

# Mixture Models for Estimating the Size of a Closed Population When Capture Rates Vary among Individuals

Robert M. Dorazio

U.S. Geological Survey, Florida Caribbean Science Center, 7920 NW 71 Street,  
Gainesville, Florida 32653, U.S.A.  
*email:* bdorazio@usgs.gov

and

J. Andrew Royle

U.S. Fish and Wildlife Service, Division of Migratory Bird Management,  
11510 American Holly Drive, Laurel, Maryland 20708, U.S.A.  
*email:* andy\_royle@fws.gov

**SUMMARY.** We develop a parameterization of the beta-binomial mixture that provides sensible inferences about the size of a closed population when probabilities of capture or detection vary among individuals. Three classes of mixture models (beta-binomial, logistic-normal, and latent-class) are fitted to recaptures of snowshoe hares for estimating abundance and to counts of bird species for estimating species richness. In both sets of data, rates of detection appear to vary more among individuals (animals or species) than among sampling occasions or locations. The estimates of population size and species richness are sensitive to model-specific assumptions about the latent distribution of individual rates of detection. We demonstrate using simulation experiments that conventional diagnostics for assessing model adequacy, such as deviance, cannot be relied on for selecting classes of mixture models that produce valid inferences about population size. Prior knowledge about sources of individual heterogeneity in detection rates, if available, should be used to help select among classes of mixture models that are to be used for inference.

**KEY WORDS:** Beta-binomial; Heterogeneity; Latent-class; Logistic-normal; Species richness.

## 1. Introduction

The Schnabel census (Schnabel, 1938) is a commonly used survey method in which individual members of a countable population are repeatedly sampled (at different times or locations), noting the presence or absence of each individual. The objective is to estimate the number of individuals in the population that are never observed, and thereby obtain an estimate of the size  $N$  of the population. The duration of the survey is typically short so that individuals are unlikely to enter or leave the population through births, deaths, immigration, or emigration.

Such capture-recapture surveys have been conducted for decades, and the literature on this subject is enormous (Schwarz and Seber, 1999). Many of the statistical models developed for estimating  $N$  include parameters that specify systematic sources of variation in an individual's rate of capture or detection. In these models, capture probabilities are assumed to differ among sampling occasions or among subsets of the population (often identified through post-stratification) that are thought to have been detected with nearly equal probability (Darroch, 1961; Otis et al., 1978; Pollock, Hines, and Nichols, 1984; Huggins, 1989; Ahlo, 1990;

Plante, Rivest, and Tremblay, 1998). When probabilities of capture differ among members of the population, failure to account for the differences can result in biased estimates of  $N$  (Otis et al., 1978). The potential for bias can be mitigated, to some extent, by careful attention to survey design (Seber, 1982, p. 495) and by partitioning the population (and the data) into groups of individuals that are thought to have similar detection rates (assuming, of course, that information in the sample allows these groups to be identified). However, even these measures may not adequately eliminate the potential for bias owing to uncontrollable and unobservable variation in the detection of individuals. For example, differences in behavior, coloration, or song of individual animals can dramatically affect their detectability. Therefore, models that account for individual variation in capture rates are needed for accurate estimation of  $N$ .

In the capture-recapture literature,  $M_h$  denotes models in which heterogeneity ( $h$ ) among individuals is assumed to be the primary source of variation in capture rates (Otis et al., 1978). Estimators of  $N$  appropriate for  $M_h$  were developed first by Burnham (1972), who considered a beta-binomial mixture model and a nonparametric, jackknife estimator

(Burnham and Overton, 1978, 1979). Subsequent models of heterogeneity have included alternative methods of estimation (Pollock and Otto, 1983; Smith and van Belle, 1984; Chao, 1989; Lloyd and Yip, 1991; Yip, 1991; Chao, Lee, and Jeng, 1992; Lloyd, 1992; Bunge and Fitzpatrick, 1993; Lee and Chao, 1994) and likelihood-based methods that parameterize capture rates in terms of differences among individuals and differences among observable covariates (Mingoti and Meeden, 1992; Darroch et al., 1993; Agresti, 1994; Norris and Pollock, 1996; Coull and Agresti, 1999; Fienberg, Johnson, and Junker, 1999; Pledger, 2000).

In these models inferences about  $N$  generally depend on an assumed pattern of variation in individual capture rates. In latent-class (or finite-mixture) models (Agresti, 1994; Norris and Pollock, 1996; Coull and Agresti, 1999; Pledger, 2000), capture rates are assumed to differ among two or more classes of individuals and to be homogeneous within each class. Because class identity is not observed, the proportion of individuals in each class must be estimated. Latent-class models seem most appropriate for populations that actually contain groups of individuals that are thought to have different rates of capture (e.g., juveniles and adults). However, even in these populations, additional variation in capture rates may exist among individuals within each group. In such cases, application of latent-class models underestimates the actual heterogeneity in capture rates and may yield unsatisfactory inferences about  $N$  (Coull and Agresti, 1999). In these populations, it seems more natural to consider models that assume a different rate of capture for each individual. Such models specify individual variation in capture rates hierarchically using continuous distributions, such as a beta distribution of latent capture probabilities (Burnham, 1972) or a normal distribution of their logits (Coull and Agresti, 1999; Fienberg et al., 1999). A practical advantage of continuous mixtures is that the number of model parameters needed to specify heterogeneity in individual rates of capture does not increase with  $N$ , thus helping to ensure that model parameters are well identified.

An interesting aspect of the  $N$ -estimation problem is that conventional goodness-of-fit statistics, such as deviance, cannot be relied on for selecting a model that produces valid inferences about  $N$ . For example, Coull, and Agresti (1999) showed that if capture probabilities are relatively low or vary greatly among individuals, the logistic-normal and latent-class models both may fit the observed data reasonably well, but provide substantially different inferences about  $N$ . The reason, of course, is that model-based estimates of  $N$  actually correspond to extrapolations for the capture histories of unobserved individuals (Fienberg, 1972), so it is not surprising that the extrapolated values are sensitive to model structure. These results suggest that prior knowledge of potential sources of heterogeneity in individual rates of capture and data-based criteria must *both* be considered when selecting models for estimating  $N$ .

There are particular  $N$ -estimation problems in which capture rates are expected to vary greatly among individuals. One problem is the estimation of species richness, the number of distinct species in a natural community of animals (Bunge and Fitzpatrick, 1993; Boulinier et al., 1998). Species iden-

tity is a natural (taxonomic) mark that may be observed in repeated samplings of a community. Natural communities of animals generally include common species that are easily detected, rare species that are difficult to find, and an assortment of species at intermediate levels of abundance whose detectabilities vary greatly. Behavioral differences among species also contribute to their variable rates of detection. Therefore, continuous mixtures seem more appropriate than finite mixtures for approximating the differences in detectability among species.

Another situation in which capture rates are expected to vary among individuals occurs in live-trapping surveys of small mammals. In these surveys traps typically are deployed using an evenly spaced grid of locations that encompasses the range of the population (Otis et al., 1978). Animals with larger foraging areas (or territories) potentially are exposed to more traps than animals with smaller foraging areas. However, differential exposure of animals can occur also if the sampling grid fails to include the entire population of animals because animals near grid boundaries are less prone to capture. Thus, a differential exposure to traps can produce heterogeneity in capture rates of individual animals, even if behavioral differences, such as trap-attraction or trap-avoidance, are absent (Seber, 1982, p. 166).

Species-richness and capture-recapture problems provide ample motivation for the development and application of continuous-mixture models. It is also desirable to have several classes of these models since estimates of  $N$  are sensitive to model structure, as noted earlier. Surprisingly, only two classes of continuous mixtures have been proposed: beta-binomial (Burnham, 1972) and logistic-normal models (Coull and Agresti, 1999; Fienberg et al., 1999). The beta-binomial model is not widely used. Burnham (1972) fitted this model to simulated and actual data sets and concluded that maximum likelihood estimates (MLEs) of  $N$  "will generally be unsatisfactory" for the range of capture probabilities and population sizes likely to be encountered in surveys of natural animal populations. Unacceptably low precision in MLEs of  $N$  and, in some cases, weak identifiability of model parameters apparently formed the basis of his conclusion and motivated development of a jackknife estimator of  $N$  (Burnham and Overton, 1978, 1979).

Our experience with the beta-binomial model suggests that Burnham's 1972 conclusions are overly pessimistic. We show here that the beta-binomial model can provide sensible estimates of  $N$  in a wide range of situations. In particular, we demonstrate the utility of this model for estimating abundance in capture-recapture surveys and for estimating species richness in surveys of breeding birds. We also compare inferences based on continuous and finite mixtures by fitting logistic-normal and latent-class models.

In Section 2, we develop a parameterization of the beta-binomial model that improves the identifiability of its parameters. Computational considerations associated with maximization of the likelihood function and estimation of confidence intervals for  $N$  also are described. In Section 3, beta-binomial, logistic-normal, and latent-class models are fitted to counts of snowshoe hare recaptures and breeding bird species. Simulation experiments are used to illustrate

the difficulty in selecting among classes of models for computing accurate estimates of  $N$ . In Section 4 simulation experiments illustrate the effects of individual variation in capture rates on beta-binomial estimates of  $N$  in populations of various sizes. Section 5 contains our conclusions and a description of how continuous-mixture models might be extended for use in spatially or temporally replicated surveys of species richness.

## 2. A Beta-Binomial Mixture for Heterogeneous Rates of Capture

We assume that a closed population of unknown size  $N$  is observed at  $T$  distinct sampling occasions and that  $n$  different individuals in the population are encountered. If we let  $x_i \in \{0, 1, \dots, T\}$  denote the number of occasions in which the  $i$ th individual is observed (where  $i = 1, \dots, N$ ), we note that  $x_i \in \{1, \dots, T\}$  for the  $n$  individuals observed in the sample and  $x_i = 0$  for the  $N - n$  remaining, unobserved individuals. Without loss of generality, hereafter we refer to the act of observing or detecting an individual as a “capture” and assume that each individual has some nonzero probability of capture at each sampling occasion.

We make two assumptions regarding the  $N$  individuals: (1) the  $\{x_i\}$  are exchangeable in the sense that captures of one individual are independent of those of another individual and (2)  $x_i | T, \alpha, \beta \stackrel{iid}{\sim} \text{BetaBinomial}(T, \alpha, \beta)$ . The second assumption implies that capture probabilities vary among individuals and that this variation can be modeled in terms of the random differences expected under a  $\text{Beta}(\alpha, \beta)$  distribution. Based on these assumptions, the contribution of the observed data  $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$  to the likelihood is

$$f(\mathbf{x} | T, \alpha, \beta) = \prod_{i=1}^n \binom{T}{x_i} \frac{\Gamma(\alpha + \beta) \Gamma(x_i + \alpha) \Gamma(T - x_i + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(T + \alpha + \beta)}. \quad (1)$$

If we let  $n_x = \sum_{i=1}^n I(x_i = x)$  denote the number of individuals observed in  $x$  of the  $T$  sampling occasions, we can express (1) in terms of a vector of sufficient statistics  $\mathbf{n} = (n_1, \dots, n_T)^t$  as follows:

$$g(\mathbf{n} | T, \alpha, \beta) = \prod_{x=1}^T \left\{ \binom{T}{x} \frac{\Gamma(\alpha + \beta) \Gamma(x + \alpha) \Gamma(T - x + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(T + \alpha + \beta)} \right\}^{n_x}. \quad (2)$$

To estimate  $\alpha$ ,  $\beta$ , and the unknown population size  $N$ , we need to form the *complete-data* likelihood by including the contribution of the  $n_0 = N - n$  unobserved individuals. We do so by conditioning on  $N$  (Sanathanan, 1972), which yields a multinomial likelihood function

$$L(N, \alpha, \beta | \mathbf{n}) = \frac{N!}{\prod_{x=0}^T n_x!} \times \prod_{x=0}^T \left\{ \binom{T}{x} \frac{\Gamma(\alpha + \beta) \Gamma(x + \alpha) \Gamma(T - x + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(T + \alpha + \beta)} \right\}^{n_x} \quad (3)$$

$$= \frac{N!}{(N - n)! \prod_{x=1}^T n_x!} \times \left\{ \frac{\Gamma(\alpha + \beta) \Gamma(\alpha) \Gamma(T + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(T + \alpha + \beta)} \right\}^{N-n} g(\mathbf{n} | T, \alpha, \beta). \quad (4)$$

It is easily shown that (4) may be expressed in the following simpler form

$$L(N, \alpha, \beta | \mathbf{n}) = \frac{cN!}{(N - n)!} \left\{ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(T + \alpha + \beta)} \right\}^N \times \{\Gamma(\alpha) \Gamma(T + \beta)\}^{N-n} \times \prod_{x=1}^T \{\Gamma(x + \alpha) \Gamma(T - x + \beta)\}^{n_x}, \quad (5)$$

where  $c = (\prod_{x=1}^T \binom{T}{x}^{n_x}) / (\prod_{x=1}^T n_x!)$  depends on observed data but not on the model parameters.

We use a transformation of the model parameters,  $\mu = \alpha / (\alpha + \beta)$ ,  $\tau = \alpha + \beta$ , and  $n_0 = N - n$ , to facilitate their interpretation and estimation. If  $p$  denotes a latent variable for the probability of capture of an individual in the population, the beta-binomial mixture implies  $E(p) = \mu$  and  $\text{Var}(p) = \mu(1 - \mu) / (\tau + 1)$ . Therefore,  $\mu$  corresponds to the average capture probability among individuals in the population and  $\tau$  corresponds to the similarity (i.e., inverse of heterogeneity) in capture probabilities among individuals. Using this parameterization, the logarithm of the likelihood function in (5) is

$$\begin{aligned} l(\mu, \tau, n_0 | \mathbf{n}) &= h(n_0 + n + 1) - h(n_0 + 1) + (n_0 + n) \\ &\quad \times [h(\tau) - h(\tau\mu) - h\{\tau(1 - \mu)\} - h(T + \tau)] \\ &\quad + n_0[h(\tau\mu) + h\{T + \tau(1 - \mu)\}] \\ &\quad + \sum_{x=1}^T n_x [h(x + \tau\mu) + h\{T - x + \tau(1 - \mu)\}], \end{aligned} \quad (6)$$

where  $h(\cdot) = \log \Gamma(\cdot)$  and the constant  $c$  is omitted.

### 2.1 Maximizing the Likelihood

The MLEs of our model's parameters must be computed by maximizing the log likelihood in (6) numerically. Our experience indicates that the speed and accuracy of these calculations can be improved greatly by using the analytic gradient  $\nabla l$  of the log likelihood (see the Appendix) and a quasi-Newton maximizing algorithm. Additional improvements in optimization can be achieved by implementing a 1–1 transformation of the model's parameters:  $\eta = \text{logit } \mu$ ,  $\theta = \log \tau$ , and  $\phi = \log n_0$ . This transformation places all parameters on a similar scale and ensures that estimates of  $\mu$ ,  $\tau$ , and  $n_0$  are confined to their admissible ranges without requiring constraints during optimization. In addition, this rescaling of the parameters generally improves the condition of Hessian matrices, and therefore improves assessments of parameter identifiability and uncertainty (see below).

Table 1

MLEs and 95% profile-likelihood-based confidence intervals of snowshoe hare abundance  $N$  based on fits of various models of the recaptures.

Model	Structure of capture prob.	$\hat{N}$	95% confidence interval	$G_1^{2a}$	df <sub>1</sub> <sup>b</sup>	$G_2^{2c}$	df <sub>2</sub> <sup>d</sup>	AIC <sup>e</sup>
Beta-binomial	$M_h$	90.8	(73.3, 219.4)	3.97	3	62.0	60	66.0
Logistic-normal	$M_h$	91.7	(74.7, 150.6)	3.72	3	61.7	60	65.7
Logistic-normal	$M_{t+h}$	91.9	(74.7, 153.0)	-	-	52.8	55	66.8
Latent-class	$M_h$	76.7	(69.1, 87.0)	0.49	2	58.5	59	64.5
Latent-class	$M_{t+h}$	76.4	(70.2, 86.4)	-	-	49.7	54	65.7
Latent-class	$M_{t \times h}$	85.2	(74.0, 106.4)	-	-	41.2	49	67.2

$G_1^{2a}$  and  $df_1^b$  denote the deviance and degrees of freedom for likelihood functions based on the sufficient statistic  $\mathbf{n}$ .  $G_2^{2c}$  and  $df_2^d$  denote the deviance and degrees of freedom for likelihood functions based on the number of individuals associated with each of the 63 observable capture histories.

<sup>e</sup>AIC = Akaike's information criterion.

The analytic gradient and Hessian of the log likelihood  $l^*$  defined under the  $(\eta, \theta, \phi)$ -parameterization are easily computed from  $\nabla l$  and  $\mathbf{H}$ , the gradient and Hessian provided in the Appendix. Given the vector of derivatives  $\mathbf{u} = (d\mu/d\eta, d\tau/d\theta, dn_0/d\phi)^t = \{\mu(1-\mu), \tau, n_0\}^t$ , application of the chain rule yields the gradient of the reparameterized log likelihood  $\nabla l^* = \nabla l \odot \mathbf{u}$ , where  $\odot$  indicates the Hadamard product (Magnus and Neudecker, 1999). It is easily verified that a second application of the chain rule yields the Hessian matrix of the reparameterized log likelihood  $l^*$

$$\mathbf{H}^* = \{\mathbf{H} + \text{diag}(\nabla l \odot \mathbf{v})\} \odot \mathbf{u}\mathbf{u}^t,$$

where  $\mathbf{v} = (1 - 2\mu, 1, 1)^t$ .

If a quasi-Newton algorithm converges to the MLE of the model parameters, it provides an approximation of  $\hat{\mathbf{H}}^*$ , the Hessian matrix evaluated at the MLE. In our experience this approximation cannot be relied on for assessments of parameter identifiability in small samples. Instead, we recommend that the analytic Hessian be used to compute the observed information matrix  $-\hat{\mathbf{H}}^*$ . The ratio of largest to smallest eigenvalues of  $-\hat{\mathbf{H}}^*$  then may be used to estimate the condition number of the information matrix and to determine whether the parameters are adequately identified by the data.

Wald-type confidence intervals for  $N$  based on asymptotic normality of the MLEs may be computed in the usual way by inverting  $-\hat{\mathbf{H}}^*$  to obtain  $\widehat{\text{SE}}(\hat{\phi})$  and adding the  $n$  individuals observed in the sample to the endpoints  $\exp\{\hat{\phi} \pm z_{\alpha/2}\widehat{\text{SE}}(\hat{\phi})\}$ . However, profile-likelihood confidence intervals for  $n_0$  (and therefore  $N$ ) may be preferable in small samples if the sampling distribution of the estimator is highly skewed (Cormack, 1992).

### 3. Examples

#### 3.1 Recaptures of Snowshoe Hares

Otis et al. (1978) describe a survey in which snowshoe hares were repeatedly captured using a  $10 \times 10$  grid of baited traps. In six consecutive days of trapping, 68 individual hares were observed, yielding the following vector of recapture frequencies:  $\mathbf{n} = (25, 22, 13, 5, 1, 2)^t$ . Recaptures of snowshoe hares associated with each of the 63 ( $= 2^6 - 1$ ) observable capture histories have been tabulated and analyzed elsewhere (Agresti, 1994; Coull and Agresti, 1999).

Three classes of mixture models (beta-binomial, logistic-normal, and latent-class) were fitted to the hare data and

used to estimate population size  $N$  (Table 1). Models were parameterized to allow capture rates to differ among individuals (or groups of individuals in finite mixtures) and, if possible, among capture occasions. For latent-class models, we report results obtained with only two support points. Models with three or more support points yield redundant estimates of support, insignificant reductions in deviance, or unstable MLEs. Similar findings are reported by Agresti (1994), who fitted quasi-symmetric ( $M_{t+h}$ ) and ordinary ( $M_{t \times h}$ ) latent-class models to these data.

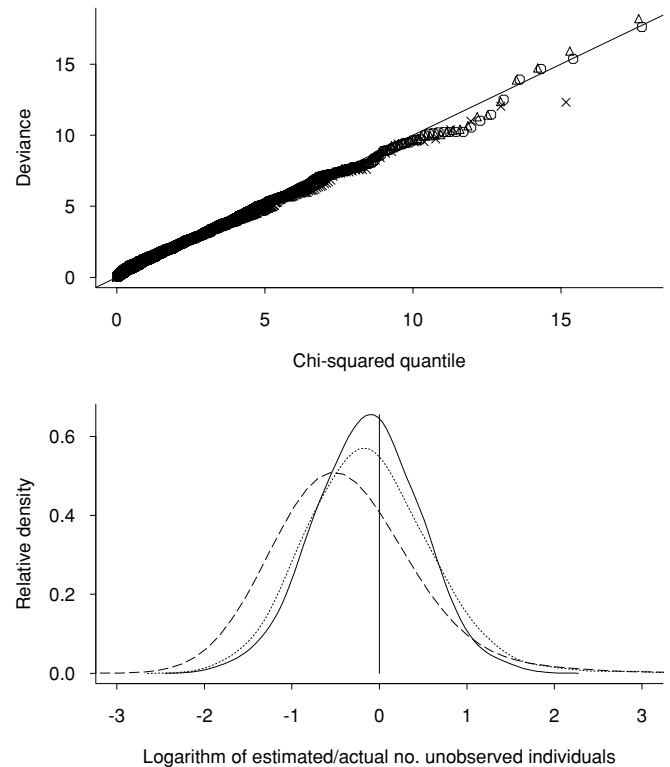
Within each class of mixture models, likelihood-ratio statistics and Akaike's information criterion (AIC) are valid diagnostics for use in model selection. For example, in logistic-normal mixtures, the reduction in deviance achieved by increasing model complexity from  $M_h$  to  $M_{t+h}$  does not appear to be significant ( $p = 0.11, \chi_5^2$ ). AIC likewise favors the simpler model  $M_h$ . Furthermore, the MLEs and confidence intervals of  $N$  are virtually identical in both logistic-normal models. Similar results are obtained by comparing latent-class models with and without time dependence. The simplest model  $M_h$  has the lowest value of AIC and appears to be the parsimonious choice. Thus, comparison of models within each of these classes suggests that heterogeneity among animals appears to be the primary source of variability in capture rates of snowshoe hares.

Likelihood-ratio statistics for comparing models with and without heterogeneity ( $M_h$  vs.  $M_0$ ) do not have simple chi-squared reference distributions because the simpler model ( $M_0$ ) falls on the boundary of the parameter space of the more complex model (Agresti, 2002, p. 525). However, the correct reference distributions are available in some cases involving specific classes of mixture models (Self and Liang, 1987; Pledger, 2000).

Selection among classes of mixture models is more difficult than selection within classes and can be crucial to obtaining valid inferences about  $N$ . This is particularly true in surveys where capture rates are relatively low and vary greatly among individuals, as appears to be the case in the survey of snowshoe hares (Coull and Agresti, 1999). Consider the results of fitting beta-binomial, logistic-normal, and latent-class models that assume differences in the rate of capture among animals but not among capture occasions (i.e.,  $M_h$ ). Each of these models appears to fit the hare data reasonably well based on the asymptotic  $\chi_{df}^2$  distribution of the residual deviance  $G_1^2$

(Table 1). However, the MLEs of  $N$  computed using the beta-binomial and logistic-normal models (90.8 and 91.7, respectively) are considerably greater than the MLE of  $N$  computed using the latent-class model (76.7). The discrepancy between estimates is not surprising because model-based estimates of  $N$  are essentially extrapolations that are sensitive to model structure, as noted in Section 1. In the case of the hare data, this sensitivity is easily demonstrated with simulation experiments. We simulated 1000 data sets for each of the three mixture models by fixing the population size at  $N = 100$  and generating random capture histories for  $T = 6$  occasions according to each model's underlying assumptions. The parameters of each data-generating model were assigned to equal the MLEs obtained by fitting the hare data ( $\hat{\mu} = 0.267$  and  $\hat{\tau} = 6.86$  for the beta binomial,  $\hat{\beta} = -1.20$  and  $\hat{\sigma} = 0.92$  for the logistic normal, and  $\hat{p}_1 = 0.295$ ,  $\hat{p}_2 = 0.973$ ,  $\hat{\pi}_1 = 0.97$ , and  $\hat{\pi}_2 = 0.03$  for the latent class). All three models were fitted to each set of simulated data. To examine the consequences of model misspecification, we compared each model-specific estimate of the number of unobserved individuals in a simulated sample, say  $\hat{n}_0$ , with the actual number of unobserved individuals  $n_0$ . Given the relatively small size of the simulated populations ( $N = 100$ ), the sampling distributions of  $\hat{n}_0$  were fairly skewed; therefore, the discrepancy between estimated and actual numbers of unobserved individuals was evaluated by examining the distribution of  $\log(\hat{n}_0/n_0)$  and using its mean as an estimate of model-specific bias. For example, Figure 1 shows that estimates of  $n_0$  obtained by fitting beta-binomial and logistic-normal models to beta-binomial-generated data are somewhat negatively biased (means =  $-0.11$  and  $-0.14$ , respectively), whereas estimates computed using the latent-class model are considerably more negatively biased (mean =  $-0.36$ ). Furthermore, quantile-quantile plots illustrate that conventional diagnostics for assessing model adequacy (deviance) are unable to detect the effects of model misspecification (Figure 1). A naive reliance on such diagnostics can lead to erroneous inferences about  $n_0$  (and therefore  $N$ ). Qualitatively similar results are obtained by fitting the three models to data generated with the logistic-normal model (Figure 2). Mean biases for the beta-binomial and logistic-normal models are small (0.02 and  $-0.02$ , respectively) compared to the negative bias of the latent-class model ( $-0.34$ ). Different results are obtained by fitting the three models to data generated using the latent-class model (Figure 3). Estimates of  $n_0$  based on beta-binomial and logistic-normal models are positively biased (means = 0.78 and 0.71, respectively), whereas estimates computed using the latent-class model are only slightly positively biased (mean = 0.08). In these simulations, however, there is considerable power to detect the effects of model misspecification using residual deviance. Chi-squared tests for goodness-of-fit (using a 5% significance level) reject the beta-binomial and logistic-normal models in about half of the simulated data sets (57% and 53%, respectively), which is substantially higher than the close-to-nominal rejection rate (6%) of the latent-class model fits.

The analyses of snowshoe hare recaptures illustrate the difficulty of computing valid inferences about  $N$  when little is known about the sources of variability in individual rates of capture. It is tempting to choose classes of models that yield narrower confidence intervals for  $N$ ; however, Coull's



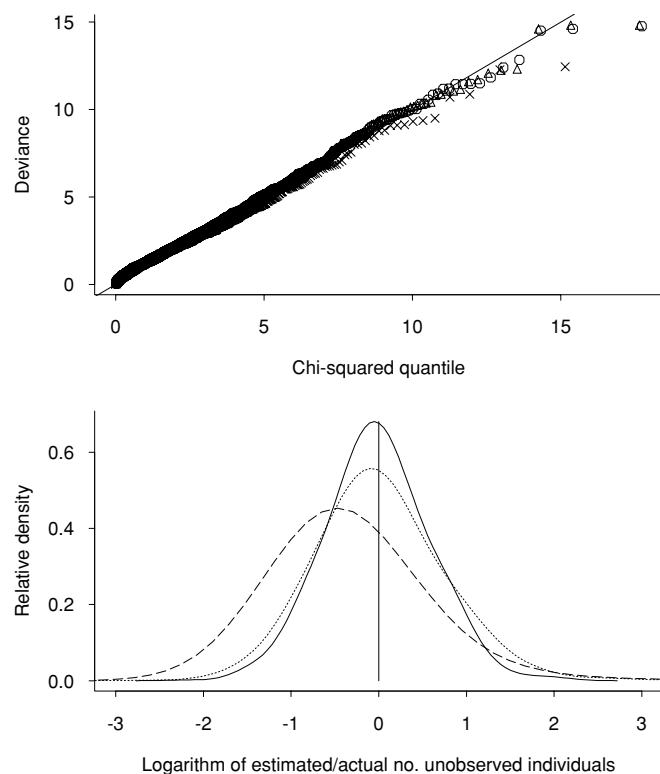
**Figure 1.** Summary of models fitted to 1000 data sets simulated using a beta-binomial model. *Lower panel:* Discrepancy between the actual number of unobserved individuals and the estimated number based on fitting beta-binomial (dotted), logistic-normal (solid), and latent-class (dashed) models. *Upper panel:* Quantile-quantile plots with 45° line superposed to compare deviances with their  $\chi^2_{df}$  reference distributions. Symbols indicate deviances of beta-binomial ( $\Delta$ ), logistic-normal ( $\circ$ ), and latent-class ( $\times$ ) models.

(1997) extensive simulation studies (summarized by Coull and Agresti (1999)) clearly reveal that this behavior can produce erroneous inferences. In the absence of additional information about possible sources of heterogeneity in the capture of individual hares, it is difficult to determine which estimate of abundance is the most accurate.

### 3.2 Species of Breeding Birds

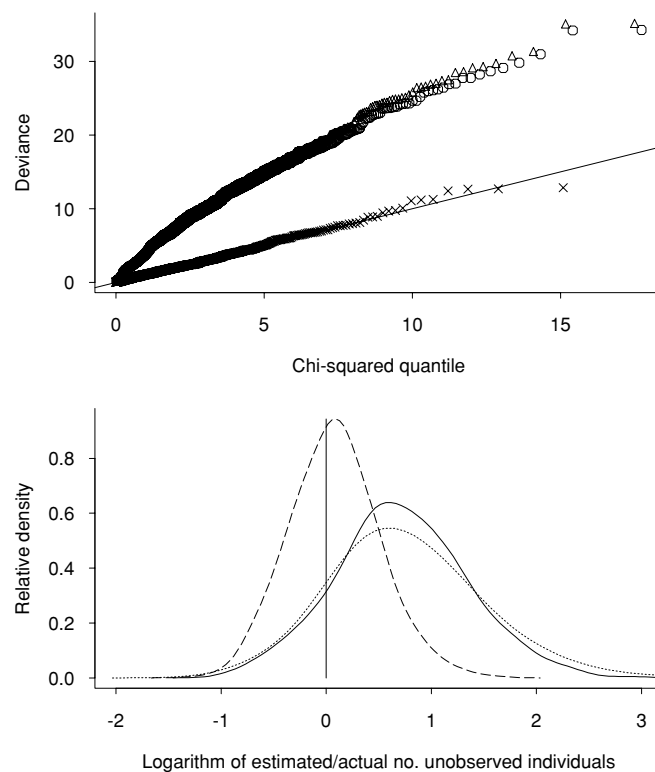
The problem of individual heterogeneity is particularly acute in the estimation of species richness because large differences in detectability of species may be induced by differences in behavior (time of activity, habitat preferences, foraging) or abundance, and by differences in the ability of observers to identify species (e.g., variable abilities in discriminating birds by song or call). Consequently, models of individual heterogeneity are not just candidates to be compared against competitors; instead they are preferred *a priori* on scientific grounds (Boulinier et al., 1998).

We illustrate the estimation of species richness using data from the North American Breeding Bird Survey (BBS). The BBS was established to provide information about changes in populations and communities of land birds at a continental



**Figure 2.** Summary of models fitted to 1000 data sets simulated using a logistic-normal model. *Lower panel:* Discrepancy between the actual number of unobserved individuals and the estimated number based on fitting beta-binomial (dotted), logistic-normal (solid), and latent-class (dashed) models. *Upper panel:* Quantile-quantile plots with 45° line superposed to compare deviances with their  $\chi^2_{df}$  reference distributions. Symbols indicate deviances of beta-binomial ( $\Delta$ ), logistic-normal ( $\circ$ ), and latent-class ( $\times$ ) models.

scale (Robbins et al., 1989; Sauer, Pendleton, and Peterjohn, 1996) and includes more than 4000 39.4-km survey routes (primary sample units) throughout North America. Details about the survey's design and implementation are described elsewhere (Robbins, Bystrak, and Geissler, 1986; Boulinier et al., 1998). For our purposes, it is sufficient to note that along each survey route an observer stops at 50 equidistant locations and records the species identity and number of all birds heard or seen within a 3-minute period. Although the BBS started in 1966, location-specific records have only been maintained since 1997. We fitted three classes of mixed models (beta-binomial, logistic-normal, and latent-class) to five years (1997–2001) of BBS species counts observed along Route 123 in Maryland (located near Davidsonville). In any single year, no species was encountered at more than 36 locations along the route. Table 2 contains the frequencies of species observed at  $\leq 36$  locations. Likelihood-ratio comparisons of models with and without parameters for location-specific rates of detection suggest that differences in detectability among species are the primary sources of heterogeneity; therefore, in Table 3 we report only the results of fitting models for  $M_h$ . For latent-class models, we report results of fitting as many latent classes as



**Figure 3.** Summary of models fitted to 1000 data sets simulated using a latent-class model. *Lower panel:* Discrepancy between the actual number of unobserved individuals and the estimated number based on fitting beta-binomial (dotted), logistic-normal (solid), and latent-class (dashed) models. *Upper panel:* Quantile-quantile plots with 45° line superposed to compare deviances with their  $\chi^2_{df}$  reference distributions. Symbols indicate deviances of beta-binomial ( $\Delta$ ), logistic-normal ( $\circ$ ), and latent-class ( $\times$ ) models.

possible without obtaining redundant points of support, negligible reductions in deviance, or unstable MLEs.

Within each class of mixed models, MLEs of species richness  $N$  are similar in each of the five years; however, the estimates differ considerably among classes (Table 3). The beta-binomial estimates of  $N$  are higher (mean = 104.3) and less precise than those computed for the logistic-normal and latent-class models (means = 81.5 and 78.7, respectively).

How do we determine the class of models that is most likely to provide valid inferences about species richness? Asymptotic distributions of data-based criteria, such as deviance, are unlikely to be helpful owing to the low counts of observed species (Table 2). However, prior information about heterogeneity in detection rates of bird species is helpful in selecting an inferential model. Since 1982, when Route 123 was first included in the BBS, 102 distinct species of birds have been observed (<http://www.mbr-pwrc.usgs.gov/cgi-bin/rtena.pl?46123>). Fourteen of these species have been observed only once in the 20-year period. In contrast, 43 of the 102 species have been detected in every year. The remaining 45 species have been observed at various intermediate levels of detection. Given the apparent wide range of detectabilities, a

**Table 2**

Frequencies of species observed at  $\leq 36$  locations along a single BBS route

No. locations	Year				
	1997	1998	1999	2000	2001
1	14	14	11	14	15
2	10	9	12	7	7
3	6	7	10	5	5
4	3	4	4	5	2
5	3	2	4	5	5
6	6	4	1	4	6
7	2	4	4	3	3
8	3	0	2	4	2
9	2	4	3	1	1
10	1	2	3	1	2
11	1	0	0	2	4
12	1	1	2	1	4
13	0	0	4	1	0
14	3	3	1	3	1
15	2	0	1	3	0
16	0	3	0	1	2
17	3	1	1	1	3
18	1	0	1	0	1
19	1	0	1	2	1
20	1	1	2	0	1
21	0	1	0	0	1
22	0	0	1	0	0
23	1	1	0	1	0
24	2	1	0	0	0
25	0	2	0	1	1
26	0	0	0	1	1
27	1	1	0	0	0
28	0	0	0	1	1
29	0	1	0	0	0
30	0	1	1	0	1
31	0	0	3	1	0
32	0	0	0	0	0
33	0	0	0	0	0
34	0	0	0	0	0
35	0	0	0	1	0
36	0	0	0	0	1
Total	67	67	72	69	71

categorization of species into a finite number of discrete groups (e.g., rare, common, ubiquitous) seems artificial and is unlikely to adequately approximate the actual species-specific differences in detectability. We therefore believe that for these data the latent-class models are scientifically untenable. Continuous mixtures, on the other hand, seem more appropriate for specifying the differences in detection of different bird species. Unfortunately, the estimates of  $N$  computed for beta-binomial and logistic-normal mixtures differ considerably (Table 3), so which of these two models is more likely to yield valid inferences? A naive, but reasonable, estimate of species richness for a particular BBS route is the cumulative number of species seen during that route's entire history of observation. For Route 123, this number equals 102 species, which is closer to the beta-binomial estimates of  $N$  than to the logistic-normal estimates. Although this kind of agreement is reassuring, the confidence intervals computed for  $N$

**Table 3**

MLEs and 95% profile-likelihood-based confidence intervals of species richness  $N$  based on fits of various models to the data in Table 2.

Model	Year	$n^a$	$\hat{N}$	95% confidence interval	$G^{2b}$	df <sup>c</sup>
Beta-binomial	1997	67	100.2	(75.3, 405.6)	26.9	47
Beta-binomial	1998	67	112.9	(78.2, $\infty$ )	37.3	47
Beta-binomial	1999	72	106.0	(81.2, 313.5)	38.6	47
Beta-binomial	2000	69	100.9	(77.1, 329.9)	25.1	47
Beta-binomial	2001	71	101.3	(78.8, 279.5)	32.0	47
Logistic-normal	1997	67	78.0	(70.6, 89.8)	29.0	47
Logistic-normal	1998	67	80.0	(71.4, 97.5)	37.9	47
Logistic-normal	1999	72	85.2	(75.8, 101.0)	38.9	47
Logistic-normal	2000	69	82.0	(73.2, 97.5)	24.1	47
Logistic-normal	2001	71	82.2	(73.8, 97.5)	35.3	47
Latent-class	1997	67	76.1	(68.7, 95.3)	20.0	42
Latent-class	1998	67	74.3	(68.1, 89.6)	30.7	42
Latent-class	1999	72	76.0	(72.1, 83.3)	29.7	42
Latent-class	2000	69	81.3	(70.8, 513.5)	20.4	40
Latent-class	2001	71	85.7	(73.9, 650.5)	28.9	40

<sup>a</sup>The number of species observed in each year's sample is denoted by  $n$ .

$G^{2b}$  and  $df^c$  denote the deviance and degrees of freedom for likelihood functions based on the sufficient statistic  $\mathbf{n}$ .

under the beta-binomial model indicate that considerable uncertainty exists in the estimates of  $N$  computed for each year. Reductions in uncertainty are obviously needed for improved inferences about  $N$ . Given the similarity in annual estimates of species richness, uncertainty might be reduced by considering the species counts observed in different years as a form of temporal replication and by developing a model that borrows strength from information observed in different years. This approach is described more fully in Section 5.2.

#### 4. Effects of Heterogeneous Capture Rates on Beta-Binomial Estimates of $N$

We conducted several simulation experiments to evaluate the performance of beta-binomial MLEs of  $N$  in populations of different numbers of individuals with different levels of heterogeneity in capture rates. In these simulations, the beta-binomial mixture was used to generate random observations of the sufficient statistic  $\mathbf{n}$ , given particular values of the model parameters and  $T$ .

Performance of the MLE obviously depends on the mean probability of capture, the heterogeneity in capture probabilities among individuals, the number of sampling occasions, and  $N$ . Generally, we expect the MLE to perform better as the proportion of the population that is observed in a sample increases. The probability that a single individual detected with probability  $p$  is observed at least once in  $T$  capture occasions is  $1 - (1 - p)^T$ . Therefore, it clear that increasing  $T$  is expected to increase the proportion of the population that is observed, even in situations where individual capture rates  $p$  are relatively small. In many practical problems  $T$  may be the only controllable feature of the survey, so it is important to consider the impact of  $T$  on the MLE's performance.

In the simulation experiments the mean probability of capture was fixed at  $\mu = 0.2$  (a relatively low rate of

**Table 4**  
*Simulated results for MLEs of  $\phi = \log n_0$  and  $\log N$  based on data generated under  $\mu = 0.5$  and  $T = 5$ .*

$N$	$\hat{E}(\phi)$	$\hat{E}(\hat{\phi})$	$\widehat{SD}(\hat{\phi})$	$\hat{E}(\log \hat{N})$	$\hat{Pr}(G^2 > \chi^2_{2,0.95})$	$\hat{E}(G^2)$	% failure
$\sigma_p = 0.078, p_{cap} = 0.961^a$							
50	0.67	0.31	0.77	3.91	0.073	2.52	15.5
100	1.36	1.14	0.48	4.60	0.076	2.37	3.1
150	1.77	1.64	0.39	5.01	0.064	2.29	2.3
250	2.28	2.18	0.31	5.52	0.065	2.20	1.0
500	2.97	2.92	0.23	6.21	0.056	2.10	0.8
2000	4.36	4.35	0.12	7.60	0.059	2.06	0.0
$\sigma_p = 0.151, p_{cap} = 0.937^a$							
50	1.15	0.61	1.08	3.91	0.054	2.26	7.7
100	1.84	1.55	0.59	4.60	0.063	2.13	1.1
150	2.24	2.05	0.46	5.00	0.069	2.15	0.5
250	2.76	2.64	0.35	5.52	0.057	2.14	0.0
500	3.45	3.40	0.24	6.21	0.058	2.05	0.0
2000	4.83	4.83	0.12	7.60	0.049	2.04	0.0
$\sigma_p = 0.224, p_{cap} = 0.893^a$							
50	1.68	1.10	1.11	3.90	0.064	2.21	1.7
100	2.37	2.12	0.63	4.60	0.048	2.02	0.1
150	2.78	2.62	0.49	5.01	0.048	2.00	0.0
250	3.29	3.21	0.36	5.52	0.046	2.02	0.0
500	3.98	3.95	0.26	6.21	0.048	2.01	0.0
2000	5.37	5.36	0.13	7.60	0.054	2.07	0.0
$\sigma_p = 0.289, p_{cap} = 0.833^a$							
50	2.12	1.57	1.14	3.90	0.070	2.34	1.6
100	2.81	2.58	0.73	4.60	0.063	2.23	0.0
150	3.22	3.08	0.58	5.01	0.056	2.14	0.0
250	3.73	3.65	0.44	5.52	0.062	2.08	0.0
500	4.42	4.40	0.30	6.22	0.058	2.09	0.0
2000	5.81	5.81	0.15	7.60	0.046	2.02	0.0
$\sigma_p = 0.353, p_{cap} = 0.754^a$							
50	2.51	1.94	1.87	3.99	0.068	2.34	2.5
100	3.20	2.93	1.08	4.63	0.062	2.12	0.8
150	3.61	3.46	0.89	5.04	0.056	2.14	0.3
250	4.12	4.01	0.65	5.53	0.060	2.09	0.0
500	4.81	4.77	0.44	6.22	0.058	2.05	0.0
2000	6.20	6.20	0.21	7.61	0.044	2.01	0.0

<sup>a</sup> $p_{cap}$  is the expected proportion of the population captured given  $\mu$ ,  $\sigma_p$ ,  $N$ , and  $T$ .

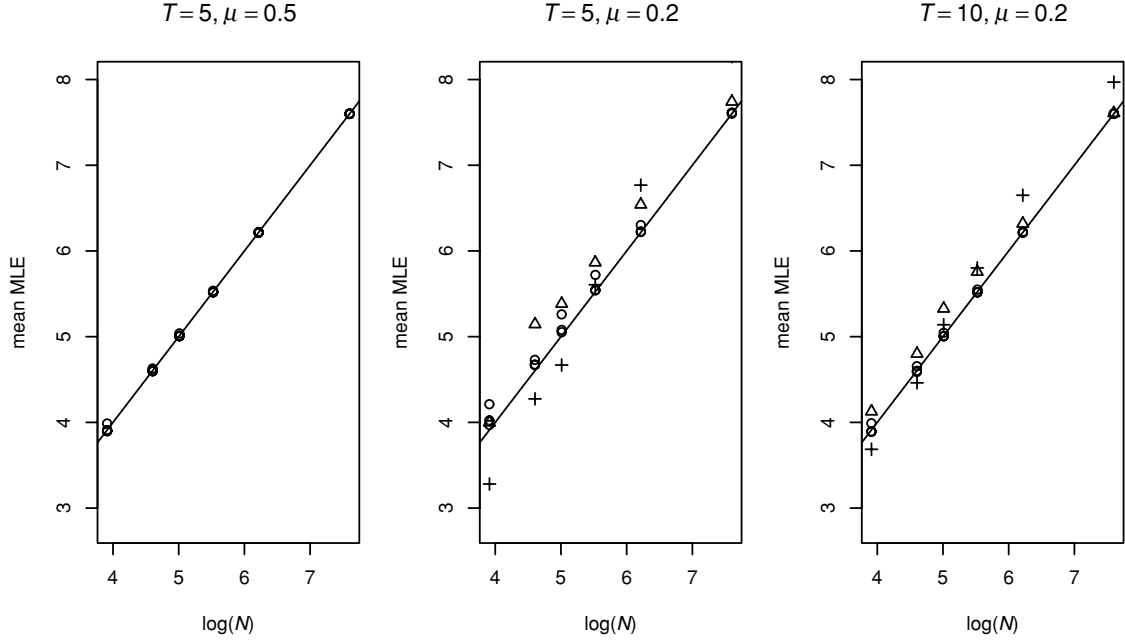
capture) or  $\mu = 0.5$ . These values do not include the entire range of mean capture probabilities that are likely to be realized in practice; however, they adequately illustrate the range of MLE behaviors that were observed in extensive simulation experiments (which are omitted to save space). Several levels of heterogeneity were considered for each value of  $\mu$  by specifying the standard deviation  $\sigma_p$  of the beta distribution of capture probabilities as follows:  $\sigma_p = \{0.3535, 0.2886, 0.2236, 0.1507, 0.0781\}$  (i.e., extreme to little heterogeneity). Under the beta distribution  $\tau = \mu(1 - \mu)/\sigma_p^2 - 1$ ,  $\alpha = \mu\tau$ , and  $\beta = \tau(1 - \mu)$ . Therefore, for  $\mu = 0.5$  the range of values specified for  $\sigma_p$  imply that  $\alpha = \beta = \{0.5, 1, 2, 5, 20\}$ . The case of  $\alpha = \beta = .5$  represents an extreme, and likely pathological, form of heterogeneity wherein most individuals of the population have either very low or very high capture probabilities. In contrast, the case of  $\alpha = \beta = 20$  represents almost no heterogeneity. In simulation experiments where  $\mu$  was fixed at 0.2, the range of values selected for  $\sigma_p$  implies that  $\alpha$  and  $\beta$  were

fixed as follows:  $\alpha = \{0.056, 0.184, 0.440, 1.208, 5.048\}$  and  $\beta = \{0.224, 0.736, 1.760, 4.830, 20.192\}$ .

Simulation experiments with  $\mu = 0.2$  included  $T = 5$  or  $T = 10$  sampling occasions, whereas experiments with  $\mu = 0.5$  included only  $T = 5$  occasions. We did not consider  $T = 10$  in the latter case because nearly all individuals in the population would have been captured, given the specified range of heterogeneity. The simulated populations included  $N = 50, 100, 150, 250, 500$ , or 2000 individuals. Although  $N = 500$  and  $N = 2000$  may be unrealistic in practice, these large populations were included to illustrate asymptotic behavior of the MLE.

For each combination of  $N$ ,  $T$ ,  $\mu$ , and  $\sigma_p$  we summarized the results of 1000 simulated data sets in which our optimization algorithm had successfully converged to the MLE of the model parameters. We also summarized the frequency of simulations where lack of convergence or weak identifiability of model parameters prevented computation of MLEs. The usual deviance statistic  $G^2$  for a multinomial likelihood based





**Figure 4.** Simulated mean of MLEs of  $\log N$  vs.  $\log N$  with  $45^\circ$  line superposed. Symbols indicate different levels of heterogeneity  $\sigma_p$ : 0.353 (+), 0.289 ( $\Delta$ ), and 0.224, 0.151, 0.078 (o).

on  $\mathbf{n}$  was computed for each of the 1000 simulated data sets. The asymptotic distribution of  $G^2$  is  $\chi^2_{T-3}$ ; therefore, in sufficiently large populations the mean of the  $G^2$  values should equal  $T - 3$ , and the proportion of  $G^2$  values that exceed a critical value of  $\chi^2_{T-3, 0.95}$  should equal 0.05.

#### 4.1. Simulation Results

Our simulations revealed that sampling distributions of the MLEs of  $N$  and  $n_0$  can be highly skewed, particularly when  $\mu = 0.2$  or  $N = 50$ . In the presence of skewness, interpretation of moments of these distributions becomes problematic. Interestingly, when MLEs are computed for data sets simulated with  $\mu = 0.5$ , the skewness in sampling distributions is either greatly reduced or absent. For consistency of presentation we summarize the MLEs of  $N$  and  $n_0$  on the logarithmic scale, which also provides a superior scaling ( $\phi = \log n_0$ ) for maximizing the likelihood function (see Section 2.1).

In simulations with  $\mu = 0.5$  and  $T = 5$ , the MLE of  $\phi$  is biased for small  $N$ , but the bias decreases rapidly as  $N$  increases and as heterogeneity decreases (Table 4). There is effectively no bias in  $\log N$ , however, owing to the high proportion of individuals that are expected to be captured (Figure 4). In simulations based on  $\mu = 0.2$  and  $T = 5$ , the MLEs of  $\log N$  are considerably biased at the two highest levels of heterogeneity, although the bias is greatly diminished at  $N = 2000$  (Figure 4). The MLEs of  $\log N$  are reasonably unbiased for  $N \geq 150$  in the three least heterogeneous cases (Table 5). In simulations where  $\mu = 0.2$  and  $T = 10$ , the bias in  $\log N$  is small in all but the most heterogeneous cases (Figure 4 and Table 6).

Reductions in bias with increases in  $N$  illustrate that the MLEs of the beta-binomial model are consistent. However, it is important to remember that maximum likelihood estima-

tion is only an asymptotic procedure, and we cannot be overly optimistic about its capabilities. For example, the simulation results indicate that the uncertainty in MLEs of  $\phi$  (expressed as  $\widehat{SD}(\hat{\phi})$  in Tables 4–6) can be large relative to  $\widehat{E}(\hat{\phi})$  if population size is small or if capture probabilities are extremely heterogeneous. In these cases, there is only limited information in the data, and it is unrealistic to expect high precision in estimates of  $N$  or narrow confidence intervals for  $N$ . Comparison of results in Tables 5 and 6 indicates that an increase in  $T$ , when feasible, can reduce the level of uncertainty by effectively increasing the proportion of the population observed in the sample.

The operating characteristics of the deviance statistic  $G^2$  indicate its *potential* utility in assessments of model adequacy. In sufficiently large populations, the simulated distributions of  $G^2$  appear to have  $\chi^2_{T-3}$  reference distributions. Unfortunately,  $G^2$  alone cannot be relied on for selecting a model that produces valid inferences about  $N$ , as shown in Section 3.1.

#### 4.2 Estimation Problems

In some of the simulated data sets, lack of convergence or weak identifiability of model parameters prevented computation of MLEs. The relative occurrence of these estimation problems is indicated in Tables 4–6 as % failure (= number of failures  $\times 100 /$  (1000 successes + number of failures)). Generally speaking, estimation problems appeared to be more prevalent in situations that led to greater bias in MLEs of  $\log N$ .

We detected two different kinds of estimation problems. The first occurs when convergence to the MLE cannot be achieved within a finite number of iterations because  $\tau$  tends to increase without bound. This kind of problem typically occurs in situations where capture rates are nearly identical for most individuals in the sample. In practice,

**Table 5**  
*Simulated results for MLEs of  $\phi = \log n_0$  and  $\log N$  based on data generated under  $\mu = 0.2$  and  $T = 5$ .*

$N$	$\hat{E}(\phi)$	$\hat{E}(\hat{\phi})$	$\widehat{SD}(\hat{\phi})$	$\hat{E}(\log \hat{N})$	$\hat{Pr}(G^2 > \chi^2_{2,0.95})$	$\hat{E}(G^2)$	% failure
$\sigma_p = 0.078, p_{cap} = 0.642^a$							
50	2.88	2.87	1.48	4.03	0.024	1.83	6.8
100	3.58	3.60	0.93	4.67	0.033	1.92	4.8
150	3.98	3.99	0.59	5.05	0.041	2.00	2.3
250	4.49	4.50	0.45	5.55	0.057	2.04	1.5
500	5.19	5.17	0.32	6.22	0.043	2.18	0.8
2000	6.57	6.57	0.17	7.60	0.055	2.10	0.0
$\sigma_p = 0.151, p_{cap} = 0.571^a$							
50	3.07	2.68	1.94	3.97	0.047	2.23	5.4
100	3.76	3.58	1.51	4.67	0.050	2.25	2.9
150	4.16	4.09	1.19	5.08	0.061	2.30	1.5
250	4.68	4.61	0.64	5.54	0.058	2.24	0.4
500	5.37	5.34	0.48	6.23	0.059	2.15	0.0
2000	6.75	6.75	0.22	7.61	0.047	1.91	0.0
$\sigma_p = 0.224, p_{cap} = 0.475^a$							
50	3.27	2.76	3.50	4.21	0.078	2.54	6.4
100	3.96	3.67	2.36	4.73	0.067	2.29	5.7
150	4.37	4.37	2.37	5.26	0.051	2.20	4.4
250	4.88	4.94	1.89	5.72	0.049	2.09	3.0
500	5.57	5.59	1.16	6.30	0.050	2.00	0.8
2000	6.96	6.96	0.33	7.62	0.045	1.96	0.0
$\sigma_p = 0.289, p_{cap} = 0.378^a$							
50	3.44	2.38	3.65	4.00	0.094	2.64	12.9
100	4.13	4.05	3.98	5.14	0.074	2.36	7.7
150	4.54	4.47	3.38	5.38	0.079	2.33	9.5
250	5.05	5.10	2.92	5.87	0.069	2.19	8.8
500	5.74	5.92	2.41	6.54	0.044	2.10	6.4
2000	7.13	7.25	1.18	7.74	0.058	1.97	1.6
$\sigma_p = 0.353, p_{cap} = 0.276^a$							
50	3.59	1.25	2.89	3.28	0.176	3.87	38.1
100	4.28	2.48	3.61	4.27	0.124	3.13	11.0
150	4.69	3.16	3.22	4.67	0.090	2.74	8.3
250	5.20	4.41	3.79	5.61	0.076	2.50	9.4
500	5.89	5.93	4.07	6.77	0.062	2.32	13.8
2000	7.28	7.74	3.47	8.27	0.059	2.17	13.2

<sup>a</sup> $p_{cap}$  is the expected proportion of the population captured given  $\mu$ ,  $\sigma_p$ ,  $N$ , and  $T$ .

diagnosis of this problem would typically lead us to consider a simpler model (e.g., one with constant probability of capture for all individuals) because only two parameters ( $\mu$  and  $n_0$ ) of the beta-binomial model are well identified by the data. The second kind of estimation problem occurs when optimization converges to an MLE, but the observed information matrix  $-\hat{H}^*$  is nearly singular. This problem typically occurs when the expected number of individuals observed in the sample is much less than  $N$  owing to low capture rates of most individuals or to an inadequate number of sampling occasions. In this case, the information in the data is not sufficient to permit inference for all of the model's parameters, and ridges or plateaus are evident in the likelihood surface. Ridges in the likelihood usually correspond to highly negative posterior correlations between  $\eta = \text{logit } \mu$  and  $\phi = \log n_0$  and may produce MLEs near the boundary of the parameter space (e.g.,  $\mu \rightarrow 0$  and  $n_0 \rightarrow \infty$ ). Problems of stability of MLEs and convergence near the boundary of the parameter space are not unique to the beta-binomial model. Coull and Agresti

(1999) note similar difficulties with the logistic-normal model fitted to a data set containing limited information. Norris and Pollock (1996) impose an arbitrary upper bound on  $N$  and a lower bound on capture probabilities to trap convergence problems in latent-class models.

## 5. Discussion

We have shown that the beta-binomial model can be used to make sensible inferences about  $N$  when heterogeneity in rates of capture or detection exists and is the primary phenomenon thought to have produced the variation in the data. The simulation results in Section 4 suggest that the beta-binomial estimator of  $N$  is asymptotically unbiased over a wide range of capture rates, including those expected of animals encountered in capture-recapture or species-richness studies. However, our simulations also reveal 2 different kinds of estimation problems. The first can occur when heterogeneity in capture rates is absent or negligible. This is a degenerate case of the model ( $\tau \rightarrow \infty$ ) where maximization of the likelihood can be

**Table 6**  
*Simulated results for MLEs of  $\phi = \log n_0$  and  $\log N$  based on data generated under  $\mu = 0.2$  and  $T = 10$ .*

$N$	$\hat{E}(\phi)$	$\hat{E}(\hat{\phi})$	$\widehat{SD}(\hat{\phi})$	$\hat{E}(\log \hat{N})$	$\hat{Pr}(G^2 > \chi_{7,0.95}^2)$	$\hat{E}(G^2)$	% failure
$\sigma_p = 0.078, p_{cap} = 0.848^a$							
50	2.03	1.75	0.65	3.90	0.007	4.33	1.5
100	2.72	2.59	0.48	4.60	0.010	4.66	0.7
150	3.13	3.03	0.39	5.01	0.008	4.83	0.2
250	3.64	3.58	0.31	5.52	0.021	5.20	0.1
500	4.33	4.30	0.21	6.21	0.020	5.37	0.0
2000	5.72	5.71	0.11	7.60	0.029	6.01	0.0
$\sigma_p = 0.151, p_{cap} = 0.738^a$							
50	2.57	2.17	0.99	3.89	0.027	6.18	0.2
100	3.27	3.07	0.61	4.59	0.037	6.52	0.0
150	3.67	3.56	0.49	5.00	0.039	6.68	0.0
250	4.18	4.12	0.37	5.52	0.045	6.93	0.0
500	4.88	4.85	0.26	6.21	0.045	7.09	0.0
2000	6.26	6.26	0.13	7.60	0.048	6.90	0.0
$\sigma_p = 0.224, p_{cap} = 0.591^a$							
50	3.02	2.56	2.11	3.99	0.067	8.02	2.5
100	3.71	3.53	1.31	4.66	0.063	7.64	1.1
150	4.12	4.02	0.91	5.05	0.064	7.58	0.5
250	4.63	4.59	0.72	5.55	0.065	7.47	0.4
500	5.32	5.29	0.41	6.22	0.058	7.21	0.0
2000	6.71	6.70	0.20	7.60	0.039	6.96	0.0
$\sigma_p = 0.289, p_{cap} = 0.450^a$							
50	3.32	2.69	3.29	4.13	0.083	8.29	9.2
100	4.01	3.76	2.68	4.80	0.092	7.75	7.1
150	4.41	4.42	2.80	5.33	0.065	7.52	6.5
250	4.92	4.98	2.16	5.76	0.053	7.27	4.5
500	5.62	5.63	1.32	6.32	0.043	7.14	1.5
2000	7.00	6.99	0.36	7.61	0.055	7.09	0.0
$\sigma_p = 0.353, p_{cap} = 0.305^a$							
50	3.55	1.82	3.58	3.69	0.065	9.05	27.0
100	4.24	2.96	3.51	4.46	0.092	8.62	10.2
150	4.65	3.91	3.68	5.14	0.090	8.23	9.2
250	5.16	4.86	3.54	5.80	0.084	7.86	11.0
500	5.85	5.92	3.52	6.65	0.069	7.49	11.1
2000	7.24	7.48	2.54	7.97	0.047	7.13	8.3

<sup>a</sup> $p_{cap}$  is the expected proportion of the population captured given  $\mu$ ,  $\sigma_p$ ,  $N$ , and  $T$ .

difficult (especially in small populations) because only two of the model's parameters ( $\mu$  and  $n_0$ ) are well identified. The obvious solution to this problem is to fit  $M_0$  (Otis et al., 1978), a model which assumes that all individuals are captured with an equal, constant probability. The second estimation problem can occur when the expected number of individuals observed in a sample is small relative to the size of the population. This situation may be induced by low rates of capture of most individuals, by an inadequate number of sampling occasions, or by some combination of these two. Whatever the cause, the beta-binomial model's parameters may be weakly identified in this situation owing to ridges or plateaus in the likelihood surface. We have found that optimizations that utilize the  $(\eta, \theta, \phi)$ -parameterization developed in Section 2.1 and the analytic gradient of the log likelihood are much less prone to this kind of estimation problem than optimizations based on the  $(\mu, \tau, n_0)$ -parameterization. In addition, evaluation of the analytic, observed information matrix provides a reliable diagnostic (the condition number) for judg-

ing whether the model's parameters are adequately identified by the data.

### 5.1 Continuous vs. Finite Mixtures

In the simulation studies in Section 3.1, we showed that inferences about  $N$  can be quite sensitive to the pattern of individual heterogeneity assumed in a mixture model. This is especially true in populations of animals with widely varying rates of capture. For example, when continuous distributions (beta or normal) were used to simulate highly variable capture rates, estimates of  $N$  obtained by fitting the latent-class model were considerably more biased than those obtained by fitting the beta-binomial or logistic-normal models (Figures 1 and 2). The reason, of course, is that the discrete distribution specified in the finite-mixture does not adequately approximate the latent distribution of capture rates. Pledger (2000) asserts that "the two-point support [in finite-mixture models] is often able to provide enough variability to model data from a distribution with more support points, even infinitely

many.” Our simulation results cast doubt on the generality of this assertion.

Adding support points potentially allows finite mixtures to better approximate a highly variable, latent distribution of capture rates; however, this is not always feasible. In Model  $M_h$ , for example, the number of support points cannot exceed  $T/2$  and remain well identified (Pledger, 2000). In addition, problems of identifiability are not always resolved by increasing the number of capture occasions. As heterogeneity in capture rates increases, finite-mixture models must “spend” the information in the data to approximate the latent distribution of capture rates. Failing to do so (e.g., by using an insufficient number of support points) may produce biased estimates of  $N$  or interval estimates of  $N$  that are too narrow and have poor coverage (Coull, 1997; Coull and Agresti, 1999). The beta-binomial and logistic-normal models obviously have an advantage in this situation. Such continuous mixtures are able to specify large variation in capture rates without increasing the number of parameters to be estimated and, in so doing, provide more accurate, if less precise, estimates of  $N$ .

In the foregoing paragraphs we do not mean to imply that finite mixtures lack utility. In populations of individuals with modest differences in capture rates, finite mixtures may be adequate. The difficulty lies in the fact that because capture rates are unobservable, one usually cannot assess the level of individual heterogeneity prior to model fitting. To make matters worse, conventional diagnostics for selecting among finite- and continuous-mixture models (deviance) are not always reliable with respect to  $N$ -estimation. We therefore stress the importance of using prior knowledge of potential sources of heterogeneity in capture rates to select classes of models for computing inferences about  $N$ . Unfortunately, unlike our example of breeding bird species (Section 3.2), such prior knowledge is not always available.

### 5.2 Extension of Mixture Models for Species Richness

Estimation of species richness is an important application of mixture models because probabilities of detection can vary greatly among different species of animals, as described in Section 3.2. In surveys of species richness, the number of repeated observations  $T$  may correspond to different times or locations, and it is entirely feasible to increase  $T$  by an arbitrarily large amount and still be able to assume that the sample includes a closed population. Changes in species richness due to evolution of new species or to extinction of existing species are unlikely to occur during the time required to complete a survey.

Increasing  $T$  as a matter of sampling design can dramatically improve estimates of species richness  $N$  because the probability that a species detected with probability  $p$  is encountered in the survey ( $= 1 - (1 - p)^T$ ) approaches unity as  $T$  becomes large. Such improvements in estimates of species richness will be most conspicuous in animal communities dominated by species that are difficult to detect (i.e., low  $p$ ), as illustrated in our simulation studies (Figure 4).

In Section 3.2, we showed that the beta-binomial model provides sensible estimates of the number of bird species along a single route of the BBS. However, uncertainty in the estimates is fairly high because the counts of bird species observed along each BBS route suggest considerable heterogeneity in probabilities of detection (Table 2). In the absence of addi-

tional route- or location-specific covariates that might help to reduce the heterogeneity, it is logical to consider whether improvements in  $N$ -estimation are possible by modeling temporal variation in species counts observed along a single BBS route or by modeling spatial variation in species counts observed along different BBS routes. The idea behind such models is to borrow strength from information observed in different years or in different locations to provide better inferences about  $N$ . Regardless of whether the beta-binomial or logistic-normal models can be extended to specify temporal or spatial dependence, further development of models with some form of hierarchical structure is clearly needed.

### RÉSUMÉ

Nous développons une paramétrisation du mélange de lois bêta et binomiale qui fournit un cadre inférentiel approprié pour estimer l'effectif d'une population fermée lorsque les probabilités de capture ou de détection varient entre individus. Nous ajustons trois classes de modèles de mélange (bêta – binomiale, logistique – normale, et modèle à classes latentes), d'une part à des données de recapture de lièvres d'Amérique pour en estimer l'abondance, et d'autre part à des comptages d'espèces d'oiseaux pour estimer la richesse spécifique. Dans les deux jeux de données les taux de détection varient davantage entre individus (lièvres ou espèces) qu'entre occasions ou lieux de capture. Les estimateurs de l'effectif de la population et de la richesse spécifique sont sensibles aux hypothèses faites sur la distribution sous-jacente des probabilités individuelles de détection. Nous montrons par simulations que les outils habituels d'évaluation de la qualité du modèle, comme la déviance, ne sont pas fiables pour choisir entre des modèles de mélanges celui qui fournit une inférence correcte de l'effectif de la population. Lorsqu'elles existent, des connaissances *a priori* sur les causes d'hétérogénéité des taux de détection doivent être utilisées pour aider à choisir quelle classe de modèles de mélanges utiliser pour l'inférence.

### REFERENCES

- Agresti, A. (1994). Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics* **50**, 494–500.
- Agresti, A. (2002). *Categorical Data Analysis*, 2nd edition. Hoboken, New Jersey: Wiley.
- Ahlo, J. M. (1990). Logistic regression in capture-recapture models. *Biometrics* **46**, 623–635.
- Boulinier, T., Nichols, J. D., Sauer, J. R., Hines, J. E., and Pollock, K. H. (1998). Estimating species richness: The importance of heterogeneity in species detectability. *Ecology* **79**, 1018–1028.
- Bunge, J. and Fitzpatrick, M. (1993). Estimating the number of species: A review. *Journal of the American Statistical Association* **88**, 364–373.
- Burnham, K. P. (1972). Estimation of population size in multiple capture-recapture studies when capture probabilities vary among animals. Ph.D. thesis, Oregon State University, Corvallis.
- Burnham, K. P. and Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**, 625–633.
- Burnham, K. P. and Overton, W. S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology* **60**, 927–936.

- Chao, A. (1989). Estimating population size for sparse data in capture-recapture experiments. *Biometrics* **45**, 427–438.
- Chao, A., Lee, S.-M., and Jeng, S.-L. (1992). Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics* **48**, 201–216.
- Cormack, R. M. (1992). Interval estimation for mark-recapture studies of closed populations. *Biometrics* **48**, 567–576.
- Coull, B. A. (1997). Subject-specific modelling of capture-recapture experiments. Ph.D. thesis, University of Florida, Gainesville.
- 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. (1961). The two-sample capture-recapture census when tagging and sampling are stratified. *Biometrika* **48**, 241–260.
- 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.
- Fienberg, S. E. (1972). The multiple-recapture census for closed populations and incomplete  $2^k$  contingency tables. *Biometrika* **59**, 591–603.
- 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.
- Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* **76**, 133–140.
- Lee, S. and Chao, A. (1994). Estimating population size via sample coverage for closed capture-recapture models. *Biometrics* **50**, 88–97.
- Lloyd, C. J. (1992). Modified martingale estimation for recapture experiments with heterogeneous capture probabilities. *Biometrika* **79**, 833–836.
- Lloyd, C. J. and Yip, P. (1991). A unification of inference from capture-recapture studies through martingale estimating functions. In *Estimating Equations*, Godambe, V. P. (ed), 65–88. Oxford: Clarendon Press.
- Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised edition. West Sussex, U.K.: Wiley.
- Mingoti, S. A. and Meeden, G. (1992). Estimating the total number of distinct species using presence and absence data. *Biometrics* **48**, 863–875.
- 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.
- Plante, N., Rivest, L. P., and Tremblay, G. (1998). Stratified capture-recapture estimation of the size of a closed population. *Biometrics* **54**, 47–60.
- Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* **56**, 434–442.
- Pollock, K. H. and Otto, M. C. (1983). Robust estimation of population size in closed animal populations from capture-recapture experiments. *Biometrics* **39**, 1035–1050.
- Pollock, K. H., Hines, J. E., and Nichols, J. D. (1984). The use of auxiliary variables in capture-recapture and removal experiments. *Biometrics* **40**, 329–340.
- 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. *Proceedings of the National Academy of Sciences (USA)* **86**, 7658–7662.
- 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.
- Schnabel, Z. E. (1938). The estimation of the total fish population of a lake. *American Mathematical Monthly* **45**, 348–352.
- Schwarz, C. J. and Seber, G. A. F. (1999). Estimating animal abundance: Review III. *Statistical Science* **14**, 427–456.
- Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters*, 2nd edition. London: Charles Griffin.
- Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.
- Smith, E. P. and van Belle, G. (1984). Nonparametric estimation of species richness. *Biometrics* **40**, 119–129.
- Yip, P. (1991). A method of inference for a capture-recapture experiment in discrete time with variable capture probabilities. *Communications in Statistics—Stochastic Models* **7**, 343–362.

Received September 2001. Revised December 2002.

Accepted December 2002.

## APPENDIX

The analytic gradient of the log-likelihood function in (6) is  $\nabla l = (\partial l / \partial \mu, \partial l / \partial \tau, \partial l / \partial n_0)^t$ . The elements of  $\nabla l$  can be evaluated using the digamma function  $\psi(\cdot) = h'(\cdot)$  as follows:

$$\begin{aligned} \frac{\partial l}{\partial \mu} = & \tau \left( (n_0 + n)[\psi\{\tau(1 - \mu)\} - \psi(\tau\mu)] \right. \\ & + n_0[\psi(\tau\mu) - \psi\{T + \tau(1 - \mu)\}] \\ & \left. + \sum_{x=1}^T n_x[\psi(x + \tau\mu) - \psi\{T - x + \tau(1 - \mu)\}] \right) \end{aligned}$$

$$\begin{aligned}\frac{\partial l}{\partial \tau} &= (n_0 + n)[\psi(\tau) - \mu\psi(\tau\mu) \\ &\quad - (1 - \mu)\psi\{\tau(1 - \mu)\} - \psi(T + \tau)] \\ &\quad + n_0[\mu\psi(\tau\mu) + (1 - \mu)\psi\{T + \tau(1 - \mu)\}] \\ &\quad + \sum_{x=1}^T n_x[\mu\psi(x + \tau\mu) + (1 - \mu)\psi\{T - x + \tau(1 - \mu)\}]\end{aligned}$$

$$\begin{aligned}\frac{\partial l}{\partial n_0} &= \psi(n_0 + n + 1) - \psi(n_0 + 1) + h(\tau) - h\{\tau(1 - \mu)\} \\ &\quad - h(T + \tau) + h\{T + \tau(1 - \mu)\}\end{aligned}$$

The elements of the Hessian matrix  $\mathbf{H}$  of the log-likelihood function in (6) can be evaluated using the trigamma function  $\psi'(\cdot) = h''(\cdot)$  as follows:

$$\begin{aligned}\frac{\partial^2 l}{\partial \mu^2} &= \tau^2 \left( -(n_0 + n)[\psi'(\tau\mu) + \psi'\{\tau(1 - \mu)\}] \right. \\ &\quad + n_0[\psi'(\tau\mu) + \psi'\{T + \tau(1 - \mu)\}] \\ &\quad \left. + \sum_{x=1}^T n_x[\psi'(x + \tau\mu) + \psi'\{T - x + \tau(1 - \mu)\}] \right)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l}{\partial \tau^2} &= (n_0 + n)[\psi'(\tau) - \mu^2\psi'(\tau\mu) \\ &\quad - (1 - \mu)^2\psi'\{\tau(1 - \mu)\} - \psi'(T + \tau)]\end{aligned}$$

$$\begin{aligned}&+ n_0[\mu^2\psi'(\tau\mu) + (1 - \mu)^2\psi'(T + \tau\{1 - \mu\})] \\ &+ \sum_{x=1}^T n_x[\mu^2\psi'(x + \tau\mu) + (1 - \mu)^2\psi'\{T - x + \tau(1 - \mu)\}]\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l}{\partial \mu \partial \tau} &= \frac{1}{\tau} \frac{\partial l}{\partial \mu} + \tau \left( (n_0 + n)[(1 - \mu)\psi'\{\tau(1 - \mu)\} - \mu\psi'(\tau\mu)] \right. \\ &\quad + n_0[\mu\psi'(\tau\mu) - (1 - \mu)\psi'\{T + \tau(1 - \mu)\}] \\ &\quad + \sum_{x=1}^T n_x[\mu\psi'(x + \tau\mu) - (1 - \mu) \\ &\quad \left. \times \psi'\{T - x + \tau(1 - \mu)\}] \right)\end{aligned}$$

$$\frac{\partial^2 l}{\partial n_0^2} = \psi'(n_0 + n + 1) - \psi'(n_0 + 1)$$

$$\frac{\partial^2 l}{\partial \mu \partial n_0} = \tau[\psi\{\tau(1 - \mu)\} - \psi\{T + \tau(1 - \mu)\}]$$

$$\begin{aligned}\frac{\partial^2 l}{\partial \tau \partial n_0} &= \psi(\tau) - \psi(T + \tau) \\ &\quad + (1 - \mu)[\psi\{T + \tau(1 - \mu)\} - \psi\{\tau(1 - \mu)\}]\end{aligned}$$