). However, this approach only facilitated the estimation of some latent features of interest and did not entirely eliminate the need to integrate the nuisance parameters from the likelihood.

We develop here a general parameterization of multinomial models with unknown index

N that yields a simple and efficient Bayesian implementation using MCMC methods. The basic strategy is to augment the observed dataset with a fixed, known number, say M, of all-zero capture histories and to model the augmented dataset as a zero-inflated version of the complete-data model using an unknown, but estimable, zero-inflation parameter. We require that $M > N$ which, in most biological applications, can be easily achieved a priori. Indeed, in some problems (e.g., in estimating the size of an animal community), the quantity M has a natural interpretation as the size of a super-community of species from which the local community is drawn. The important implication of our data-augmentation and modeling strategy is that by fixing the size of the dataset, we gain considerable simplicity in implementation, even for very complex models. Consider the heterogeneity model M_h described previously. The proposed data augmentation yields a dataset composed of the observed detection frequencies $y_1, y_2, \ldots, y_n$ and the augmented zeros $y_{n+1}, \ldots, y_M$. Given the augmented dataset, one may adopt a conventional likelihood-based framework for inference wherein the likelihood of the augmented dataset is a zero-inflated binomial mixture [see Section 3, p. 72, and Royle (2006) for examples]. Alternatively, one may adopt a Bayesian framework wherein straightforward applications of Gibbs sampling may be used to compute inferences for N and other model parameters (see Section 2.1).

Our zero-inflated parameterization of the multinomial model and our use of data augmentation were inspired originally by consideration of site-occupancy models in ecology, which we describe in Section 2. However, an appealing Bayesian motivation can also be developed by considering a natural, noninformative prior distribution for N. We describe this in Section 2.1. We provide three applications using data augmentation involving estimating the size of an animal population (Section 3), estimating the prevalence of diabetes in a human population (Section 4), and modeling animal community structure for the estimation of species richness (Section 5). We conclude with a short discussion of our results in Section 6.

2. SITE-OCCUPANCY MODELS

A problem of increasing interest in ecology is estimation of the probability of occurrence (or "site occupancy") of a species that is detected imperfectly. Suppose each of M randomly selected sites is sampled T times yielding a binary sequence $\mathbf{x}_i$ $(i = 1, \ldots, M)$ that indicates whether the species is detected or not on each sampling occasion. (Note that the capture-history concept described earlier in Section 1 applies equally well to the sequence of detections.) At occupied sites where the species is present, we assume that one observes $x = 1$ with probability p and $x = 0$ with probability $1 - p$. At unoccupied sites we assume that one observes $x = 0$ with probability 1. If the T observations from each site are made independently, it is sufficient to reduce the binary sequence of detections to the number of times the species is detected in T attempts, that is, $y_i = \sum_{j=1}^{T} x_{ij}$ for the ith site.

Suppose the species is detected (i.e., $y > 0$) at n of the M sites and that the goal is to estimate the mean probability of occurrence of the species in question. The obvious estimator n/M is biased owing to imperfect detectability. MacKenzie et al. (2002) noted that the likelihood of the detection frequencies, $\{y_i\}$, is a product of zero-inflated binomial densities

$$L(p, \psi|\{y_i\}) = \prod_{i=1}^{M} \psi \text{ Binomial}(y_i|T, p) + I(y_i = 0)(1 - \psi), \tag{2.1}$$

where ψ denotes the parameter of interest, mean site occupancy.

What distinguishes this problem from the conventional N-estimation problem is that, here, the all-zero capture histories are observed, and the size of the dataset (M) is fixed. However, in some instances, there is interest in estimating the *number* of occupied sites from among the M sampled sites. Let this quantity be denoted by N. To facilitate estimation of N, we develop the likelihood conditional on N. Note that sites can be classified into three classes: occupied and detected, occupied and not detected, and unoccupied. This yields a trinomial distribution for the three frequencies n, $N-n$ and $M-N$ which can be rearranged to yield the following specification of the joint likelihood

$$\begin{aligned} &L_J(N, p, \psi \mid \mathbf{y}, M, n) \\ &\quad = \left[\frac{N!}{(N-n)!}\, p^{\sum_{i=1}^{n} y_i}\,(1-p)^{T\cdot N-\sum_{i=1}^{n} y_i}\right]\left[\frac{M!}{N!(M-N)!}\,\psi^{N}(1-\psi)^{M-N}\right] \end{aligned} \tag{2.2}$$

being the distribution of the site-specific detection frequencies and the observed number of occupied sites, n, conditional on N (in the first set of square brackets), and the conditional distribution of N given M (in the second set of square brackets). We note that the first term is the likelihood for N and p that arises in closed population sampling (Williams, Nichols, and Conroy 2002, p. 299) which we will denote as $L_0(N, p \mid \mathbf{y})$. If interest is limited to inference about the population mean ψ, we can remove N from (2.2) by summation to obtain the zero-inflated likelihood (2.1). Alternatively, we might compute the maximum-likelihood estimate of (p, N, ψ) by maximizing the joint likelihood (2.2) directly. Finally, if we are interested only in estimating N, we may integrate ψ from the likelihood, say using a Uniform(0, 1) prior distribution. If we do this, we obtain a truncated likelihood

$$L_T(N, p \mid \mathbf{y}, M) = L_0(N, p \mid \mathbf{y})\,\frac{I(N \le M)}{M+1}, \tag{2.3}$$

which yields an estimate of N that is equivalent to that obtained by maximizing the untruncated likelihood L_0 subject to $N \le M$. Note that $\hat{N}$ obtained by maximizing L_0 is consistent. Thus, we might as well ignore the second component of the likelihood, which is the solution for estimating ψ suggested by Nichols and Karanth (2002).

The factorization in (2.2) establishes an approximate duality between estimating the number of occupied sites, when the all-zero capture histories are observed, and estimating the population size when they are not. Moreover, (2.2) motivates our data augmentation scheme for estimating N. That is, one could apply this duality in the opposite direction—given capture–recapture data wherein the zero detection frequencies are not observed, one may augment the data with an arbitrarily large (relative to n) number of all-zero capture histories and then estimate the proportion of those zeros that were exposed to sampling (i.e., estimate site occupancy).

As a final comment, our factorized likelihood ties nicely into conventional finite-population sampling ideas that exploit "superpopulation" models (Hedayat and Sinha 1991). In our case the superpopulation model is the second component of the likelihood in (2.2), and the issue is whether one is interested in estimating the superpopulation parameter, ψ, or its finite-population manifestation, N.

2.1 A BAYESIAN ANALYSIS OF CLOSED POPULATION MODELS

We have suggested that, when confronted with a population size estimation problem, one can augment the observed detection frequencies with a number of zeros and focus on the problem of estimating a zero-inflation parameter (or its complement) instead of the multinomial index N. But, is adding zeros to the dataset innocuous? Here, we describe a sense in which it is innocuous. We consider the classic problem of estimating the size of a closed population of individuals that are vulnerable to a constant probability of capture p on each of T sampling occasions. (This scenario is designated as "Model M_0" in the capture–recapture literature.) We show that a Bayesian formulation of this model provides a natural and intuitive justification for the data augmentation scheme described in the previous section. Indeed, we prove that the site-occupancy model in which zeros are observed is equivalent to a population-size estimation model in which the zeros are *not* observed, when N is given a discrete uniform prior.

Suppose T samples of a population of size N yields a sample of $n \leq N$ individuals, each observed with detection frequency $y_i > 0$ $(i = 1, \ldots, n)$. Conditioning on N yields a complete-data likelihood $L_0(N, p \mid \mathbf{y})$ (defined earlier as a factor of (2.2)) that includes the detection frequencies of both observed and unobserved individuals. Consider a Bayesian analysis with independent priors on p and N. The posterior distribution is

$$\pi(N, p \mid \mathbf{y}) \propto L_0(N, p \mid \mathbf{y})[N][p].$$

Assume that $[p] \sim \text{Uniform}(0, 1)$. Selection of an objective prior for $[N]$ is less obvious; however, one prior that seems reasonable is a proper, discrete uniform distribution on all integers between 0 and M, which we denote by Duniform$(0, M)$. In this case

$$\pi(N, p \mid \mathbf{y}) \propto L_T(N, p \mid \mathbf{y}, M),$$

that is, the posterior is proportional to the truncated likelihood derived earlier (in (2.3)).

Note that our discrete uniform prior for N also may be constructed hierarchically. Assume N has a binomial prior distribution with index M and success parameter ψ, and assume $[\psi] \sim \text{Uniform}(0, 1)$. This two-stage prior specification yields a model that is, precisely, the factorization of the site occupancy model given in (2.2). Removing ψ from the joint prior $[N|M, \psi][\psi]$ by integration yields our Duniform$(0, M)$ prior for N, that is,

$$\int_0^1 \text{Binomial}(N|M, \psi)\, \mathrm{d}\psi = \text{Duniform}(0, M).$$

Thus, data augmentation of the model is equivalent to a Bayesian formulation of model M_0 that assumes a uniform prior for p and a discrete uniform prior for N.

3. ESTIMATING THE SIZE OF AN ANIMAL POPULATION

We provide an introductory example using the famous snowshoe hare data originally examined by Otis et al. (1978) and analyzed compulsively by statisticians ever since (Cormack 1989; Agresti 1994; Coull and Agresti 1999; Dorazio and Royle 2003). In the hare survey $n = 68$ individuals were captured in six days of trapping with encounter frequencies $\mathbf{n} = (25, 22, 13, 5, 1, 2)'$ where $n_y = \sum_{i=1}^{n} I(y_i = y)$. It is simple enough to obtain the maximum of the likelihood

$$L_0(N, p \mid \mathbf{y}) = \frac{N!}{(N-n)!}\, p^{\sum_i y_i}\, (1-p)^{T \cdot N - \sum_i y_i}$$

which yields $\hat{N} = 74.7$ and $\hat{p} = 0.32$. A Wald-type confidence interval for N is [69.9, 79.5], which is not likely to be of much use as the data appear to exhibit heterogeneity (Otis et al. 1978) and the likelihood is fairly skewed.

Using data augmentation, we introduce a large number of zeros bringing the size of the dataset to M. Given M, we can obtain the MLEs of ψ and p by maximizing (2.1). Adopting this approach, with 100 zeros added to the dataset (and thus $M = 168$), yields $\hat{p} = 0.32$ and $\hat{\psi} = 0.449$. Note that the conditional distribution of $n_0 = N - n$ given the data is binomial with index $M - n$ and parameter $\frac{(1-p)^T \psi}{(1-p)^T \psi + (1-\psi)}$. One could thus use the MLEs to obtain a prediction of n_0 in a manner consistent with classical notions of predicting random effects (i.e., best unbiased prediction). Alternatively, a more direct likelihood solution is obtained by maximizing the joint likelihood (2.2) as a function of the three parameters. This yields $\hat{N} = 74.7$ as before.

We now develop a Bayesian analysis of the augmented snowshoe hare dataset. Let $\{z_i\}$ $(i = 1, \ldots, M)$ denote a collection of partially observed, latent variables, which indicate whether a hare in the superpopulation is available to be captured. We assume $[z_i|\psi] \sim$ Bernoulli(ψ); thus, ψ denotes the probability that an individual hare belongs to those individuals in the superpopulation that are available to be captured. For the n hares observed in the sample, we know $z = 1$; however, the availabilities of the unobserved hares (for whom $y = 0$) are unknown. One convenient way to model these zero-inflated outcomes is to specify y_i conditional on the latent value of z_i as follows:

$$[y_i \mid z_i, p] \sim \text{Binomial}(T, z_i p).$$

We choose this representation because it yields a convenient implementation in the popular software package WinBUGS (Gilks, Thomas, and Spiegelhalter 1994). (The WinBUGS project is located at *http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml*.) The model is completed using our distributional assumption about the latent variables

$$[z_i \mid \psi] \sim \text{Bernoulli}(\psi)$$

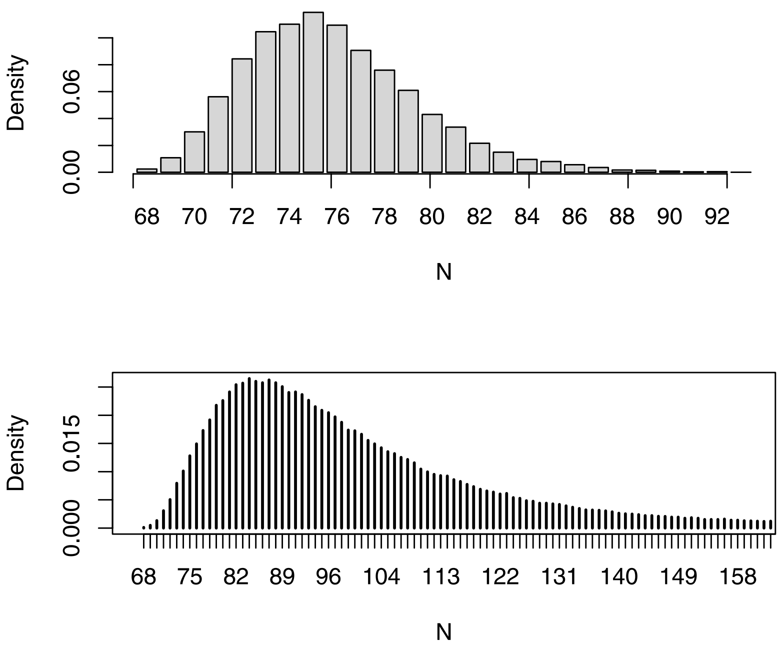

Figure 1. Posterior distribution of N for the hare data under the constant-p (upper panel) and logit-normal heterogeneity models (lower panel).

and by specifying priors for the remaining parameters p and ψ. We choose independent Uniform(0, 1) priors for both of these parameters to specify prior indifference in their magnitudes. Our specification of the model yields an efficient Gibbs sampler composed of the following three steps:

(1) Draw p from a beta distribution with parameters $\sum_{i=1}^{M} y_i z_i + 1$ and $T \sum_{i=1}^{M} z_i - \sum_{i=1}^{M} y_i z_i + 1$.

(2) For each unobserved value of z_i (for which $y_i = 0$), draw a Bernoulli random variable with probability

$$\Pr(z_i = 1 \mid y_i = 0) = \frac{(1-p)^T \psi}{(1-p)^T \psi + (1-\psi)}.$$

(3) Draw ψ from a beta distribution with parameters $1 + \sum_{i=1}^{M} z_i$ and $1 + M - \sum_{i=1}^{M} z_i$.

Note that the posterior distribution of N can be obtained simply by calculating $N = \sum_{i=1}^{M} z_i$ at each step of the Gibbs sampler. For these data the estimated posterior distribution of N has a median of 75.0 and a 95% credible interval of [70, 84] (Figure 1).

We return now to model M_h, which assumes heterogeneity in capture probabilities among individuals in the population. This class of models has received considerable attention in recent years (Norris and Pollock 1996; Coull and Agresti 1999; Fienberg et al. 1999; Pledger 2000; Basu and Ebrahimi 2001; Tardella 2002; Dorazio and Royle 2003; Link 2003). The conventional method (based on integrated likelihood) of fitting heterogeneity

```
model; {
   mup ~ dnorm(0,.01)
   taup ~ dgamma(.01,.01)
   psi ~ dunif(0,1)
   for(i in 1:M){
     z[i]~ dbin(psi,1)
     b[i] ~ dnorm(mup,taup)
     logit(p[i]) <- b[i]
     mu[i]<- p[i]*z[i]
     x[i] ~ dbin(mu[i],T)
   }
   N<-sum(z[]) }
```

Figure 2. WinBUGS model specification for the logit-normal model of heterogeneity.

models can be found in many of these references. Here we provide a Bayesian analysis of a heterogeneity model fit to the snowshoe hare data using our proposed data augmentation scheme.

As before, we introduce a set of partially observed, Bernoulli outcomes $\{z_i\}$. The model consists of a conditional likelihood $[y_i|z_i, p_i] \sim \text{Binomial}(T, z_i p_i)$, a model for the latent indicator variables $[z_i|\psi] \sim \text{Bernoulli}(\psi)$, and a Uniform(0, 1) prior on ψ. To complete the model, we specify heterogeneity in capture probabilities using the logit-normal specification of Coull and Agresti (1999) in which $\log(p_i/(1-p_i)) = b_i$ with $[b_i \mid \mu_p, \sigma^2] \sim \text{Normal}(\mu_p, \sigma_p^2)$. The model is completed by specifying prior distributions for μ_p and σ_p^2. We assumed $[\mu_p] \sim \text{Normal}(0, 100)$, and used a gamma prior on $\tau_p = 1/\sigma_p^2$ with parameters $\alpha = 0.01$ and $\beta = 0.01$. This parameterization has mean and variance α/β and α/β^2, respectively. We implemented the model in the freely available software package WinBUGS. This is hardly more difficult than providing a pseudo-code description of the model. The WinBUGS model specification for the logit-normal heterogeneity model is given in Figure 2. Relevant discussion of conventional likelihood analysis of the data-augmented heterogeneity model can be found in Royle (2006).

Summaries of the posterior distribution were calculated from four independent Markov chains initialized with random starting values, run 200,000 times after a 10,000 burn-in and resampling every four draws. This yielded four sets of 50,000 posterior draws. We computed the Brooks-Gelman-Rubin convergence diagnostic (Gelman and Rubin 1992; Brooks and Gelman 1997) using the output from the four chains. Values of this statistic, referred to as the scale reduction factor, near 1.0 indicate convergence. For the hare data, multivariate potential scale reduction factor was 1.017, and for each parameter the potential scale reduction factors were all less than 1.001 excepting for τ which had a scale reduction factor of 1.015.

As before, the posterior of N may be obtained by calculating $N = \sum_{i=1}^{M} z_i$ at each step of the Gibbs sampler. For the snowshoe hare data the posterior mean and median of N under this model were 104.1 and 95, respectively. The posterior distribution is shown in Figure 1 (lower panel). The results are slightly different than reported by others due to the skew of the posterior distribution. For example, Dorazio and Royle (2003) estimated N from the integrated likelihood and reported $\hat{N} = 91.7$ (74.7, 150.6) under the same logistic-normal model. The estimated (posterior means) parameters of the heterogeneity distribution were $\widehat{E[p]} = 0.21$ and $\hat{\sigma} = 1.127$ indicating substantial heterogeneity and relatively low detection probabilities.

4. ESTIMATING THE PREVALENCE OF DIABETES

We consider here an analysis of data from an epidemiological survey designed to estimate the prevalence of diabetes among residents of a small town in northern Italy. The data, originally reported and analyzed by Bruno et al. (1994), include records of $n = 2{,}069$ residents that were cross-classified into each of 15 ($= 2^4 - 1$) categories depending on whether the residents' names appeared on each of four lists: (1) prescriptions (a computerized list of insulin and hypoglycemic prescriptions); (2) reimbursements (all residents who requested a reimbursement for insulin and reagent strips); (3) clinics (all patients diagnosed as diabetics by the local clinic and/or family physicians); and (4) hospitals (all patients discharged with a primary or secondary diagnosis of diabetes in all private and public hospitals in the region). Therefore, each resident's records may be summarized using a binary vector $\mathbf{x}$ of $T = 4$ elements to indicate whether an individual appeared ($x = 1$) or did not appear ($x = 0$) on each list. The objective of the analysis is to estimate the number of residents with diabetes that did not appear on any of the four lists.

The diabetes data have been analyzed using a variety of models (Yip et al. 1995; Biggeri, Stanghellini, Merletti, and Marchi 1999; Fienberg et al. 1999; Bartolucci and Forcina 2001). It is not our intention here to complete an exhaustive analysis of these data. Instead, we demonstrate how data augmentation may be used to fit a Bayesian version of a Rasch model quite easily. The particular model under consideration was developed by Fienberg et al. (1999, see Section 4), whose Gibbs sampler is more challenging to implement than our approach because it requires the numerical approximation of some integrals that cannot be evaluated as part of the MCMC sampling.

We begin by summarizing our data-augmentation version of the Rasch model. First, we create a superpopulation of residents by adding an arbitrarily large, but known, number of all-zero vectors to the n observed vectors. Therefore, our augmented data includes an $M \times T$ matrix $\mathbf{X}$, whose first n rows are observed and whose last $M - n$ rows are unobserved (i.e., $\mathbf{x}_i = \mathbf{0}$, $i = n + 1, \ldots, M$). We introduce a latent indicator variable z_i which takes the value 1 if an individual in the superpopulation is a diabetic that potentially could have been recorded in the multiple-list survey; otherwise $z_i = 0$. We assume that $\{z_i\}$ are independent, Bernoulli-distributed random variables indexed by parameter ψ. Obviously, z_i is observed for $i = 1, \ldots, n$, but not otherwise. By introducing the superpopulation of latent variables $\{z_i\}$ into the model, we effectively transform the problem of estimating N into the equivalent problem of estimating $\sum_{i=1}^{M} z_i$, which, of course, depends on the estimated value of ψ.

Let p_{ij} denote the probability that the ith individual in the superpopulation appears on the jth list ($j = 1, \ldots, T$). If the ith individual is a diabetic, we assume

$$[x_{ij} \mid p_{ij}, z_i = 1] \sim \text{Bernoulli}(p_{ij})$$

If the ith individual is not diabetic ($z_i = 0$), we assume $x_{ij} = 0$ ($\forall j$) with probability 1. Following the Rasch model, the heterogeneity in p_{ij} is partitioned into an individual component b_i and a list component β_j, which are additive on the logit scale: $\log(p_{ij}/(1-p_{ij})) = b_i+\beta_j$. The list component β_j is identical for all individuals. Variation among the individual components, on the other hand, is specified with a zero-centered normal distribution: $[b_i \mid \sigma^2] \sim \text{Normal}(0, \sigma^2)$. Our data-augmentation version of the Rasch model is completed by specifying prior distributions for ψ, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_T)'$ and σ^2. Indifference in the magnitude of ψ is specified by assuming a uniform prior $[\psi] \sim \text{Uniform}(0, 1)$. Following Fienberg et al. (1999), we assume $[\boldsymbol{\beta}] \sim \text{Normal}(\mathbf{0}, 10\mathbf{I})$, $[\sigma^2] \sim \text{InvGamma}(1, 1)$, and mutual independence $[\psi, \boldsymbol{\beta}, \sigma^2] = [\psi][\boldsymbol{\beta}][\sigma^2]$. In addition to these priors, we also carried out a Bayesian analysis using the following priors $\text{logit}^{-1}(b_j) \sim \text{Uniform}(0, 1)$ and $\sigma \sim \text{Uniform}(0, 100)$ (Gelman 2005).

Bayesian analysis by data augmentation under both sets of priors was conducted by augmenting the $n = 2{,}069$ observed vectors of data with 2,000 all-zero vectors. Thus, the superpopulation size here is $M = 4{,}069$. For comparison, we also computed the MLE of model parameters by maximizing the likelihood having removed the individual effects by numerical integration (Coull and Agresti 1999). Inferences about population size N (and other parameters) were essentially identical among these modes of inference and priors, owing to the large sample size. For example, the MLE of N was $\hat{N} = 2{,}697$, whereas the posterior median of N was 2,693 using the priors of Fienberg et al. (1999), and 2,699 under the uniform prior specification. These estimates are remarkably similar to those computed by Fienberg et al. (1999), who reported a posterior mode of $\hat{N} = 2{,}693$ and a 95% credible interval of [2,567, 2,906].

Subsequently, we discuss some salient points of the analysis based on the Fienberg et al. (1999) priors, as results were similar under the uniform prior specification. Summaries of the posterior distribution were calculated from four independent Markov chains initialized with random starting values, run 50,000 times after a 10,000 burn-in and resampling every five draws. This yielded four sets of 10,000 posterior draws. We computed the Brooks-Gelman-Rubin convergence diagnostic (Gelman and Rubin 1992; Brooks and Gelman 1997) using the output from the four chains. Values of this statistic, referred to as the scale reduction factor, near 1.0 indicate convergence. For the diabetes data, the multivariate potential scale reduction factor was 1.001, and for each parameter the potential scale reduction factors were all less than 1.002.

The posterior mass of N was located far from the upper support of 4,069, with the maximum N of 3,148 observed in the posterior sample. This suggests that the chosen value of M was adequate. The 95% interval for N based on the 2.5% and 97.5% percentiles of the posterior was (2,535, 2,915). The posterior median of σ was 1.235 (95% interval: (1.046, 1.454)). The posterior means of β_j were (-0.412, -3.285, 0.808, -2.031) indicating substantial variation in detectability among lists (the posterior intervals did not overlap for any of the four parameters).

5. ESTIMATING SPECIES RICHNESS WITH A MULTISPECIES, SITE-OCCUPANCY MODEL

We consider here a complex multinomial model with unknown index. In this example the index N denotes the number of species in an animal community. Unlike the previous examples, the community is sampled in a manner that allows us to model the probabilities of detection *and* occurrence of individual species. Specifically, we consider surveys wherein each of J sites is visited K times and the identities of all species detected during each visit are noted. The purpose of temporal replication at each sample location is to provide the information needed to estimate the probability of detecting a species, given that it is present, separately from its probability of occurrence. However, the total duration of the survey must be sufficiently short that N may safely be assumed to remain constant in the time required to complete the survey. Therefore, the traditional "closure" assumption, which precludes an addition (or subtraction) of species in the community as a consequence of local colonization (or extinction) events, is assumed to be satisfied.

This sampling protocol requires a slight modification of the notation used in our earlier examples of population surveys. In particular, let x_{ij} denote the number of times that species i $(= 1, \ldots, N)$ is detected in $K > 1$ visits to site j $(= 1, \ldots, J)$. Thus, x_{ij} is not binary as in our earlier examples. At the completion of the survey, suppose $n < N$ distinct species are actually detected. By allowing $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iJ})'$ to denote the vector of J site-specific observations of species i $(i = 1, \ldots, n)$, we note that $\mathbf{x}_i = \mathbf{0}$ for the $N - n$ undetected species in the community for which $i = n + 1, \ldots, N$. Therefore, for reasons that will become clear shortly, we have combined the observed data and the all-zero vectors into an ordered $N \times J$ matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}, \ldots, \mathbf{x}_N)'$. Using the same ordering, we also define an $N \times J$ matrix $\mathbf{Z}$ of binary indicators, whose elements denote the presence $(z_{ij} = 1)$ or absence $(z_{ij} = 0)$ of species i at site j. Note that $\mathbf{Z}$ is only partially observed. A species must be present at a site before it can be detected; therefore, $z_{ij} = 1$ if $x_{ij} > 0$. However, if $x_{ij} = 0$, there are two mutually exclusive possibilities for the value of z_{ij}: (1) species i is present at site j but undetected $(z_{ij} = 1)$, or (2) species i is absent at site j $(z_{ij} = 0)$.

Dorazio and Royle (2005a) developed a conventional, multinomial model of the observed portion of $\mathbf{X}$ and showed that allowing the occurrence of species to vary spatially (i.e., among sample sites) through the definition of $\mathbf{Z}$ provides a straightforward, but computationally intensive, mechanism for estimating N and other quantities of ecological interest. Here we use data augmentation and develop an alternative parameterization of the multinomial model that is more amenable to analysis by MCMC sampling. Importantly, our reparameterized model can be specified in the BUGS language (Gilks et al. 1994) and easily fitted using the freely available WinBUGS software (*http://www.mrc-bsu.cam.ac.uk/bugs/*).

To motivate the use of data augmentation, we note that if N was known, the hierarchical model for estimating species occurrence and detection parameters would be complete without requiring special considerations in the likelihood function for an "undetected" portion of the community. Furthermore, there would be no difficulty in fitting the model using MCMC. The difficulty with N being unknown is that the dimension of the parameter vectors associated with species occurrence and detection changes each time another MCMC draw of the parameter N is computed. To obtain a model in which the dimension of the parameters is constant, we create a super-community of species, one that comprises the n observed species and an arbitrarily large, but known, number of unobserved species for which $\mathbf{x}_i = \mathbf{0}$ ($i = n+1, n+2, \ldots, N, N+1, \ldots, M$). The super-community size M is fixed, and thus the dimension of the parameter vectors is constant (i.e., not a function of N). In taking this approach, we do not directly estimate N as a parameter. Instead, we introduce an additional latent indicator variable, say w_i, which takes the value 1 if a species in the super-community is a member of the N species that are vulnerable to sampling and 0 otherwise. We assume that $\{w_i\}$ are independent, Bernoulli-distributed random variables indexed by parameter Ω. Obviously, w_i is observed for $i = 1, \ldots, n$, but not otherwise. By introducing the super-community of latent variables $\{w_i\}$ into the model, we effectively transform the problem of estimating N into the equivalent problem of estimating $\sum_{i=1}^{M} w_i$, which, of course, depends on the estimated value of Ω.

Our reparameterized model does require that M be assigned a sufficiently high value; however, in practice it is a simple matter to assess the adequacy of any particular choice of M. Recall that $\mathrm{E}(N) = M\,\Omega$ and that Ω is bounded between 0 and 1; therefore, estimates of Ω will necessarily decline as higher values of M are chosen. If the assigned value of M is too low, the posterior distribution of Ω will be concentrated near the upper limit of its support, and we risk underestimating the true value of N. The obvious solution is to increase M until the posterior of Ω is centered well below its upper limit. However, higher values of M also imply higher computational costs, so some care is advised in assigning too high a value to M. We note that both N and M have practical interpretations in this problem, N being the number of species present on the *sampled* units, and M being the number of species in some (potentially hypothetical) super-community, for which the sample units were (randomly) selected.

5.1 Modeling Occurrence and Detection

We model the augmented $M \times J$ matrix $\mathbf{X}$ using the same assumptions described by Dorazio and Royle (2005a), except that now the indicator variables $\{w_i\}$ are included in the model. Specifically, if the ith species in the super-community is available to be sampled (i.e., $w_i = 1$), its occurrence at site j is modeled as a Bernoulli outcome with probability ψ_{ij},

$$[z_{ij} \mid \psi_{ij},\, w_i = 1] \sim \text{Bernoulli}(\psi_{ij})$$

otherwise, we assume $z_{ij} = 0$ with probability 1. Likewise, if the ith species is present at site j (i.e., $z_{ij} = 1$), we assume a binomial(K, θ_{ij}) distribution for the number of detections x_{ij},

$$[x_{ij} \mid \theta_{ij},\, z_{ij} = 1] \sim \text{Binomial}(K, \theta_{ij})$$

otherwise, we assume $x_{ij} = 0$ with probability 1. We use a hierarchical model to specify variation in occurrence and detection probabilities among species. The effects of species- and site-specific differences in rates of occurrence and detection are parameterized on the logit scale as follows: logit $\psi_{ij} = u_i + \alpha_j$ and logit $\theta_{ij} = v_i + \beta_j$, where u_i and v_i denote species-level effects, and α_j and β_j denote site-level effects. The species-level effects are assumed to be centered at zero; therefore, α_j denotes a logit-scale parameter for the mean probability of occurrence among all species at site j, and β_j denotes a logit-scale parameter for the mean probability of detection among all species at site j. A linear combination of parameters and site-level covariates may be substituted for α_j or β_j, assuming of course that such covariates are available and are thought to be informative about the magnitude of ψ_{ij} or θ_{ij}. In the absence of site-level covariates (as in our example dataset; see Section 5.2), we assume that α_j and β_j have constant values, say α and β, at each of the J sites.

Species-specific differences in the probabilities of occurrence and detection are modeled by specifying a parametric form for the joint distribution of u_i and v_i. For example, we assume $[u_i, v_i \mid \mathbf{\Sigma}] \sim \text{Normal}(\mathbf{0}, \mathbf{\Sigma})$, which allows us to specify the heterogeneity in occurrence and detection among species using only a few parameters (specifically, σ_u^2, σ_v^2, and σ_{uv}, the unique elements of the 2×2 matrix $\mathbf{\Sigma}$). We complete our reparameterized model by assuming mutually independent prior distributions for Ω, α, β,and $\mathbf{\Sigma}$. In particular, we assume a Uniform(0,1) prior for Ω, $\text{logit}^{-1}(\alpha)$, and $\text{logit}^{-1}(\beta)$; Inverse-Gamma(a,b) priors for σ_u^2 and σ_v^2 ($a = 0.1$ and $b = 10$ denote shape and scale parameters, respectively); and a Uniform(-1,1) prior for the correlation parameter $\rho_{uv} = \sigma_{uv}/\sigma_u\sigma_v$. The Inverse-Gamma($\epsilon$, $1/\epsilon$) distribution, for some small ϵ, is often used as a default or objective prior of variance parameters, particularly in models that maintain conjugacy between the prior and posterior distributions (e.g., see Carlin and Louis 2000, p. 149). Similarly, we use uniform distributions to specify our prior indifference in the mean probabilities of detection and occurrence, in the Ω parameter, and in the correlation parameter ρ_{uv}.

5.2 Analysis of an Avian Community

We describe here an analysis of birds observed along a single route in Maytown, Alabama. This route is part of a continental-scale survey which includes more than 4000 roadside routes located in North America (Robbins et al. 1986, 1989; Sauer, Pendleton, and Peterjohn 1996). Each route is 39.4 km and contains 50 equally spaced sites. At each of these sites an observer records the number and identity of each species detected (visually or aurally) within a three-minute period. In most years each roadside route is visited only once annually; however, in 1991 several routes were sampled repeatedly during the breeding season to evaluate the variation in bird counts both between and within sites. The data used in our analysis were collected at one of these routes, which was visited by the same observer on 11 different days in the month of June.

Seventy-five species of birds were detected along this route, and there was considerable variation among species in the observed frequencies of detection at each site (Dorazio and Royle 2005a). To compute the posterior distribution of species richness, we augmented the observed 75×50 matrix of species- and site-specific detection frequencies with 100 all-zero detection frequencies. On fitting our model of the augmented dataset, we found that the estimated size of the community exceeds the number of species observed in the sample by a substantial margin (Figure 3). In fact, the posterior probability that the avian community comprises only $N = 75$ species is essentially zero, and the estimated median

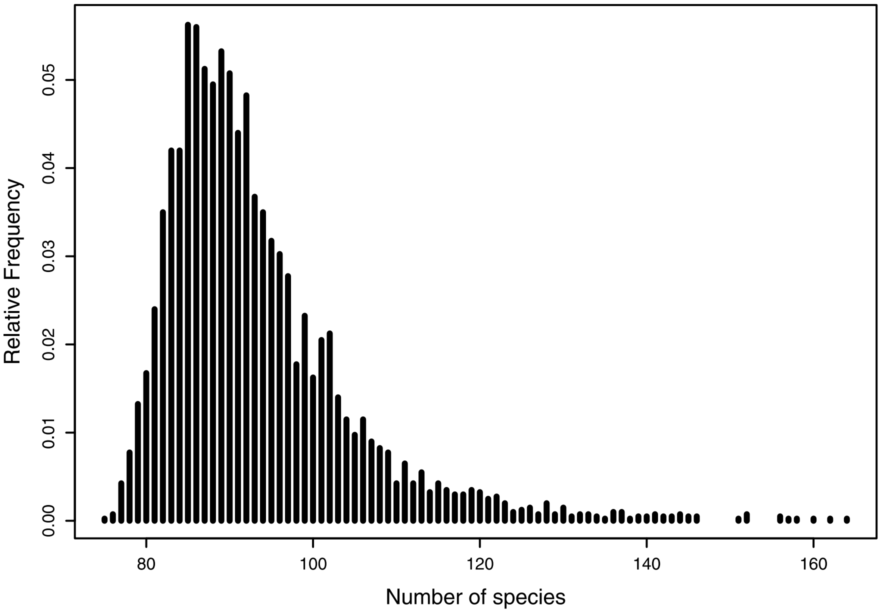

Figure 3. Posterior distribution of species richness in a community of breeding birds.

and mean values of species richness are 90.0 and 93.0, respectively.

These results are consistent with our estimated levels of heterogeneity in species occurrence and detection, which suggest that detection failures in many bird species are attributed to low rates of occurrence, as opposed to simply low rates of detection. In other words, a substantial portion of the community includes relatively uncommon species, so it is not surprising that the estimated total number of species in the community exceeds the number of species observed in the sample.

6. DISCUSSION

Multinomial models wherein the multinomial index is unknown are common in many disciplines including ecology, where the focus is usually on estimating the size of an animal population, or the number of species in a community. Similar problems arise in other disciplines including epidemiology, census adjustment, and computer science. In this article, we develop a data augmentation scheme for the analysis of multinomial models with unknown index. Under this approach, the observed dataset is augmented with a large number of all-zero capture histories yielding a fixed dataset of size $M > N$. The augmented dataset is modeled as a zero-inflated version of the complete-data model using an unknown, but estimable, zero-inflation parameter which essentially plays the role of the multinomial index parameter N. Choice of M can influence parameter estimates if too small a value is used. We advocate that M be chosen by trial-and-error, such that mass of N is not concentrated near M. Higher values of M also imply greater computational costs, so some care is advised in assigning too high a value to M a priori.

Our initial motivation for adopting data augmentation in this class of problems was purely pragmatic, and arose by consideration of a related class of models, known as site occupancy models in ecology (described in Section 2), in which the “all zero” encounter histories are observed. However, formal Bayesian arguments can be used to motivate data augmentation. In particular, a discrete uniform prior for the unknown multinomial index yields precisely the motivating site occupancy model in which zeros are observed. Thus, augmenting a dataset with zeros merely induces a discrete uniform prior for N, a consequence that appears fairly innocuous so far as prior specification is concerned.

Another approach for dealing with the variable dimension parameter space in this class of multinomial models is to regard N as a fixed model index and to treat its estimation as a model-selection problem, using methods such as those suggested by Carlin and Chib (1995) and King and Brooks (2001). The method of Carlin and Chib (1995), a trans-dimensional Gibbs sampler, has recently been implemented by Durban and Elston (2005) in the analysis of mark–recapture models. This Gibbs sampler requires an analyst to select candidate generating distributions and “pseudo-priors” which, at this point, seems to be as much an art as a science. Identification of objective and appropriate pseudo-priors or candidate generating distributions is not straightforward, nor are specific choices evidently innocuous (Link and Barker 2006). Thus, converting the N-estimation problem to one of model-selection yields a method of dealing with a problem that does not exist under our data augmentation scheme (i.e., the problem of moving from one model to another). Our use of data augmentation is consistent in form and function with classical motivations and methods of data augmentation (Tanner and Wong 1987; Gelman 2004). In particular, we use data augmentation to simplify the calculations involved in simulation-based methods of Bayesian analysis *and* to reparameterize the multinomial model using parameters that have a direct interpretation. In this reparameterized model the problem of estimating N is transformed into the equivalent problem of estimating a sum of latent indicator variables, each of which denotes the status of an individual’s population membership (i.e., whether the individual is a member of the population exposed to sampling), or “occupancy” status under the motivating class of models. Thus, our use of data augmentation establishes a conceptual connection between classes of multinomial models with unknown index and the class of site-occupancy models described in Section 2.

Furthermore, our approach is not an example of the trans-dimensional Gibbs sampler proposed by Carlin and Chib (1995). Like Durban and Elston (2005), we do fix the size of the dataset by augmenting it with a collection of “all zero” encounter histories. However, we take advantage of the fact that the augmented data yields a *reparameterization* of the model so that N is no longer present as a formal parameter and so that one prior distribution (on logit-scale detection) may be specified for all individuals in the superpopulation, regardless of whether they are observed or unobserved (= augmented), and regardless of whether the unobserved individuals are members of the population exposed to sampling (i.e., those labeled $n+1$ to N). Furthermore, our approach does not require “tuning parameters” or trial runs of the algorithm because the concept of a pseudo-prior, which is apparently necessary for convergence of trans-dimensional Gibbs samplers, does not apply to our model. In our

```
model {
   psi~dunif(0,1)
   mulogit~dnorm(0,0.1)
   tau.th~dgamma(0.5,0.5)
   sigma2.th <- 1/tau.th
   sigma.th <- sqrt(sigma2.th)
   tau.a ~ dgamma(0.5,0.5)

   sigma2.a <- 1/tau.a
   sigma.a <- sqrt(sigma2.a)
   N<-sum(z[])
   for (j in 1:J) {
     alpha[j] ~ dnorm(0, tau.a)
   }
   for (i in 1:M) {
     z[i]~dbern(psi)
     theta[i]~dnorm(0,tau.th)
     for (j in 1:J) {
       logit(p[i, j]) <- mulogit + theta[i] + alpha[j]
       muY[i,j]<-z[i]*p[i,j]
       Y[i, j] ~ dbern(muY[i,j])
     }
   }
}
```

Figure 4. WinBUGS *model specification for fitting the Rasch model to the snowshoe hare data.*

opinion the trans-dimensional Gibbs sampler lacks flexibility and generality. For example, we cannot envision a direct implementation of the model described in Section 5 using the method of Durban and Elston (2005). In contrast, our approach, which combines data augmentation and reparameterization, can be extended to fit more complex models with no additional challenges (see, e.g., Dorazio, Royle, Söderström, and Glimskär 2006). The practical difference between the trans-dimensional Gibbs sampler as applied by Durban and Elston (2005) and our model of augmented data is evident in Figure 4, where we provide WinBUGS model specification for fitting the Rasch model to the snowshoe hare data analyzed by Durban and Elston (2005), using their notation and prior specification. (The WinBUGS model specification needed to implement the trans-dimensional Gibbs sampler is referenced in their article). We note that most of their code is rendered unnecessary if one adopts our reparameterized model of the augmented data.

The main advantage of the proposed data augmentation approach over other methods of analyzing such models is that it permits a straightforward Bayesian implementation, even in very complex models, because the size of the dataset is fixed. Indeed, the data augmentation approach described here is a direct consequence of our seeking a practical implementation of the animal community model described in Section 5. Implementation of this model using data augmentation in the popular software package WinBUGS requires little more than a dozen instructions defining the probability model, as described by Dorazio et al. (2006). Contrast this with the effort put forth by Dorazio and Royle (2005a) or in related models by Fienberg et al. (1999), or using the model-selection approach of Durban and Elston (2005).

We have demonstrated the application of this approach to classical animal population size estimation problems (including those with heterogeneity in detectability among individuals), the Rasch model for estimating the number of diabetes cases, and also a more complicated model of animal community structure containing heterogeneous occurrence and detection probabilities. We believe that this approach should yield efficient solutions to other problems, such as models in which individual covariates influence detectability (Royle 2007), or open population ("Jolly-Seber") models with heterogeneity.

Finally, one interesting consequence of the Bayesian justification is that it suggests the number of observed zeros obtained from site occupancy surveys provides little information about the number of occupied sites. Indeed, we noted a factorization of the site occupancy likelihood (Equation 2.2), which implies, when M is large, that the zero observations can be discarded with little loss of efficiency. However, in some situations, the slight information provided by the upper bound on N may be important. For example, in models for estimating the size of a closed population, Link (2003) showed that population size, N, is not identifiable across classes of mixture distributions for p. This phenomenon appears less important in site occupancy models (Royle 2006), which seems logical in the context of the discrete uniform prior on N that the observed zeros imply. That is, when the upper bound (M) is not too large, this can introduce enough information to mitigate the Link (2003) problem. Thus, while Dorazio and Royle (2005b) suggested resolving the identifiability problem by judicious choice of priors for p, some attention might also be paid to accommodating prior information on N.

[Received August 2005. Revised May 2006.]

REFERENCES

Agresti, A. (1994), "Simple Capture–Recapture Models Permitting Unequal Catchability and Variable Sampling Effort," *Biometrics*, 50, 494–500.

Bartolucci, F., and Forcina, A. (2001), "Analysis of Capture-Recapture Data with a Rasch-Type Model Allowing for Conditional Dependence and Multidimensionality," *Biometrics*, 57, 714–719.

Basu, S., and Ebrahimi, N. (2001), "Bayesian Capture-Recapture Methods for Error Detection and Estimation of Population Size: Heterogeneity and Dependence," *Biometrika*, 88, 269–279.

Biggeri, A., Stanghellini, E., Merletti, F., and Marchi, M. (1999), "Latent Class Models for Varying Catchability and Correlation Among Sources in Capture-Recapture Estimation of the Size of a Human Population," *Statistica Applicata*, 11, 563–586.

Brooks, S. P., and Gelman, A. (1997), "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455.

Bruno, G., Biggeri, A., Merletti, F., LaPorte, R., McCarty, D. J., and Pagano, G. (1994), "Applications of Capture-Recapture to Count Diabetes," *Diabetes Care*, 17, 548–556.

Carlin, B. P., and Chib, S. (1995), "Bayesian Model Choice via Markov Chain Monte Carlo," *Journal of the Royal Statistical Society*, Series B, 57, 473–484.

Carlin, B. P., and Louis, T. A. (2000), *Bayes and Empirical Bayes Methods for Data Analysis* (2nd ed.), Boca Raton, FL: Chapman and Hall.

Cormack, R. M. (1989), "Log-Linear Models for Capture-Recapture," *Biometrics*, 45, 395–413.

Coull, B. A., and Agresti, A. (1999), "The Use of Mixed Logit Models to Reflect Heterogeneity in Capture-Recapture Studies," *Biometrics*, 55, 294–301.

Darroch, J. N., Fienberg, S. E., Glonek, G. F. V., and Junker, B. W. (1993), "A Three-Sample Multiple-Recapture Approach to Census Population Estimation with Heterogeneous Catchability," *Journal of the American Statistical Association*, 88, 1137–1148.

Dorazio, R., Royle, J., Söderström, B., and Glimskär, A. (2006), "Estimating Species Richness and Accumulation by Modeling Species Occurrence and Detectability," *Ecology*, 87, 842–854.

Dorazio, R. M., and Royle, J. A. (2003), "Mixture Models for Estimating the Size of a Closed Population When Capture Rates vary Among Individuals," *Biometrics*, 59, 351–364.

——— (2005a), "Estimating Size and Composition of Biological Communities by Modeling the Occurrence of Species," *Journal of the American Statistical Association*, 100, 389–398.

——— (2005b), Rejoinder to "The Performance of Mixture Models in Heterogeneous Closed Population Capture-Recapture," *Biometrics*, 61, 874–876.

Durban, J. W., and Elston, D. A. (2005), "Mark-Recapture With Occasion and Individual Effects: Abundance Estimation Through Bayesian Model Selection in a Fixed Dimensional Parameter Space," *Journal of Agricultural, Biological, and Environmental Statistics*, 10, 291–305.

Fienberg, S. E., Johnson, M. S., and Junker, B. W. (1999), "Classical Multilevel and Bayesian Approaches to Population Size Estimation Using Multiple Lists," *Journal of the Royal Statistical Society of London*, Series A, 163, 383–405.

Gelman, A. (2004), "Parameterization and Bayesian Modeling," *Journal of the American Statistical Association*, 99, 537–545.

——— (2005), "Prior Distributions for Variance Parameters in Hierarchical Models," *Bayesian Analysis*, 1, 1–19.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–511.

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994), "A Language and Program for Complex Bayesian Modelling," *The Statistician*, 43, 169–178.

Hedayat, A. S., and Sinha, B. K. (1991), *Design and Inference in Finite Population Sampling*, New York: John Wiley.

King, R., and Brooks, S. P. (2001), "On the Bayesian Analysis of Population Size," *Biometrika*, 88, 317–336.

Link, W. A. (2003), "Nonidentifiability of Population Size from Capture-Recapture Data with Heterogeneous Detection Probabilities," *Biometrics*, 59, 1123–1130.

Link, W., and Barker, R. J. (2006), "Model Weights and the Foundations of Multi-Model Inference," *Ecology*, (in press).

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., and Langtimm, C. A. (2002), "Estimating Site Occupancy Rates When Detection Probabilities are Less Than One," *Ecology*, 83, 2248–2255.

Nichols, J. D., and Karanth, K. U. (2002), "Statistical Concepts: Assessing Spatial Distributions," in *Monitoring Tigers and Their Prey: A Manual for Researchers, Managers and Conservationists in Tropical Asia*, eds. K. U. Karanth and J. D. Nichols, Bangalore, India: Centre for Wildlife Studies, pp. 29–38.

Norris, J. L. III, and Pollock, K. H. (1996), "Nonparametric MLE Under Two Closed Capture-Recapture Models with Heterogeneity," *Biometrics*, 52, 639–649.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978), "Statistical Inference From Capture Data on Closed Animal Populations," *Wildlife Monographs*, 62, 1–135.

Pledger, S. (2000), "Unified Maximum Likelihood Estimates for Closed Capture-Recapture Models Using Mixtures," *Biometrics*, 56, 434–442.

Robbins, C. S., Bystrak, D., and Geissler, P. H. (1986), "The Breeding Bird Survey: Its First Fifteen Years, 1965–1979," Resource Publication 157, United States Fish and Wildlife Service, Washington, D.C.

Robbins, C. S., Sauer, J. R., Greenberg, R. S., and Droege, S. (1989), “Population Declines in North American Birds that Migrate to the Neotropics,” in *Proceedings of the National Academy of Sciences (USA)*, 86, pp. 7658–7662.

Royle, J. (2006), “Site Occupancy Models with Heterogeneous Detection Probabilities,” *Biometrics*, 62, 97–102.

——— (2007), “Analysis of Capture-Recapture Models With Individual Covariates,” unpublished manuscript.

Sanathanan, L. (1972), “Estimating the Size of a Multinomial Population,” *Annals of Mathematical Statistics*, 43, 142–152.

Sauer, J. R., Pendleton, G. W., and Peterjohn, B. G. (1996), “Evaluating Causes of Population Change in North American Insectivorous Songbirds,” *Conservation Biology*, 10, 465–478.

Tanner, M. A., and Wong, W. H. (1987), “The Calculation of Posterior Distributions by Data Augmentation” (with discussion), *Journal of the American Statistical Association*, 82, 528–550.

Tardella, L. (2002), “A New Bayesian Method for Nonparametric Capture-Recapture Models in Presence of Heterogeneity,” *Biometrika*, 89, 807–817.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002), *Analysis and Management of Animal Populations*, San Diego, CA: Academic Press.

Wolter, K. M. (1986), “Some Coverage Error Models for Census Data,” *Journal of the American Statistical Association*, 81, 338–346.

Yip, P. S. F., Bruno, G., Tajima, N., Seber, G. A. F., Buckland, S. T., Cormack, R. M., Unwin, N., Chang, Y. F., Fienberg, S. E., Junker, B. W., LaPorte, R. E., Libman, I. M., and McCarty, D. J. (1995), “Capture-Recapture and Multiple-Record Systems Estimation. I: History and Theoretical Development,” *American Journal of Epidemiology*, 142, 1047–1058.