PENNSTATE

Phone:(814)865-1220 / Fax:(814)863-7114

Center for Likelihood Studies
Department of Statistics

The Pennsylvania State University
422 Thomas Building
University Park, PA 16802

Identifiability and One-Sided Inference in the Species Problem

By Changxuan Mao and Bruce G. Lindsay

Technical Report #01-11-14

2001

Center for Likelihood Studies
Department of Statistics
The Pennsylvania State University
University Park, PA 16802

Identifiability and one-sided inference in the species problem

Chang Xuan Mao and Bruce G. Lindsay

Chang Xuan Mao (Corresponding author)

Graduate student

Interdepartmental Group in Biostatistics

University of California, Berkeley

367 Evans Hall

Berkeley, CA 94720-3860

Email:cmao@stat.berkeley.edu

Bruce G. Lindsay

Distinguished Professor of Statistics

Department of Statistics

Pennsylvania State University

326 Thomas Building

University Park, PA 16802-2111.

Identifiability and one-sided inference in the species problem

Chang Xuan Mao and Bruce G. Lindsay

November 14, 2001

Abstract

In the classical species problem, the unknown N number of distinct classes in a population can be estimated by sampling individuals from the population. There are several estimation procedures available when the abundance levels of classes are allowed to be different. In this paper, it will be shown in a Poisson mixture model that N is identifiable with heterogeneous class abundances. However, the only way to construct a confidence interval with a target coverage $1 - \alpha$ is to let the upper confidence limit be infinity. That is, estimation of number of classes is a one-sided inference issue in the sense of Donoho (1988) in that it is possible to obtain sensible lower confidence limits for any target confidence coverage but informative upper confidence limits are not available. Finally, we extend our discussion on identifiability and one-sided inference issues on population size estimation in capture-recapture studies.

Key words Number of species; Population size; Capture-recapture; One-sided inference; Poisson mixture; Binomial mixture.

1 Introduction

Since Fisher, Corbet and Williams (1943) established a statistical model for estimation the number of butterfly species, the species problem has attracted many researchers. A wide variety of applications have been found, such that the term "species" have been endowed with many new meanings. The "species" may represent taxa in a ecological community (Colwell and Coddington 1994), distinct records in a database (Arnold and Beaver 1988), dies used in the minting process (Esty 1983), words known by an author (Efron and Thisted 1976), executions in Vietnam (Bickel and Yahav 1985), and expressed genes in a cDNA library (Mao and Lindsay 2001); see Bunge and Fitzpatrick (1993) for a review.

The statistical description is that the individuals in a target population are classified into unknown N distinct classes being indexed by 1, 2, ..., N and a random sample of individuals is drawn from the population which is used to make inference about N. Let x_i be the number of individuals from the ith class that are present in the sample, called the *frequency* of the ith class. Let Ω be the set of all nonnegative integers and Ω_0 be the set of all positive integers. For j in Ω, let n_j be the number of classes that occur with frequency j and

n_+ be the number of distinct classes observed in the sample, that is,

$$n_j = \sum_{i=1}^{N} I(x_i = j) \text{ and } n_+ = \sum_{i=1}^{N} I(x_i > 0) = \sum_{j \in \Omega_0} n_j.$$

Note that n_0 is not observed and $N = n_0 + n_+$. Let $\mathbf{n}$ be the vector of observed frequency counts $\{n_j : j \in \Omega_0\}$. The issue of interest is estimation of N based on the observed frequency counts $\mathbf{n}$.

In this paper each x_i is assumed to follow a Poisson distribution with mean λ_i. Given the λ_i's, the x_i's are assumed to be independent and constitute a *Poisson sample*. Let s be the total number of individuals in the sample, that is, $s = \sum_{i=1}^{N} x_i$. Conditioning on s, the x_i's follow a multinomial distribution with index N and cell probabilities π_i's with $\pi_i = \lambda_i / (\sum_{i=1}^{N} \lambda_i)$. Thus the Poisson sample becomes a *multinomial sample* by conditioning. Note that π_i represents the relative abundance of the ith class in the population.

There are a number of estimation procedures proposed in the literature either based on the Poisson sample or the multinomial sample. In a Poisson model, the λ_i's may take distinct values. These λ_i's are often assumed to arise as a random sample from a *latent distribution* so that marginally the x_i's come from a *Poisson mixture*. The Poisson mixture model was used by Efron and Thisted (1976), Mao and Lindsay (2001), Norris and Pollock (1998), Ord and Whitemore (1986) and Zelterman (1988) among others. Alternative approaches are based on the multinomial sample, such as the nonparametric models in Chao (1984) and Chao and Lee (1992) in which the π_i's are allowed to take distinct values.

Although there has been so much work done for the species problem, I. J. Good's gave the following remark saying that " I don't believe it is usually possible to estimate the number of species, but only an appropriate lower bound to that number. This is because there is nearly always a good chance that there are a very large number of extremely rare species"; see Bunge and Fitzpatrick (1993). This motivates the authors to consider that it is possible that there is no informative upper confidence limit for N can be found for N based on the observed frequency counts $\mathbf{n}$.

This paper is organized as follows. In Section 2, we formulate our Poisson mixture model. In Section 3, the identifiability issues in the Poisson mixture model will be addressed. Then we discuss some moment results which will be used to discuss one-sided inference issues. In Section 5, we will show that upper confidence limit is often non-informative for a fixed target coverage of an confidence interval. In Section 6, our discussion on identifiability issues and one-sided inference will be extended to capture-recapture studies.

2 The Poisson mixture model

Let $f(x;\lambda)$ be the Poisson density with mean λ with respect to counting measure on Ω, where

$$f(x;\lambda) = e^{-\lambda}\lambda^x/x!, \quad x \in \Omega, \quad \lambda \in (0, +\infty).$$

To form a mixture model, we assume that λ is from a latent distribution $Q(\lambda)$ and that x has a Poisson distribution conditioning on λ.

The marginal density of x, called the *mixture distribution*, is denoted by $f(x; Q)$, which can be written as

$$f(x; Q) = \int f(x; \lambda) \mathrm{d}Q(\lambda), \quad x \in \Omega.$$

Given N and Q, the joint density of the frequencies x_i's can be written as

$$\prod_{i=1}^{N} f(x_i; Q) = \prod_{j \in \Omega} f(j; Q)^{n_j}.$$

The joint density of the frequency counts $\{n_j : j \in \Omega\}$ is given by

$$\frac{N!}{\prod_{j \in \Omega} n_j!} \prod_{j \in \Omega} f(j; Q)^{n_j}.$$

Since n_0 is not observed, we consider the density of $(n_+, \mathbf{n})$ which is denoted by $L(N, Q)$. Note that we have suppressed the dependence of $L(N, Q)$ on $(n_+, \mathbf{n})$ for notational convenience. It is the observed data likelihood, where

$$L(N, Q) = \frac{N! [f(0; Q)]^{N - n_+} \prod_{j \in \Omega_0} f(j; Q)^{n_j}}{(N - n_+)! \prod_{j \in \Omega_0} n_j!} \tag{1}$$

There are two parameters in $L(N, Q)$, an integer-valued parameter N and an infinite dimensional parameter Q. Note that $L(N, Q)$ is the product of a binomial density for n_+ and a multinomial density for $\mathbf{n}$ given n_+. The density of $\mathbf{n}$ given n_+ is denoted by $L_c(Q)$, which can be written as

$$L_c(Q) = \frac{n_+!}{\prod_{j \in \Omega_0} n_j!} \prod_{j \in \Omega_0} \left[\frac{f(j; Q)}{1 - f(0; Q)} \right]^{n_j}. \tag{2}$$

Also, the dependence of $L_c(Q)$ on $\mathbf{n}$ is suppressed. The density of n_+ is denoted $L_m(N, Q)$ and given by

$$L_m(N, Q) = \frac{N! f(0; Q)^{N - n_+} [1 - f(0; Q)]^{n_+}}{n_+! (N - n_+)!}. \tag{3}$$

The dependence of $L_m(N,Q)$ on n_+ is suppressed. It is clear that

$$L(N,Q) = L_m(N,Q)L_c(Q).$$

Let $\mathcal{Q}$ be the set of all latent distributions for Poisson densities. The model that we are considering is

$$\mathcal{M} = \{L(N,Q) : N \in \Omega_0, Q \in \mathcal{Q}\}.$$

We may only consider submodels of the model $\mathcal{M}$. There are several parametric submodels discussed in the literature such as negative binomial and inverse Gauss Poisson. The latent distribution Q may be assumed to have finite number of support points; see Mao and Lindsay (2001) for nonparametric estimation about N, which says consistent estimators exists for finite Poisson mixture models.

3 Identifiability issues

We will discuss identifiability issues about the model $\mathcal{M}$. It is useful to reformulate the model $\mathcal{M}$ in order to show that the parameters in $\mathcal{M}$ are identifiable. To do so, for each Q in $\mathcal{Q}$, let P be a re-weighted measure from Q defined by

$$\mathrm{d}P(\lambda) = e^{-\lambda}\mathrm{d}Q(\lambda)/[\int(1-e^{-\lambda})\mathrm{d}Q(\lambda)]. \tag{4}$$

Let $\mathcal{P}$ be the set of all such measures P which are reweighted from latent distributions Q in $\mathcal{Q}$. It is clear that via (4), we have defined a mapping ψ from $\mathcal{Q}$ to $\mathcal{P}$ where $P = \psi(Q)$ for each Q in $\mathcal{Q}$. From (4), we have

$$\mathrm{d}Q(\lambda) = e^{\lambda}\mathrm{d}P(\lambda)/[\int e^{\lambda}\mathrm{d}P(\lambda)]. \tag{5}$$

Here, the latent distribution $Q(\lambda)$ is also a re-weighted version of $P(\lambda)$. It is clear that ψ is a bijection from $\mathcal{Q}$ to $\mathcal{P}$, and via (5) we have also defined a mapping $\phi = \psi^{-1}$ with $Q = \phi(P)$ for each P in $\mathcal{P}$. We can think that P and Q as alternating ways to represent the unknown latent distribution in our model $\mathcal{M}$ and it will be useful to move back and forth between these representations.

For any finite measure ν on $(0, +\infty)$, define $\mu_x(\nu)$ by

$$\mu_x(\nu) = \int \lambda^x \mathrm{d}\nu(\lambda), \quad x \in \Omega$$

where $\mu_x(\nu)$ is the xth moment of ν when $x \in \Omega_0$ and $\mu_0(\nu)$ is the total mass of ν. From simple algebra, we obtain the relationship between the moments of P and the mixture density $f(x;Q)$ as follows:

Lemma 1 *For each x in Ω, we have*

$$\mu_x(P) = x! f(x;Q)/[1 - f(0;Q)] \tag{6}$$

An important quantity in our analysis is the odds of getting a zero frequency, $\mu_0(P) = f(0;Q)/[1 - f(0;Q)]$, which will be written as $\theta(Q)$ or θ with the dependence on Q being suppressed. Given θ, the density of n_+ can be regarded as a likelihood for N, where

$$L_m(N;\theta) = \frac{N!}{n_+!(N-n_+)!} \frac{\theta^{N-n_+}}{(1+\theta)^N} \tag{7}$$

which is obtained from (3) using (6). The maximum likelihood estimator given θ for N is given by

$$\widehat{N}(\theta) = n_+ + n_+\theta. \tag{8}$$

See Lindsay and Roeder (1987). If an estimator for θ is plugged into $\widehat{N}(\theta)$, then we obtain an estimator for N called the pseudo maximum likelihood estimator by Gong and Samaniego (1981).

From (8), it is clear that estimation of N requires to develop sensible estimators for θ. The density of $\mathbf{n}$ given n_+ can be regarded as likelihood for P to make inference for P and hence θ. Substituting (6) into (2), we have

$$L_c(\phi(P)) = \frac{n_+!}{\prod_{j\in\Omega_0} n_j!} \prod_{x\in\Omega_0} \left[\frac{\mu_x(P)}{x!}\right]^{n_x}. \tag{9}$$

It is clear that the moment $\mu_x(P)$ of P is identifiable for each x in Ω_0 from (9). The following theorem is about P and its identifiable moments.

Theorem 2 *The measure P is uniquely determined by its moment sequence $\{\mu_x(P) : x \in \Omega_0\}$.*

The proof is simple. Let $\mathrm{d}P_*(\lambda) = \lambda \mathrm{d}P(\lambda)$, which yields $\mathrm{d}P(\lambda) = \lambda^{-1}\mathrm{d}P_*(\lambda)$. For $t < 1$, the moment generating function of $P_*(\lambda)$ satisfies

$$\int e^{\lambda t}\mathrm{d}P_*(\lambda) = \frac{\int e^{\lambda(t-1)}\lambda \mathrm{d}Q(\lambda)}{\int (1-e^{-\lambda})\mathrm{d}Q(\lambda)} < +\infty.$$

Thus $\{\mu_x(P_*) : x \in \Omega\}$ uniquely determines $P_*(\lambda)$ and the theorem holds by noting that $\mu_x(P_*) = \mu(x+1; P)$ for x in Ω.

Theorem 2 shows that P is identifiable. Now we are prepared to discuss the identifiability of N and Q.

Corollary 3 *The parameters N and Q are identifiable.*

The proof is easy. Since P is identifiable, it is clear that θ is identifiable due to (6) and Q is identifiable due to (5). Then N is identifiable via (7) since θ is identifiable.

4 Some moment results

It is clear that from (7) that our ability to make inference on N is closely related to our ability to infer $\theta = \mu_0(P)$ based on the conditional likelihood (9). To study the question, we first develop results in which we will see to what extent, $\theta = \mu_0(P)$ can be determined by $\{\mu_x(P) : x \in \Omega_0\}$, in particular, how $\mu_0(P)$ varies when $\{\mu_x(P) : x \in \Omega_0\}$ varies.

To facilitate the discussion, we will endow $\mathcal{P}$ with a distance defined by

$$d(P_1, P_2) = \sum_{x \in \Omega_0} |\mu_x(P_1) - \mu_x(P_2)|, P_1, P_2 \in \mathcal{P}.$$

Note that the distance is possible to take value $+\infty$. It is clear that the distance is well-defined based on Theorem 2. The essence using moments to define distance is that we want to know how the moments will affect the total mass.

A trivial lower bound for $\mu_0(P)$ is zero, which is of no interest. A simple lower bound for $\mu_0(P)$ is $\beta(P) = \mu_1^2(P)/\mu_2(P)$, that is, $\mu_0(P) \geq \beta(P)$. Note that from Cauchy-Schwartz inequality, we have

$$(\int \lambda \mathrm{d}P(\lambda))^2 \leq \int \lambda^2 \mathrm{d}P(\lambda) \times \int \mathrm{d}P(\lambda)$$

with $\mu_0(P) = \beta(P)$ if and only if P is degenerate. For the lower bound $\beta(P)$, the following result is clear.

Lemma 4 *For each ϵ with $0 < \epsilon < \beta(P)$, there exists $\delta > 0$, such that*

$$\mu_0(P') \geq \beta(P') > \beta(P) - \epsilon, \quad \forall P' \in \mathcal{P} \text{ with } d(P, P') < \delta.$$

The proof is straightforward and omitted. This lemma says that when P' is close to P, its total mass possesses a nontrivial lower bound, which is estimable. For more elaborate lower bounds in the case that P is non-degenerate the readers are referred to Mao and Lindsay (2001).

However, no matter how small $d(P, P')$ is, there exists particular P' which has total mass $\mu_0(P')$ much larger than $\mu_0(P)$ at any specified level. In fact, we have

Lemma 5 *For each t and s in $(0, 1)$, there exists P' in $\mathcal{P}$, such that*

$$d(P, P') < Ct \text{ but } \mu_0(P') - \mu_0(P) = -\log s$$

with C being a constant. The proof is constructive and is given in the Appendix.

Corollary 6 *Fixed P, for any $\epsilon > 0$ and any $\delta > 0$, there exists P' such that*

$$d(P, P') < \epsilon \text{ and } \mu_0(P') - \mu_0(P) > \delta.$$

Here is the proof. Let $t_* = \min(\epsilon/C, 1)$ and $s_* = e^{-\delta}$. Then

$$d(P, P_{s,t}) < \epsilon \text{ and } \mu_0(P_{s,t}) - \mu_0(P) > \delta, s < s_*, t < t_*.$$

The corollary holds by taking P' to be any such $P_{s,t}$.

The moment results in Lemma 4 and Corollary 6 suggest that one-sided inference issues may arise in the species problem. In Bahadur and Savage (1956), it was shown that there is no approach to constructing confidence intervals for the mean nonparametrically. Donoho (1988) considered one-sided inference about functionals of a density and showed that for many interesting functionals such as the number of modes and the number of support points of a finite mixture, it is impossible to get two-sided confidence intervals. After studying several examples, Donoho (1988) concluded that "...one can only get one-sided information, unless prior or external information is available", in particular, when "the quantity of interest is a measure of the complexity of a system." In Pfanzagl (1998) various results on the nonexistence of confidence procedures were summarized. The ideas in Bahadur and Savage (1956) and Donoho (1988) are adapted here to analyze the one-sided inference in the species problem.

5 One-sided inference

We start from considering conditional confidence inference about θ in the conditional distribution (9). As before, let $\mu_0(P)$ be written as $\theta(Q)$. Let $\hat{\theta}_u$ be a $(1-\alpha)$-level upper confidence limit for $\theta(Q)$, where probability calculations are made conditionally on n_+. We will allow $\hat{\theta}_u = +\infty$ if needed to obtain the target confidence coverage. By this we mean that

$$P_Q\{\hat{\theta}_u \geq \theta(Q)\} \geq 1-\alpha, \quad \forall Q \in \mathcal{Q} \tag{10}$$

The claim is that this implies

Theorem 7 $P_Q\{\hat{\theta}_u = +\infty\} \geq 1-\alpha, \quad Q \in \mathcal{Q}.$

The proof is given in the Appendix. The theorem says that the upper confidence limit for the total mass can take the noninformative value $+\infty$ with probability no less than $1-\alpha$ uniformly over $\mathcal{Q}$.

We next discuss confidence inference about N in the model $\mathcal{M}$ unconditionally, which are more desirable since N is the parameter of primary interest. However, the result about N in next theorem is a little bit weaker than that about the total mass θ in Theorem 7.

Further discussion requires the introduction of the total variation distance. Let $(\mathcal{X}, \rho)$ be a metric space with ρ being the metric. Let H_1 and H_2 be two probabilities on the measurable space $(\mathcal{X}, \rho, \mathcal{B})$ with $\mathcal{B}$ being Borel σ-field generated by all open sets. The total variation distance of H_1 and H_2 is denoted by $\tau(H_1, H_2)$ and given by

$$\tau(H_1, H_2) = 2 \sup_{B \in \mathcal{B}} |H_1(B) - H_2(B)|.$$

Let ν be a σ-finite measure on $(\mathcal{X}, \rho, \mathcal{B})$ dominating H_1 and H_2. It is quite simple to show that $\tau(H_1, H_2)$ equals the L_1 distance of $\mathrm{d}H_1/\mathrm{d}\nu$ and $\mathrm{d}H_2/\mathrm{d}\nu$, the densities (Radon-Nikodym derivatives) of H_1 and H_2 with respect to ν, (see Bickel et al. 1998), that is,

$$\tau(H_1, H_2) = \int |\mathrm{d}H_1/\mathrm{d}\nu - \mathrm{d}H_2/\mathrm{d}\nu| \mathrm{d}\nu.$$

The two alternating expressions for the total variation distance are also important to the proofs for Theorem prop:five and the next theorem.

Let $\widehat{N}_u$ be a $(1-\alpha)$-level upper confidence limit for N. We will also allow $\widehat{N}_u = +\infty$ in order to obtain the target confidence coverage. We consider $\widehat{N}_u$ that satisfies

$$P_{N,Q}\{\widehat{N}_u \geq N\} \geq 1-\alpha, \quad \forall N \in \Omega_0, Q \in \mathcal{Q}. \tag{11}$$

Let $\gamma = 1 - f(0; Q)$ and $\eta(N,\gamma)$ be the total variation distance between a binomial density with index N and probability γ and its Poisson approximation with mean parameter $N\gamma$. Based on (11), we have the following result about the upper confidence limit $\widehat{N}_u$.

Theorem 8 $P_{N,Q}\{\widehat{N}_u = +\infty\} \geq 1-\alpha-\eta(N,\gamma)/2, \quad N \in \Omega_0, Q \in \mathcal{Q}.$

See the Appendix for the proof.

Note that the probability lower bound in Theorem 8 depends on N and γ, which is not like the probability lower bound in Theorem 7. However, in Theorem 8, we can choose N and γ such that $\eta(N,\gamma)$ can be arbitrary small. Table 5 gives some numeric examples about $\eta(N,\gamma)$. It turns out that $\eta(N,\gamma)$ depends on γ much more than on N. The numeric results also show that when a large proportion of classes are observed, the probability that the upper confidence limit takes infinity will be small. It is consistent with our intuition.

$N\backslash\gamma$	0.0200	0.0300	0.0400	0.0500	0.0600	0.0700	0.0800	0.0900
100	0.0091	0.0152	0.0205	0.0252	0.0295	0.0351	0.0409	0.0464
1000	0.0097	0.0147	0.0198	0.0249	0.0299	0.0352	0.0404	0.0457
10000	0.0098	0.0147	0.0198	0.0248	0.0299	0.0351	0.0403	0.0456
$N\backslash\gamma$	0.1000	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000
100	0.0517	0.1083	0.1731	0.2456	0.3332	0.4372	0.5672	0.7283
1000	0.0510	0.1079	0.1722	0.2458	0.3321	0.4360	0.5659	0.7398
10000	0.0510	0.1079	0.1722	0.2459	0.3321	0.4359	0.5657	0.7396

Table 1: The total variation distance between a binomial density with index N and probability γ and its Poisson approximation with mean $N\gamma$.

6 Capture-recapture studies

The inference about the number of classes can be based on datasets with only presence or absence information for each class available at M independent observing occasions. Each class is present or absent independently at each occasion with the same probability over occasions. Let x_i be the number of occasions on which the ith class present and π_i be the probability for the ith class to be present. Then x_i is a binomial random variable with index M and odds $\lambda_i = \pi_i/(1-\pi_i)$. The x_i's are independent given the λ_i's and constitute a *binomial sample*. The x_i's are also called *frequencies*. It is clear that n_+ and the *frequency counts* n_j's can be defined as before.

Such binomial samples also arise from capture-recapture studies with M capture occasions. Capture-recapture experiments are useful tools to estimate the size of a target population which consists of N individuals. In a capture-recapture experiment, x_i is the number of captures of the ith individual provided that the individuals are indexed by 1, 2, ..., N. When we assume that each individual has the same probability to be captured across capture occasions, then each x_i is a binomial random variable. For capture-recapture studies, Chao (2001) is a comprehensive review.

Now let Ω^b be the set of nonnegative integers no greater than M and Ω_0^b be the set of natural numbers no greater than M. Also let $f^b(x;\lambda)$ be the binomial density with odds λ with respect to counting measure on Ω^b, where

$$f^b(x;\lambda) = [1+\lambda]^{-M}\lambda^x/x!, \quad x \in \Omega^b, \quad \lambda \in (0, +\infty).$$

We can also form a mixture model by assuming that λ is from a latent distribution $Q^b(\lambda)$ and that x has a binomial distribution conditioning on λ. The marginal density of x is $f^b(x;Q^b) = \int f^b(x;\lambda)\mathrm{d}Q^b(\lambda)$. Given N and Q^b, the joint density of the frequencies x_i's can be written as

$$\prod_{i=1}^{N} f^b(x_i;Q^b) = \prod_{j\in\Omega^b} f^b(j;Q^b)^{n_j}.$$

The joint density of the frequency counts $\{n_j : j \in \Omega^b\}$ is given by

$$\frac{N!}{\prod_{j\in\Omega^b} n_j!} \prod_{j\in\Omega^b} f^b(j;Q^b)^{n_j},$$

which yields the the observed data likelihood $L(N, Q^b)$, where

$$L(N,Q^b) = \frac{N![f^b(0;Q^b)]^{N-n_+}\prod_{j\in\Omega_0^b} f^b(j;Q^b)^{n_j}}{(N-n_+)!\prod_{j\in\Omega_0^b} n_j!}. \qquad (12)$$

Note that $L(N, Q^b)$ is also the product of a binomial density for n_+, denoted by $L_m(N, Q^b)$, and a multinomial density for $\mathbf{n}$ given n_+, denoted by $L_c(Q)$, where

$$L_c(Q^b) = \frac{n_+!}{\prod_{j \in \Omega_0^b} n_j!} \prod_{j \in \Omega_0^b} \left[\frac{f^b(j; Q^b)}{1 - f^b(0; Q^b)} \right]^{n_j} \tag{13}$$

and

$$L_m(N, Q^b) = \frac{N! f^b(0; Q^b)^{N-n_+} [1 - f^b(0; Q^b)]^{n_+}}{n_+!(N - n_+)!}. \tag{14}$$

Also we have

$$L(N, Q^b) = L_m(N, Q^b) L_c(Q^b).$$

Let $\mathcal{Q}^b$ be the set of all latent distributions for binomial densities $f^b(x; \lambda)$. The model under consideration is

$$\mathcal{M}^b = \{L(N, Q^b) : N = 1, 2, \ldots, Q^b \in \mathcal{Q}^b\}.$$

The parametric submodel discussed in the literature is beta-binomial; see Burnham and Overton (1978). In Mao and Lindsay (2001) Q^b was assumed to have finite number of support points.

It is clear our binomial mixture model for binomial samples are similar to our Poisson mixture model for Poisson samples by noting the structure similarity between (12), (13), (14), and (1), (2), (3) respectively. However, the model $\mathcal{M}^b$ is not identifiable. The reason is that we have only finitely many identifiable moments for P^b corresponding to Theorem 2 where P^b is the reweighted measure of Q^b defined by

$$\mathrm{d}P^b(\lambda) = (1 + \lambda)^{-M} \mathrm{d}Q^b(\lambda) / [\int (1 - (1 + \lambda)^{-M}) \mathrm{d}Q^b(\lambda)].$$

In order to create an identifiable model, we must consider submodels of the model $\mathcal{M}^b$. One way to do so is to assume that P^b is in the class of measures uniquely determined by its their finite moment sequences, to be specific, from the first moment up to Mth moment. Such an assumption allows us to develop similar one-sided inference results as what we have for a Poisson mixture model by noting that our one-sided inference discussion does not dependent on any specific property of the Poisson distribution and Poisson mixture provided that we have identifiable models.

7 Discussion

Although in general two-sided confidence intervals are more desirable than one-sided confidence interval, our analyses show that a two-sided confidence interval for a target coverage probability will often take the noninformative upper confidence limit and become a one-sided confidence interval. The data with finite sample size do allow us to set appropriate lower confidence limits for the number of classes while informative upper confidence limits are usually hopeless unless we have informative external to the data. Two-sided confidence intervals are actually sensible for parameters only equal to N under explicit or implicit restrictions on the latent distribution. In the literature, two-sided confidence intervals for N were often reported, for example, see Chao (1984) and Chao and Lee (1992) in nonparametric settings. However, the claimed coverage should be questionable. Two-sided confidence intervals via bootstrap were discussed in Mao and Lindsay (2001) where the confidence intervals were not symmetric about the estimate and the upper confidence limits were often extremely large. It is clear that such phenomena are not surprising as a target coverage requires the upper confidence limit to be noninformative.

References

Arnold, B. and Beaver, R. J. (1988). Estimation of the number of classes in a population. *Biometrical Journal*, 30:413–424.

Bahadur, R. and Savage, L. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics*, 27:1115–1122.

Bickel, P., Klassen, C., Ritov, Y., and Wellner, J. (1998). *Efficient and adaptive estimation for semiparametric models.* Johns Hopkins University Press, Baltimore.

Bickel, P. J. and Yahav, J. A. (1985). On estimating the number of unseen species: how many executions were there? Technical report, Department of Statistics, University of California, Berleley.

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. and Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals (corr: V68 p345). *Biometrika*, 65:625–634.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, 11:265–270.

Chao, A. (2001). An overview of closed capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics*, 6:138–155.

Chao, A. and Lee, S. M. (1992). Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, 87:210–217.

Colwell, R. K. and Coddingtion, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc . Lond. B.*, 345:101–118.

Donoho, D. (1988). One-sided inference about functionals of a density. *The Annals of Statistics*, 16:1390–1420.

Efron, B. and Thisted, R. (1976). Estimating the number of unseen species: how many words did shakespeare know? *Biometrika*, 63:435–447.

Esty, W. W. (1983). A normal limit law for a nonparametric estimator of the coverage of a random sample. *The Annals of Statistics*, 11:905–912.

Fisher, R., Corbet, A. S., and Williams, C. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal ecology*, 12:42–58.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *Annals of Statistics*, 9:861–869.

Lindsay, B. and Roeder, K. (1987). A unified treatment of integer parameter models. *Journal of American Statistical Association*, 82:758–764.

Mao, C. X. and Lindsay, B. G. (2001). Moment-based nonparametric estimators for the number of classes in a population. Technical Report 01-06-15, The Likelihood Center, Department of Statistics, Pennsylvania State University.

Norris, J. L. I. and Pollock, K. H. (1998). Non-parametric mle for Poisson species abundance models allowing for heterogeneity between species. *Environmental and Ecological Statistics*, 5:391–402.

Ord, J. K. and Whitmore, G. . (1986). The Poisson-inverse Gaussian distribution as a model for species abundance. *Communications in Statistics, Part A – Theory and Methods*, 15:853–871.

Pfanzagl, J. (1998). The nonexistence of confidence sets for discontinuous functionals. *Journal of statistical planning and inference*, 75:9–20.

Zelterman, D. (1988). Robust estimation in truncated discrete distributions with application to capture-recapture experiments. *Journal of statistical planning and inference*, 18:225–237.

Appendix: Proofs

Proof for Lemma 5

The proof is constructive. For each t in $(0,1)$, define a function g_t on $(0,+\infty)$ as follows:

$$g_t(\lambda) = \left(\begin{array}{cc} t^{-1} & \lambda \in (0,t] \\ \lambda^{-1} & \lambda \in (t,1] \\ 0 & \lambda \in (1,+\infty) \end{array} \right.$$

Let G_t be the measure with density g_t with respect to Lebesgue measure on $(0,+\infty)$. Note that G_t is in $\mathcal{P}$. We have

$$\mu_0(G_t) = \int g_t(\lambda)\mathrm{d}\lambda = \int_0^t t^{-1}\mathrm{d}\lambda + \int_t^1 \lambda^{-1}\mathrm{d}\lambda = 1 - \log t$$

and when x is in Ω_0,

$$\mu_x(G_t) = \int \lambda^x g_t(\lambda)\mathrm{d}\lambda = \int_0^t t^{-1}\lambda^x\mathrm{d}\lambda + \int_t^1 \lambda^{x-1}\mathrm{d}\lambda = \frac{1}{x} - \frac{t^x}{x(x+1)}.$$

For each s and t in $(0,1)$, we have

$$\mu_0(G_{st}) - \mu_0(G_t) = -\log s$$

and

$$d(G_t, G_{st}) = \sum_{x\in\Omega_0} \left|\frac{t^x}{x(x+1)} - \frac{(st)^x}{x(x+1)}\right| = \sum_{x\in\Omega_0} \frac{(1-s^x)t^x}{x(x+1)} < Ct$$

with $C = \sum_{x\in\Omega_0}[x(x+1)]^{-1}$.

For each P in $\mathcal{P}$, define $P_{s,t}$ as follows:

$$P_{s,t} = P + G_{st} - G_t. \tag{15}$$

Note that $P_{s,t}$ is a well-defined measure on $(0, +\infty)$ and is in $\mathcal{P}$. It is clear that

$$\mu_x(P_{s,t}) - \mu_x(P) = \mu_x(G_{s,t}) - \mu_x(G_t) = -\log s > 0 \tag{16}$$

and

$$d(P, P_{s,t}) = d(G_t, G_{st}) < Ct \tag{17}$$

Therefore, the lemma holds.

Proof for Theorem 7

Let $Q_{s,t} = \phi(P_{s,t})$ with $P_{s,t}$ defined in (15) and $P = \psi(Q)$. Now we consider the total variation distance between $L_c(Q)$ and $L_c(Q_{s,t})$. Let $C_{\mathbf{n}}^{n_+} = n_+![\prod_{j\in\Omega_0} n_j!]^{-1}$. We have

$$\begin{aligned}
\tau(L_c(Q), L_c(Q_{s,t})) &= \sum_{\mathbf{n}} \left| C_{\mathbf{n}}^{n_+} \prod_{x\in\Omega_0} \left[\frac{\mu_x(P)}{x!}\right]^{n_x} - C_{\mathbf{n}}^{n_+} \prod_{x\in\Omega_0} \left[\frac{\mu_x(P_{s,t})}{x!}\right]^{n_x} \right| \\
&= \sum_{\mathbf{n}} C_{\mathbf{n}}^{n_+} \prod_{x\in\Omega_0} \left[\frac{\mu_x(P)}{x!}\right]^{n_x} \left\{ \prod_{x\in\Omega_0} \left[\frac{\mu_x(P_{s,t})}{\mu_x(P)}\right]^{n_x} - 1 \right\} \\
&\leq \sum_{\mathbf{n}} C_{\mathbf{n}}^{n_+} \prod_{x\in\Omega_0} \left[\frac{\mu_x(P)}{x!}\right]^{n_x} \left\{ \prod_{x\in\Omega_0} \left[1 + \frac{Ct}{\mu_x(P)}\right]^{n_+} - 1 \right\} \\
&\leq \epsilon(n_+, t)
\end{aligned}$$

where

$$\epsilon(n_+, t) = \sum_{\mathbf{n}} C_{\mathbf{n}}^{n_+} \prod_{x\in\Omega_0} \left[\frac{\mu_x(P)}{x!}\right]^{n_x} \left\{ \left[1 + \frac{Ct}{\min\{\mu_x(P) : x \in [1, n_+] \cap \Omega_0\}}\right]^{n_+^2} - 1 \right\}. \tag{18}$$

Note that (17) has been used in the derivation. Also, it is clear that $\epsilon(n_+, t)$ goes to zero when t goes to zero for each fixed n_+.

Since it is clear that

$$|P_Q\{\hat\theta_u \geq \theta(Q_{s,t})\} - P_{Q_{s,t}}\{\hat\theta_u \geq \theta(Q_{s,t})\}| \leq \sup_{B\in\mathcal{B}} |P_Q(B) - P_{Q_{s,t}}(B)|,$$

we have

$$P_Q\{\hat\theta_u \geq \theta(Q_{s,t})\} \geq P_{Q_{s,t}}\{\hat\theta_u \geq \theta(Q_{s,t})\} - \tau(L_c(Q), L_c(Q_{s,t}))/2$$

From (10) and (16), we then obtain

$$P_Q\{\hat\theta_u \geq \theta(Q) - \log s\} \geq 1 - \alpha - \epsilon(n_+, t)/2.$$

The desired result can be obtained by letting s and t goes to zero.

Proof for Theorem 8

For fixed N and Q, consider the following mapping from $(0, 1)$ to $(0, +\infty)$:

$$s \longmapsto K = K(s) = -N \log s/(1 + \theta(Q)).$$

This gives an inverse mapping from $(0, +\infty)$ to $(0, 1)$

$$K \longmapsto s = s(K) = \exp(-K(1 + \theta(Q)/N)).$$

It is clear that s goes to zero if and only if K goes to infinity and

$$N/[1 + \theta(Q)] = (N + K)/[1 + \theta(Q_{s,t})] \tag{19}$$

where $Q_{s,t} = \phi(P_{s,t})$ with $P_{s,t}$ being defined in (15) and $P = \psi(Q)$.

Now we will only consider s in the set $\mathcal{S} = \{\exp(-K(1+\theta(Q)/N)) : K \in \Omega_0\}$. Such a restriction forces $K = K(s)$ to be a natural number. Note that s in $\mathcal{S}$ and K in Ω_0 are related such that the marginal distributions $L_m(N+K, Q_{s,t})$ and $L_m(N,Q)$ have the same mean because of (19). It is clear that

$$\sum_{n_+} |L_m(N+K, Q_{s,t}) - L_m(N,Q)| \longrightarrow \eta(N,\gamma) \text{ as } K \longrightarrow +\infty. \quad (20)$$

Then calculate the total variation distance $\tau(L(N,Q), L(N+K, Q_{s,t}))$ between the densities $L(N,Q)$ and $L(N+K, Q_{s,t})$.

$$\begin{aligned}
& \tau(L(N,Q), L(N+K, Q_{s,t})) \\
= & \sum_{(n_+,\mathbf{n})} |L_m(N+K, Q_{s,t})L_c(Q_{s,t}) - L_m(N,Q)L_c(Q)| \\
\leq & \sum_{(n_+,\mathbf{n})} L_m(N,Q)|L_c(Q_{s,t}) - L_c(Q)| \\
& + L_c((Q_{s,t}))|L_m(N+K, Q_{s,t}) - L_m(N,Q)| \\
\leq & \sum_{(n_+,\mathbf{n})} L_m(N,Q)|L_c(Q_{s,t}) - L_c(Q)| \\
& + (L_c(Q) + |L_c(Q_{s,t}) - L_c(Q)|)|L_m(N+K, Q_{s,t}) - L_m(N,Q)|
\end{aligned}$$

So we have

$$\begin{aligned}
\tau(L(N,Q), L(N+K, Q_{s,t})) \leq & \sum_{n_+} (1+\epsilon(n_+,t))|L_m(N+K, Q_{s,t}) - L_m(N,Q)| \\
+ & \sum_{n_+} L_m(N,Q)\epsilon(n_+,t) \quad (21)
\end{aligned}$$

where $\epsilon(n_+,t) = 0$ if $n_+ = 0$ and $\epsilon(n_+,t)$ is defined in (18) when $n_+ \neq 0$.

Since it is clear that

$$|P_{N,Q}\{\widehat{N}_u \geq N+K)\} - P_{N+K,Q_{s,t}}\{\widehat{N}_u \geq N+K)\}| \leq \sup_{B\in\mathcal{B}} |P_{N,Q}(B) - P_{N+K,Q_{s,t}}(B)|,$$

we have

$$P_{N,Q}\{\widehat{N}_u \geq N+K)\} \geq P_{N+K,Q_{s,t}}\{\widehat{N}_u \geq N+K)\} - \tau(L(N,Q), L(N,Q_{s,t}))/2.$$

Letting t and s go to zero we obtain the theorem using (11), (18), (20) and (21).