



Nonparametric Estimators for the Number of Classes
in Multiple Populations

By CHANGXUAN MAO and BRUCE G. LINDSAY

Technical Report #01-06-30

2001

Center for Likelihood Studies

DEPARTMENT OF STATISTICS

THE PENNSYLVANIA STATE UNIVERSITY

UNIVERSITY PARK, PA 16802

Nonparametric estimators for the number of classes in multiple populations

Changxuan Mao and Bruce G. Lindsay

Changxuan Mao
Graduate student
Department of Statistics
Pennsylvania State University
326 Thomas Building
University Park, PA 16802

Bruce G. Lindsay
Distinguished Professor of Statistics
Department of Statistics
Pennsylvania State University
326 Thomas Building
University Park, PA 16802.

Corresponding author: Changxuan Mao

Nonparametric estimators for the number of classes in multiple populations

Changxuan Mao and Bruce Lindsay*

June 30, 2001

Abstract

Consider a collection of populations, each of them with infinitely many individuals. All individuals are classified into N disjoint classes. The statistical goal is the estimation of N , the total number of distinct classes in these populations from independent abundance-based or incidence-based random subsamples drawn from each population. There exists a collection of estimators by linear extrapolation of a multivariate log ratio function, which include several estimators developed in the literature, such as the well-known Lincoln-Peterson estimator and some triple system estimators used in census undercount adjustment. A second kind of estimation procedure is based on frequency-profile combinations, which allows the estimation procedure for one population to be applied directly. The asymptotic behavior and a bootstrap confidence inference methodology are presented for each estimator. Three datasets from epidemiology, ecology and genomics are studied for illustration.

*Changxuan Mao is a graduate student and Bruce Lindsay is Distinguished Professor of Statistics in the Department of Statistics, Pennsylvania State University, Thomas Building, University Park, PA 16802. Thanks to the Wild Bird Society of Hsin-Chu and the Institute of Genomic Research. Thanks to Drs. Anne Chao and Robert Colwell. Professor Lindsay's research was supported by National Science Foundation grant DMS-9870193.

KEY WORDS: Number of species; Number of shared species; Population size; Capture-recapture; Poisson mixture; Multivariate Poisson mixture; Multivariate binomial mixture; Moment estimator.

1 INTRODUCTION

Suppose there are K populations. Each of these populations has infinitely many individuals. The individuals are classified into some disjoint classes. Suppose there are totally N distinct classes which are present in at least one of these K populations. For each population, a random subsample is taken. The subsamples are assumed to be independent. If the number of individuals from each class is recorded in the subsample, the subsample is called an *abundance-based* subsample. If only presence or absence information of each class is available in the subsample, the subsample is called an *incidence-based* subsample. The issue of interest is the estimation of N .

The authors were motivated to address estimation in multiple populations by certain problems in genomic studies. When scientists sequence sampled clones from cDNA libraries made from different tissues, they are interested in the overlap of the expressed genes in different tissues. One overlap index is the number of genes expressed in both tissues (Gower 1985), denoted by N_{12} . Here, N represents the total number genes expressed in at least one of two tissues. Let N_1 (N_2) be the number of genes expressed in the first (second) tissue. Then $N_{12} = N_1 + N_2 - N$. There are many approaches for the estimation of N_1 (N_2), see Bunge and Fitzpatrick (1993). Let \hat{N}_1 (\hat{N}_2) be an estimator for N_1 (N_2). If we can estimate N by \hat{N} , then N_{12} can be estimated by $\hat{N}_{12} = \hat{N}_1 + \hat{N}_2 - \hat{N}$.

This estimation problem is the genomic version of the problem of estimating of the number of species present in at least one of two ecological communities, which has been proposed by ecologists and regarded as a challenge to statisticians for a long time (Colwell and Coddington 1994).

While the estimation of the number of distinct classes in one population has received extensive research (Bunge and Fitzpatrick 1993), relatively little attention has been paid to the estimation of the number of distinct classes in at least one of multiple populations. Only Chao et al. (2000), an updated version of Chen et al. (1995), presented a generalized coverage approach for the estimation of the number of species present in both of two communities from two abundance-based subsamples.

In ecological species collection projects, a target community is often divided into some sub-communities in which different sampling methods are used to access species (Longino, Coddington and Colwell 2001). One type of dataset is made up of multiple abundance-based subsamples, which have never been treated statistically (Colwell, personal communication, April 2001).

In this paper, several new nonparametric methods are developed for the estimation of the number of distinct classes N in at least one of multiple populations. We start from introducing some necessary concepts and notations. Let the classes be indexed by $1, 2, \dots, N$, and let the populations and subsamples be indexed by $1, 2, \dots, K$, where the k th subsample is from the k th population. We first consider abundance-based subsamples. For each i and each k , let $x_i^{(k)}$ be the number of individuals from the i th class in the k th subsample, called the *frequency* of the i th class in the k th subsample, $i = 1, 2, \dots, N$, $k = 1, 2, \dots, K$. Each $x_i^{(k)}$ is assumed to be an independent Poisson random variable with mean $\lambda_i^{(k)}$, where $\lambda_i^{(k)}$ is the population abundance of the i th class in the k th population. For each fixed k , the $x_i^{(k)}$'s constitute a *Poisson sample*. Together, the $x_i^{(k)}$'s constitute a *multiple Poisson sample*.

We next consider incidence-based subsamples. Suppose for the k th population, there are $M^{(k)}$ observing occasions. At each occasion, each class is present or absent independently. Let $y_{i,j}^{(k)}$ be an indicator variable of presence, that is, $y_{i,j}^{(k)} = 1$ if the i th class is present on the j th occasion in the k th population, else $y_{i,j}^{(k)} = 0$, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, M^{(k)}$, $k = 1,$

2, ... K. For each fixed pair of (j, k) , the $y_{i,j}^{(k)}$'s constitute a *Bernoulli sample*. For each fixed k , the $y_{i,j}^{(k)}$'s constitute a *multiple Bernoulli sample*.

Let

$$p_{i,j}^{(k)} = P(y_{i,j}^{(k)} = 1) \quad \text{and} \quad x_i^{(k)} = \sum_{j=1}^{M^{(k)}} y_{i,j}^{(k)}.$$

That is, $x_i^{(k)}$ is the number of occasions when the i th class is present in the k th subsample, called the *frequency* of the i th class in the k th subsample.

For each fixed pair of (i, k) , if we further assume that

$$p_{i,j}^{(k)} = p_i^{(k)}, \quad j = 1, 2, \dots, M^{(k)},$$

then each $x_i^{(k)}$ is an independent binomial random variable with success probability $p_i^{(k)}$ and index $M^{(k)}$. For each fixed k , the $x_i^{(k)}$'s constitute a *binomial sample*. Together, the $x_i^{(k)}$'s constitute a *multiple binomial sample*. The odds ratio $\lambda_i^{(k)} = p_i^{(k)} / (1 - p_i^{(k)})$ is also used to represent a binomial distribution.

Note that for the multiple binomial sample, if all $M^{(k)}$'s equal 1, the $x_i^{(k)}$'s, which are identical to the $y_{i,1}^{(k)}$'s now, constitute a *multiple Bernoulli sample*. This special case has been studied thoroughly in the literature, see Chao (2001).

Let

$$\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(K)})'.$$

This is called the *frequency-profile* of the i th class in the whole sample.

The complete data form a matrix as follows:

$$\begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(K)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(K)} \\ \vdots & \vdots & \ddots & \vdots \\ x_N^{(1)} & x_N^{(2)} & \dots & x_N^{(K)} \end{bmatrix}.$$

For each k , let $\Omega^{(k)}$ be the sample space for frequencies in the k th subsample, $k = 1, 2, \dots, K$. Let Ω be the Cartesian product of these sample spaces, that is, $\Omega = \prod_{k=1}^K \Omega^{(k)}$, which is the sample space for frequency-profiles.

For each $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_K)'$ in Ω , let $n_{\boldsymbol{\eta}}$ be the number of classes that occur with frequency-profile $\boldsymbol{\eta}$, called the *frequency-profile counts*. Let $\mathbf{0}$ be the origin of \mathcal{R}^K , that is $\mathbf{0} = (0, 0, \dots, 0)'$. Note that $n_{\mathbf{0}}$ is the number of classes that are absent in all subsamples. Let n be the number of classes present in at least one of these subsamples. It follows that $N = n + n_{\mathbf{0}}$. Finally, n is observed but $n_{\mathbf{0}}$ is not, so estimating N is equivalent to predicting $n_{\mathbf{0}}$.

The degenerate Poisson or binomial distribution, with $\lambda_i^{(k)} = 0$, is allowed. Note that if $x_i^{(k)}$ follows a degenerate distribution, it means the i th class is absent in the k th population.

Mao and Lindsay (2001) developed a sequence of moment-based nonparametric estimators for the number of distinct classes in one population. A special case of the estimators in Mao and Lindsay (2001) can be obtained by linear extrapolation of a log ratio function. These methodologies will be extended to multiple populations in this paper. Both the multiple Poisson sample and the multiple binomial sample are addressed from mixture model point of view. The estimators are nonparametric because a joint latent distribution is assumed, but not modeled directly.

In Section 2, we develop mixture models for both the multiple Poisson sample and the multiple binomial sample. It will be shown that the two mixture models have the same mathematical structure, as seen in Mao and Lindsay (2001) for single populations. In Section 3, we separate the likelihood of observed data and show that there are two parameter of interest, the parameter of interest N and a nuisance parameter. We point out that it is possible to obtain pseudo maximum likelihood estimators for N . In Section 4, a collection of estimators by linear extrapolation are developed. For the special case, the multiple Bernoulli sample, we develop new estimators from linear extrapolation and also show that several estimators in the literature are linear extrapolation estimators. In Section 5, we give a review of estimation of one-population models in Mao and Lindsay (2001). Then we introduce an easy method for the multiple

Poisson sample, that is, combining the subsamples and then applying the estimation procedure in a one-population Poisson model, which is based on the reproductivity of Poisson distributions. We next introduce methods to form moment sequences of finite measures on the positive half of the real line \mathcal{R}^+ from the multivariate mixture densities and develop estimators based on the moment sequences. It turns out that for the multiple Poisson sample, the moment-based procedure yields the same estimators as the subsample combination. Both linear extrapolation estimators and moment-based estimators are pseudo maximum likelihood estimators with the nuisance estimator estimated by moment methods. In Section 6, we study three real datasets to illustrate the procedures. In the first example, the number of students infected by Hepatitis A virus in a college is estimated. The second is a bird census dataset, from which the number of bird species that appeared in at least one of two communities is estimated. The last dataset is from genomic research. We will estimate the number of total genes expressed in at least one of two tissues of tomatoes.

2 THE MIXTURE MODELS

We first consider the representations of the Poisson density and the binomial density. Let x be a Poisson random variable given the mean λ or binomial random variable given the odds ratio λ . That is,

$$x|\lambda \sim f(x; \lambda),$$

where $f(x; \lambda)$ is a Poisson density with mean λ or a binomial density with index M and odds ratio λ . The density $f(x; \lambda)$ with respect to counting measure on \mathcal{R} , takes the following form:

$$f(x; \lambda) = c(\lambda)h(x)\lambda^x, \quad x \in \Omega, \quad \lambda \in [0, +\infty).$$

Here, $c(\lambda)$ is the normalization function and Ω is the sample space. The representation arises by letting

$$\Omega = \{0, 1, 2, \dots\} \quad (\text{Poisson}) \quad \text{or} \quad \Omega = \{0, 1, 2, \dots, M\} \quad (\text{binomial}),$$

$$[c(\lambda)]^{-1} = \exp(-\lambda) \quad (\text{Poisson}) \quad \text{or} \quad [c(\lambda)]^{-1} = [1 + \lambda]^{-M} \quad (\text{binomial}),$$

and

$$h(x) = \frac{I(x \in \Omega)}{x!} \quad (\text{Poisson}) \quad \text{or} \quad h(x) = \frac{M! I(x \in \Omega)}{(M-x)! x!} \quad (\text{binomial}).$$

where $I(E)$ is the indicator function of the event E .

Let

$$\boldsymbol{\lambda} = (\lambda^{(1)}, \lambda^{(2)}, \dots, \lambda^{(K)})'.$$

The density of a frequency-profile \mathbf{x} for one class is given by

$$f(\mathbf{x}; \boldsymbol{\lambda}) = \prod_{k=1}^K f(x^{(k)}; \lambda^{(k)}) = \mathbf{h}(\mathbf{x}) \prod_{k=1}^K c(\lambda^{(k)}) [\lambda^{(k)}]^{x^{(k)}}$$

with respect to counting measure on \mathcal{R}^K , where

$$\mathbf{h}(\mathbf{x}) = \prod_{k=1}^K h(x^{(k)}).$$

To form a mixture model, we assume that $\boldsymbol{\lambda}$ is a random vector from a K -variate latent distribution $\mathbf{Q}(\boldsymbol{\lambda})$ with no mass on the origin. That is,

$$\boldsymbol{\lambda} \sim \mathbf{Q}(\boldsymbol{\lambda}) \quad \text{and} \quad P_{\mathbf{Q}}(\{\boldsymbol{\lambda} = \mathbf{0}\}) = 0.$$

Conditioning on $\boldsymbol{\lambda}$, the frequency profile \mathbf{x} is a multivariate Poisson or multivariate binomial random vector. The marginal density of \mathbf{x} is called a *mixture density*, denoted by $f(\mathbf{x}, \mathbf{Q})$, and defined by

$$f(\mathbf{x}; \mathbf{Q}) = E_{\mathbf{Q}} f(\mathbf{x}; \boldsymbol{\lambda}) = \int f(\mathbf{x}; \boldsymbol{\lambda}) d\mathbf{Q}(\boldsymbol{\lambda}).$$

The joint density of all frequency-profiles is given by

$$\prod_{i=1}^N f(\mathbf{x}_i; \mathbf{Q}) = \prod_{\boldsymbol{\eta} \in \Omega} f(\boldsymbol{\eta}; \mathbf{Q})^{n_{\boldsymbol{\eta}}}.$$

The resulting joint marginal density of the frequency-profile counts $\{n_{\boldsymbol{\eta}}\}_{\boldsymbol{\eta} \in \Omega}$ is given by

$$\frac{N!}{\prod_{\boldsymbol{\eta} \in \Omega} n_{\boldsymbol{\eta}}!} \prod_{\boldsymbol{\eta} \in \Omega} f(\boldsymbol{\eta}; \mathbf{Q})^{n_{\boldsymbol{\eta}}}. \quad (1)$$

That is, $\{n_{\boldsymbol{\eta}}\}_{\boldsymbol{\eta} \in \Omega}$ arise from a multinomial distribution with index N and cell probabilities $f(\boldsymbol{\eta}; \mathbf{Q})$'s.

Remark 1 *It is important to note that by allowing arbitrary \mathbf{Q} , we are allowing dependence between the components of $\boldsymbol{\lambda}$. Thus, although the frequencies $x_i^{(k)}$ and $x_i^{(l)}$ in two populations (k, l) are conditionally independent given the population abundances $\lambda_i^{(k)}$ and $\lambda_i^{(l)}$, the pair $(\lambda_i^{(k)}, \lambda_i^{(l)})$ can be correlated. This might be expected. For example, if a bird census dataset was collected from two adjacent communities, we expect the abundances of a bird species in the two communities to be similar.*

3 PSEUDO MAXIMUM LIKELIHOOD ESTIMATION

Estimating N and \mathbf{Q} simultaneously seems very difficult. A simpler approach is to separate the estimation of N and that of \mathbf{Q} , which leads to the pseudo maximum likelihood estimation.

Let

$$\boldsymbol{\Omega}_0 = \boldsymbol{\Omega} - \{\mathbf{0}\}.$$

We have given the “complete” data likelihood for $\{n_\eta : n_\eta \in \boldsymbol{\Omega}\}$ in (1). Since n_0 is not observed, the observed data likelihood is

$$\frac{N!}{(N-n)! \prod_{\eta \in \boldsymbol{\Omega}_0} n_\eta!} [f(\mathbf{0}; \mathbf{Q})]^{N-n_0} \prod_{\eta \in \boldsymbol{\Omega}_0} f(\eta; \mathbf{Q})^{n_\eta}. \quad (2)$$

The observed likelihood in (2) can be further factored into a binomial density for n times a multinomial density for $\{n_\eta\}_{\eta \in \boldsymbol{\Omega}_0}$ given n . The density of $\{n_\eta\}_{\eta \in \boldsymbol{\Omega}_0}$ given n can be written as

$$\frac{n!}{\prod_{\eta \in \boldsymbol{\Omega}_0} n_\eta!} \prod_{\eta \in \boldsymbol{\Omega}_0} \left[\frac{f(\eta; \mathbf{Q})}{1 - f(\mathbf{0}; \mathbf{Q})} \right]^{n_\eta}. \quad (3)$$

The density of n is given by

$$\frac{N!}{n!(N-n)!} f(\mathbf{0}; \mathbf{Q})^{N-n} [1 - f(\mathbf{0}; \mathbf{Q})]^n. \quad (4)$$

Note that there are two parameters in (4), the parameter of interest N and a nuisance parameter $f(\mathbf{0}; \mathbf{Q})$. Let ϑ be the corresponding odds ratio,

that is, $\vartheta = f(\mathbf{0}; \mathbf{Q})/[1 - f(\mathbf{0}; \mathbf{Q})]$. The density of n given in (4) can be expressed in terms of ϑ , denoted by $L(N; \vartheta)$, and written as

$$L(N; \vartheta) = \frac{N!}{n!(N-n)!} \left(\frac{\vartheta}{\vartheta+1} \right)^{N-n} \left(\frac{1}{\vartheta+1} \right)^n .$$

When ϑ is known, $L(N; \vartheta)$ can be used as a likelihood to infer N . From Lindsay and Roeder (1987), the maximum likelihood estimator for N , is given by

$$\hat{N} = n + n\vartheta. \tag{5}$$

If we substitute an estimator $\hat{\vartheta}$ for ϑ into (5), we obtain an estimator for N , called a *pseudo-likelihood estimator* (Gong and Samaniego 1981). Note that since N is an integer, technically, we should use $[\hat{N}]$ as an estimator for N , where $[\hat{N}]$ represents the largest integer no greater than \hat{N} (Lindsay and Roeder 1987).

In the following sections, we will find moment estimators for the nuisance parameter ϑ . That is, the methods in this paper give pseudo-likelihood estimators for N with the nuisance parameter ϑ estimated by moment methods.

4 LINEAR EXTRAPOLATION ESTIMATORS

In the one-population models described by Mao and Lindsay (2001), a log ratio function was defined to be ratio of the Poisson mixture density or the binomial mixture density to $h(x)$. The empirical version of the log ratio function can be plotted in order to detect heterogeneity. An estimator was also derived from the plot by linear extrapolation. We extend these ideas to a multivariate log ratio function. Although it is hard to plot such a function, it can be used to develop estimators through linear extrapolation.

4.1 The estimators

Let $R(\boldsymbol{\eta})$ be the *multivariate ratio function* defined by

$$R(\boldsymbol{\eta}) = f(\boldsymbol{\eta}; \mathbf{Q})/\mathbf{h}(\boldsymbol{\eta})$$

Proposition 1 *The function $\log R(\boldsymbol{\eta})$ is convex in $\boldsymbol{\eta}$ and it is linear if and only if \mathbf{Q} is degenerate at one point.*

See the appendix for the proof.

Since N is unknown, we define a translated version of $\log R(\boldsymbol{\eta})$ as follows:

$$H(\boldsymbol{\eta}) = \log N + \log R(\boldsymbol{\eta}) = \log[Nf(\boldsymbol{\eta}; \mathbf{Q})/\mathbf{h}(\boldsymbol{\eta})]. \quad (6)$$

Proposition 2 *The function $H(\boldsymbol{\eta})$ is convex in $\boldsymbol{\eta}$ and it is linear if and only if \mathbf{Q} is degenerate at one point.*

Thus the plot $(\boldsymbol{\eta}, H(\boldsymbol{\eta}))$ represents a convex surface when \mathbf{Q} is not degenerate.

Suppose $n_{\boldsymbol{\eta}} \neq 0$. If we replace $f(\boldsymbol{\eta}; \mathbf{Q})$ with its unbiased “estimator” $n_{\boldsymbol{\eta}}/N$ in (6), we have an “estimator” $\hat{H}(\boldsymbol{\eta})$ for $H(\boldsymbol{\eta})$ given by

$$\hat{H}(\boldsymbol{\eta}) = \log n_{\boldsymbol{\eta}}/\mathbf{h}(\boldsymbol{\eta}).$$

Note that $\hat{H}(\mathbf{0})$ is “estimated” by n_0 since $\mathbf{h}(\mathbf{0}) = 1$. Now consider the special case in which \mathbf{Q} is degenerate at some point $\boldsymbol{\lambda}$. We can develop a collection of estimators for N which would be tight for this case and which, by convexity, estimate lower bounds for N in mixture settings.

We can write, in the degenerate case,

$$H(\boldsymbol{\eta}) = \log[N + \sum_{k=1}^K c(\lambda^{(k)})] + \sum_{k=1}^K \log \lambda^{(k)} \times \eta_k = \alpha_0 + \sum_{k=1}^K \alpha_k \eta_k,$$

where

$$\alpha_0 = \log[N + \sum_{k=1}^K c(\lambda^{(k)})] \text{ and } \alpha_k = \log \lambda^{(k)}, \quad k = 1, 2, \dots, K.$$

Let

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_K)'$$

Then

$$H(\boldsymbol{\eta}) = \alpha_0 + \boldsymbol{\alpha}'\boldsymbol{\eta},$$

and in particular,

$$H(\mathbf{0}) = \alpha_0. \quad (7)$$

The plot $(\boldsymbol{\eta}, H(\boldsymbol{\eta}))$ presents a linear surface, which is determined by the $K + 1$ parameters, the α_i 's.

Note that the $H(\boldsymbol{\eta})$'s can be used to express the α_i 's. Then $H(\mathbf{0})$ can be expressed through other $H(\boldsymbol{\eta})$'s in multiple ways. That is, there are many functionals F satisfying

$$H(\mathbf{0}) = F(H(\boldsymbol{\eta}^{(1)}), H(\boldsymbol{\eta}^{(2)}), \dots, H(\boldsymbol{\eta}^{(L)})), \quad (8)$$

for a set of frequency-profiles, $\boldsymbol{\eta}^{(i)}$ in $\boldsymbol{\Omega}_0$, $i = 1, 2, \dots, L$, which can be obtained by solving for the α_k 's in the linear equation system:

$$H(\boldsymbol{\eta}^{(i)}) = \alpha_0 + \boldsymbol{\alpha}'\boldsymbol{\eta}^{(i)}, \quad i = 1, 2, \dots, L.$$

Suppose $n_{\boldsymbol{\eta}^{(i)}} \neq 0$, $i = 1, 2, \dots, L$, and $n_0 \neq 0$. Plugging these estimators in (8), we obtain an empirical version as follows:

$$\hat{H}(\mathbf{0}) = F(\hat{H}(\boldsymbol{\eta}^{(1)}), \hat{H}(\boldsymbol{\eta}^{(2)}), \dots, \hat{H}(\boldsymbol{\eta}^{(L)})). \quad (9)$$

We can obtain a predictor for n_0 from (9) given by

$$\tilde{n}_0 = \exp[F(\hat{H}(\boldsymbol{\eta}^{(1)}), \hat{H}(\boldsymbol{\eta}^{(2)}), \dots, \hat{H}(\boldsymbol{\eta}^{(L)}))],$$

which yields an estimator for N as follows:

$$\hat{N}_F = n + \tilde{n}_0 = n + \exp[F(\hat{H}(\boldsymbol{\eta}^{(1)}), \hat{H}(\boldsymbol{\eta}^{(2)}), \dots, \hat{H}(\boldsymbol{\eta}^{(L)}))].$$

The subscript F in \hat{N}_F represents the specific functional form F . Such estimators will be called *linear extrapolation estimators*. There exists a collection of linear extrapolation estimators, which depend on how one expresses $H(\mathbf{0})$ through other $H(\boldsymbol{\eta})$'s.

Note that these estimators are consistent for N when \mathbf{Q} is degenerate at a point $\boldsymbol{\lambda}$. In the general mixture model, the estimator \hat{N}_F actually is a consistent estimator for the parameter N_F given by

$$N_F = N(1 - f(\mathbf{0}; \mathbf{Q}) + \exp[F(H(\boldsymbol{\eta}^{(1)}), H(\boldsymbol{\eta}^{(2)}), \dots, H(\boldsymbol{\eta}^{(L)}))]).$$

Usually $N_F/N < 1$ when \mathbf{Q} is not degenerate at a specific point; it is a function of $f(\mathbf{0}; \mathbf{Q})$ and the $f(\boldsymbol{\eta}^{(i)}; \mathbf{Q})$'s which does not depend on N .

Our motivation for using linear extrapolation estimators needs some elaboration. When \mathbf{Q} is degenerate at $\boldsymbol{\lambda}$, the maximum likelihood estimator for N can be easily derived. The maximum likelihood estimator for N should be less variable than a linear extrapolation estimator if \mathbf{Q} is degenerate at $\boldsymbol{\lambda}$. However, the maximum likelihood estimator is negatively biased when \mathbf{Q} is not degenerate at a point. Extrapolating $H(\mathbf{0})$ using the log ratio function can generate estimators with smaller bias; see Mao and Lindsay (2001).

It is clear that the linear extrapolation estimators are also pseudo-maximum likelihood estimators. The nuisance parameter ϑ is estimated by $\hat{\vartheta}_F$, where

$$\hat{\vartheta}_F = n^{-1} \exp[F(\hat{H}(\boldsymbol{\eta}^{(1)}), \hat{H}(\boldsymbol{\eta}^{(2)}), \dots, \hat{H}(\boldsymbol{\eta}^{(L)}))].$$

Finally, we point out that whether there exists the “best” one among all linear extrapolation estimators according to some optimality requirements is still a subject under investigation.

4.2 Confidence inference

Let ξ_F denote N_F/N , which is assumed to be independent of N . Let

$$\mathbf{f}_F = (f(\mathbf{0}; \mathbf{Q}), f(\boldsymbol{\eta}^{(1)}; \mathbf{Q})', \dots, f(\boldsymbol{\eta}^{(L)}; \mathbf{Q})').$$

Then we have

$$\xi_F = \xi_F(\mathbf{f}_F).$$

Let

$$\sigma_F^2 = \frac{\partial \xi_F}{\partial \mathbf{f}_F'} (\text{diag} \mathbf{f}_F - \mathbf{f}_F \mathbf{f}_F') \frac{\partial \xi_F}{\partial \mathbf{f}_F}.$$

Then

$$(N\sigma_F^2)^{-1/2}(\widehat{N}_F - N_F) \xrightarrow{d} N(0, 1),$$

and

$$(N\xi_F^2/\sigma_F^2)^{1/2}(\log \widehat{N}_F - \log N_F) \xrightarrow{d} N(0, 1),$$

as N goes to infinity. It is clear that the width of the confidence interval for N_F is proportional to \sqrt{N} while only the width of the confidence interval for $\log N_F$ shrinks at standard rate $1/\sqrt{N}$.

If \widehat{N}_F is tight for N , then ξ_F and σ_F^2 can be consistently estimated and we can construct confidence intervals based on the asymptotic results. When \widehat{N}_F is a lower bound for N , then ξ_F and σ_F^2 cannot be consistently estimated. Such confidence intervals may still be used but the actual coverage for N is questionable. However, the lower confidence limits for N_F are necessarily conservative lower limits for N .

As an alternative approach, bootstrap re-sampling can be used to get confidence intervals. Consider the multinomial distribution for all frequency-profiles counts in (1). If it is estimated, we can take bootstrap re-samples from the estimated multinomial distribution. For the linear extrapolation estimator, we consider a collapsed multinomial distribution.

Let

$$n_+ = N - n_0 - \sum_{i=1}^L n_{\eta^{(i)}}.$$

and

$$f(+; \mathbf{Q}) = 1 - f(0; \mathbf{Q}) - \sum_{i=1}^L f(\eta^{(i)}; \mathbf{Q}).$$

Then $\{n_0, n_+, n_{\eta^{(i)}}, i = 1, 2, \dots, L\}$ can be regarded as a random sample from a multinomial distribution with index N and cell probabilities $f(0; \mathbf{Q})$, $f(+; \mathbf{Q})$ and $f(\eta^{(i)}; \mathbf{Q})$, $i = 1, 2, \dots, L$. Suppose we have obtained the estimator \widehat{N} for N , say \widehat{N}_F here. The estimators for the cell probabilities are given by $\widehat{f}(0; \mathbf{Q}) = 1 - n/\widehat{N}$, $\widehat{f}(+; \mathbf{Q}) = n_+/\widehat{N}$, and $\widehat{f}(\eta^{(i)}; \mathbf{Q}) = n_{\eta^{(i)}}/\widehat{N}$, $i = 1, 2, \dots, L$.

4.3 An example

We define a linear extrapolation estimator for illustration. Let \mathbf{e}_k be k th standard basis vector of \mathcal{R}^K , $k = 1, 2, \dots, K$. That is, the k th entry in \mathbf{e}_k equals 1 and all other entries are zero. Let $\mathbf{1} = (1, 1, \dots, 1)'$ in \mathcal{R}^K . It is clear that

$$H(\mathbf{1}) = \alpha_0 + \sum_{k=1}^K \alpha_k \text{ and } H(\mathbf{e}_k) = \alpha_0 + \alpha_k, \quad k = 1, 2, \dots, K. \quad (10)$$

From (7) and (10), we have

$$H(\mathbf{0}) = (K - 1)^{-1} \left[\sum_{k=1}^K H(\mathbf{e}_k) - H(\mathbf{1}) \right]. \quad (11)$$

The empirical version of (11) is given by

$$\hat{H}(\mathbf{0}) = (K - 1)^{-1} \left[\sum_{k=1}^K \hat{H}(\mathbf{e}_k) - \hat{H}(\mathbf{1}) \right],$$

which yields a predictor for \tilde{n}_0 as

$$\tilde{n}_0 = [n_1^{-1} \prod_{k=1}^K n_{\mathbf{e}_k}]^{1/(K-1)}.$$

The estimator for N is denoted by $\hat{N}_F = \hat{N}_e$ and is given by

$$\hat{N}_e = n + [n_1^{-1} \prod_{k=1}^K n_{\mathbf{e}_k}]^{1/(K-1)}.$$

The corresponding ξ_F and σ_F^2 are denoted by ξ_e and σ_e^2 respectively.

Let

$$\theta_e = [f(\mathbf{1}; \mathbf{Q})^{-1} \prod_{k=1}^K f(\mathbf{e}_k; \mathbf{Q})]^{1/(K-1)}.$$

Then

$$\xi_e = 1 - f(\mathbf{0}; \mathbf{Q}) + \theta_e,$$

and

$$\sigma_e^2 = f(\mathbf{0}; \mathbf{Q}) + [\theta/(K - 1)]^2 [1/f(\mathbf{1}; \mathbf{Q}) + \sum_{k=1}^K 1/f(\mathbf{e}_k; \mathbf{Q})] - [\theta_e - f(\mathbf{0}; \mathbf{Q})]^2.$$

4.4 The multiple Bernoulli sample

We will focus on the multiple Bernoulli sample, the special case of the multiple binomial sample with all $M^{(k)}$ equal one. The models for this special case is the same as the models used for population size estimation, in which K subsamples are taken from a closed population with unknown size N . The individuals are indexed by $1, 2, \dots, N$. The presence or absence information of each individual is recorded in each subsample, that is, $x_i^{(k)}$ is the presence indicator of the i th individual in the k th subsample. For example, the subsamples may be obtained from capture-recapture experiment in ecological studies or multiple census lists in epidemiological studies or population census.

We first develop some new linear extrapolation estimators using the frequency-profiles near the origin and then show that several estimators in the literature can be obtained by linear extrapolation quite easily.

Note that now all frequency-profiles are composed of 1's and 0's. For $k \neq l$, let \mathbf{e}_{kl} equal $\mathbf{e}_k + \mathbf{e}_l$. There are K frequency profiles with only one 1 as their components (the \mathbf{e}_k 's) and $K(K-1)/2$ frequency profiles with exactly two 1's as their components (the \mathbf{e}_{kl} 's). We will use the $H(\boldsymbol{\eta})$'s with $\boldsymbol{\eta}$ being among the \mathbf{e}_k 's and \mathbf{e}_{kl} 's to express $H(\mathbf{0})$. It is clear that

$$H(\mathbf{e}_{kl}) = \alpha_0 + \alpha_k + \alpha_l, \quad k, l = 1, 2, \dots, K, k \neq l. \quad (12)$$

Then

$$\sum_{1 \leq k < l \leq K} H(\mathbf{e}_{kl}) = [K(K-1)/2]\alpha_0 + (K-1) \sum_{k=1}^K \alpha_k.$$

From (7), (10) and (12), we obtain

$$H(\mathbf{0}) = 2\{K^{-1} \sum_{k=1}^K H(\mathbf{e}_k) - [K(K-1)]^{-1} \times \sum_{1 \leq k < l \leq K} H(\mathbf{e}_{kl})\}. \quad (13)$$

Note that (13) and (11) are the same for a multiple Bernoulli sample when $K = 2$.

Suppose all of the $n_{\mathbf{e}_k}$'s and the $n_{\mathbf{e}_{kl}}$'s are not zero. The empirical

version of (13) is given by

$$\widehat{H}(\mathbf{0}) = 2\{K^{-1} \sum_{k=1}^K \widehat{H}(\mathbf{e}_k) - [K(K-1)]^{-1} \times \sum_{1 \leq k < l \leq K} \widehat{H}(\mathbf{e}_{kl})\},$$

which yields

$$\tilde{n}_0 = \prod_{k=1}^K n_{\mathbf{e}_k}^{2/K} \times \prod_{1 \leq k < l \leq K} n_{\mathbf{e}_{kl}}^{-2/[K(K-1)]}.$$

The estimator is denoted by \widehat{N}_t and given by

$$\widehat{N}_t = n + \prod_{k=1}^K n_{\mathbf{e}_k}^{2/K} \times \prod_{1 \leq k < l \leq K} n_{\mathbf{e}_{kl}}^{-2/[K(K-1)]}.$$

The corresponding ξ_F and σ_F^2 are denoted by ξ_t and σ_t^2 respectively.

Let

$$\theta_t = \prod_{k=1}^K [f(\mathbf{e}_k; \mathbf{Q})]^{2/K} \times \prod_{1 \leq k < l \leq K} [f(\mathbf{e}_{kl}; \mathbf{Q})]^{-2/[K(K-1)]}.$$

Then

$$\xi_t = 1 - f(\mathbf{0}; \mathbf{Q}) + \theta_t,$$

and

$$\begin{aligned} \sigma_t^2 &= (2\theta_t/K)^2 \left[\sum_{k=1}^K 1/f(\mathbf{e}_k; \mathbf{Q}) + 1/(K-1)^2 \times \sum_{1 \leq k < l \leq K} 1/f(\mathbf{e}_{kl}; \mathbf{Q}) \right] \\ &+ f(\mathbf{0}; \mathbf{Q}) - [\theta_t - f(\mathbf{0}; \mathbf{Q})]^2. \end{aligned}$$

Our next goal is to show that some estimators in the literature are linear extrapolation estimators although the original methodologies from which they have been developed are different.

When $K = 2$, the Lincoln-Peterson estimator is well-known (Peterson 1896 and Lincoln 1930), which is given by

$$\widehat{N}_{lp} = n + \frac{n_{(1,0)}n_{(0,1)}}{n_{(1,1)}}.$$

The Lincoln-Peterson estimator is also called the *dual-system estimator* and used for census undercount adjustment (Wolter 1986 and Hogan 1993). The Lincoln-Peterson estimator is a linear extrapolation estimator, which can be obtained from the following equation:

$$H((0, 0)) = H((1, 0)) + H((0, 1)) - H((1, 1)). \quad (14)$$

Note that the Lincoln-Peterson estimator is a special case of the linear extrapolation estimators \widehat{N}_t and \widehat{N}_e because (13) and (11) are identical to (14) for two Bernoulli subsamples.

Bishop, Fienberg and Holland (1975) gave three estimators for N when $K = 3$. Let \widehat{N}_{bfh1} , \widehat{N}_{bfh2} and \widehat{N}_{bfh3} be the first estimator, the second and the third respectively, which are given by

$$\widehat{N}_{bfh1} = n + \frac{n_{(0,0,1)} \times [n_{(0,1,0)} + n_{(1,1,0)} + n_{(1,0,0)}]}{n_{(1,0,1)} + n_{(1,1,1)} + n_{(0,1,1)}},$$

$$\widehat{N}_{bfh2} = n + \frac{n_{(0,0,1)} n_{(1,0,0)}}{n_{(1,0,1)}},$$

and

$$\widehat{N}_{bfh3} = n + \frac{n_{(1,1,1)} n_{(1,0,0)} n_{(0,1,0)} n_{(0,0,1)}}{n_{(1,0,1)} n_{(1,1,0)} n_{(0,1,1)}}.$$

These estimators are linear extrapolation estimators, which can be derived from the following equations respectively:

$$H((0, 0, 0)) = H((0, 0, 1)) + \log \frac{e^{H((0,1,0))} + e^{H((1,1,0))} + e^{H((1,0,0))}}{e^{H((1,0,1))} + e^{H((1,1,1))} + e^{H((0,1,1))}},$$

$$H((0, 0, 0)) = H((0, 0, 1)) + H((1, 0, 0)) - H((1, 0, 1)),$$

and

$$\begin{aligned} H((0, 0, 0)) &= H((1, 1, 1)) + H((1, 0, 0)) + H((0, 1, 0)) + H((0, 0, 1)) \\ &- H((1, 1, 0)) - H((1, 0, 1)) - H((0, 1, 1)). \end{aligned} \quad (15)$$

The estimator \widehat{N}_{bfh2} is called a *triple system estimator* in Darroch et al. (1993) where it was used for census undercount adjustment. The estimator \widehat{N}_{bfh1} is also a triple system estimator given by Zaslavsky and Wolfgang (1990). Besides \widehat{N}_{bfh1} and \widehat{N}_{bfh2} , Zaslavsky and Wolfgang (1990) gave several other triple system estimators. For example, \widehat{N}_{zw1} , \widehat{N}_{zw2} , \widehat{N}_{zw3} and \widehat{N}_{zw4} were four estimators developed in Zaslavsky and Wolfgang (1990), which can be written as follows:

$$\widehat{N}_{zw1} = n + \frac{n_{(0,0,1)} \times [n_{(1,0,0)} + n_{(0,1,0)}]}{n_{(1,0,1)} + n_{(0,1,1)}},$$

$$\widehat{N}_{zw2} = n + \frac{n_{(0,1,1)} n_{(1,0,1)} n_{(0,1,0)} n_{(1,0,0)}}{n_{(0,0,1)} n_{(1,1,1)} n_{(1,1,0)}},$$

$$\widehat{N}_{zw3} = n + \frac{n_{(1,0,0)}[n_{(1,0,1)} + n_{(0,0,1)}n_{(1,0,0)}]}{n_{(1,0,1)} + n_{(1,1,1)}n_{(1,1,0)}},$$

and

$$\widehat{N}_{zw4} = n - n_{(0,0,1)} + \frac{n_{(0,1,1)}n_{(1,0,1)} [n_{(1,0,1)} + n_{(1,0,0)}][n_{(0,1,1)} + n_{(0,1,0)}]}{n_{(1,1,1)}n_{(0,0,1)} n_{(1,1,1)} + n_{(1,1,0)}}.$$

The estimators \widehat{N}_{zw1} , \widehat{N}_{zw3} , \widehat{N}_{zw2} and \widehat{N}_{zw4} can be obtained from the following four equations respectively:

$$H((0, 0, 0)) = H((0, 0, 1)) + \log \frac{e^{H((1,0,0))} + e^{H((0,1,0))}}{e^{H((1,0,1))} + e^{H((0,1,1))}},$$

$$\begin{aligned} H((0, 0, 0)) &= H((0, 1, 0)) + H((1, 0, 1)) + H((0, 1, 1)) + H((1, 0, 0)) \\ &\quad - H((0, 0, 1)) - H((1, 1, 0)) - H((1, 1, 1)), \end{aligned}$$

$$H((0, 0, 0)) = H((1, 0, 0)) + \frac{e^{H((1,0,1))} + e^{H((1,0,0))} + e^{H((0,0,1))}}{e^{H((1,1,0))} + e^{H((1,0,1))} + e^{H((1,1,1))}},$$

and

$$\begin{aligned} e^{H((0,0,0))} &= -e^{H((0,0,1))} + \frac{e^{[H((0,1,1))+H((1,0,1))]}]}{e^{[H((1,1,1))+H((0,0,1))]}]} \\ &\quad \times \frac{[e^{H((1,0,1))} + e^{H((1,0,0))}][e^{H((0,1,1))} + e^{H((0,1,0))}]}{e^{H((1,1,1))} + e^{H((1,1,0))}}. \end{aligned}$$

So \widehat{N}_{zw1} , \widehat{N}_{zw2} , \widehat{N}_{zw3} and \widehat{N}_{zw4} are also linear extrapolation estimators.

Fienberg (1972) developed an estimator for general K . Let

$$|\boldsymbol{\eta}| = \sum_{k=1}^K \eta_k.$$

The estimator in Fienberg (1972) is given by

$$\widehat{N}_f = n + \prod_{\boldsymbol{\eta} \in \Omega_0 \text{ with odd } |\boldsymbol{\eta}|} n_{\boldsymbol{\eta}} \times \prod_{\boldsymbol{\eta} \in \Omega_0 \text{ with even } |\boldsymbol{\eta}|} n_{\boldsymbol{\eta}}^{-1}.$$

It is clear that

$$\begin{aligned} &\sum_{\boldsymbol{\eta} \in \Omega_0 \text{ with odd } |\boldsymbol{\eta}|} H(\boldsymbol{\eta}) - \sum_{\boldsymbol{\eta} \in \Omega_0 \text{ with even } |\boldsymbol{\eta}|} H(\boldsymbol{\eta}) - H(\mathbf{0}) \\ &= \sum_{\boldsymbol{\eta} \in \Omega \text{ with odd } |\boldsymbol{\eta}|} H(\boldsymbol{\eta}) - \sum_{\boldsymbol{\eta} \in \Omega \text{ with even } |\boldsymbol{\eta}|} H(\boldsymbol{\eta}) = 0 \end{aligned}$$

Then we have

$$H(\mathbf{0}) = \sum_{\boldsymbol{\eta} \in \Omega_0 \text{ with odd } |\boldsymbol{\eta}|} H(\boldsymbol{\eta}) - \sum_{\boldsymbol{\eta} \in \Omega_0 \text{ with even } |\boldsymbol{\eta}|} H(\boldsymbol{\eta}), \quad (16)$$

which will yield \widehat{N}_f . Thus, the estimator \widehat{N}_f is a linear extrapolation estimator. Note that $\widehat{N}_f = \widehat{N}_{b_{fh3}}$ for $K = 3$ and $\widehat{N}_f = \widehat{N}_{lp}$ for $K = 2$ since (14) and (15) are special cases of (16).

Finally, we conclude that all linear extrapolation estimators are consistent for the population size N in capture-recapture studies if we assume that capture probabilities vary only among occasions, not among individuals. In fact, such an assumption holds if and only if the latent distribution Q is degenerate at a point.

5 MOMENT-BASED ESTIMATORS

The preceding estimators were all designed to be consistent estimators of N when \mathbf{Q} is degenerate at a point but only removed some bias when \mathbf{Q} is not degenerate at a point. We now turn to develop estimators for the general mixture case.

5.1 Review of estimation in one population

The one-population model in Mao and Lindsay (2001) will play a significant role in our procedures. We briefly review the estimation of the number of classes in one population, which is the special case in which K equals one. We can simplify notations by omitting the index for populations. Then the x_i 's are the frequency for N classes the n_x 's are the frequency counts, and Q will be the latent distribution. Note that the Poisson distribution and the binomial distribution cannot be degenerate and Q can not put mass on zero.

To develop the estimators for the case in which Q is not degenerate, Mao and Lindsay (2001) developed a mathematical result about approximation of the total mass of a finite measure on \mathcal{R}^+ through its higher moments. Let ν be a finite measure on \mathcal{R}^+ and let $\mu(x)$ be its x th moment, that is,

$$\mu(x) = \int \lambda^x d\nu(\lambda), \quad x = 0, 1, \dots$$

Let $\mathcal{M} = \{\mu(x)\}_{x=0}^{+\infty}$ be the set of all moments, called a *moment sequence*.

Suppose ν has T support points where T is a natural number or infinite.

For any natural number p , let

$$\mathbf{b}_p(\mathcal{M}) = (\mu(1), \mu(2), \dots, \mu(p))',$$

and

$$A_p(\mathcal{M}) = \begin{bmatrix} \mu(2) & \mu(3) & \cdots & \mu(p+1) \\ \mu(3) & \mu(4) & \cdots & \mu(p+2) \\ \vdots & \vdots & \ddots & \vdots \\ \mu(p+1) & \mu(p+2) & \cdots & \mu(2p) \end{bmatrix}.$$

For notational simplicity, set

$$A_p = A_p(\mathcal{M}), \quad \mathbf{b}_p = \mathbf{b}_p(\mathcal{M}),$$

when \mathcal{M} is the moment sequence of ν . For each p , $0 < p$, and $p \leq T$ when T is finite, let

$$\mu_p = \mathbf{b}_p' [A_p]^{-1} \mathbf{b}_p,$$

called a *p*th order lower bound for $\mu(0)$, which is the unique solution to the equation

$$|H_p(y)| = 0 \quad \text{where} \quad H_p(y) = \begin{pmatrix} y & \mathbf{b}_p' \\ \mathbf{b}_p & A_p \end{pmatrix}.$$

It was shown in Mao and Lindsay (2001) that the sequence of the μ_p 's is monotonically increasing when p increases. Additionally, when T is finite, $\mu_T = \mu(0)$. When T is infinite and the moment generating function of ν exists, the μ_p 's converges to $\mu(0)$. That is, the μ_p 's constitute a lower bound sequence of $\mu(0)$ when the $\mu(x)$'s are a sequence of moments of ν .

Let

$$d\nu(\lambda) = c(\lambda)dQ(\lambda).$$

Then

$$\mu(x) = f(x; Q)/h(x),$$

which can be "estimated" by

$$\hat{\mu}(x) = n_x/[Nh(x)].$$

The $\hat{\mu}(x)$'s constitute an empirical version of moment sequence \mathcal{M} , denoted by $\widehat{\mathcal{M}}$. Let

$$\widehat{A}_p = A_p(\widehat{\mathcal{M}}) \text{ and } \widehat{\mathbf{b}}_p = \mathbf{b}_p(\widehat{\mathcal{M}}).$$

When \widehat{A}_p is positive definite, let

$$\hat{\mu}_p = \widehat{\mathbf{b}}_p'[\widehat{A}_p]^{-1}\widehat{\mathbf{b}}_p.$$

The $\hat{\mu}_p$'s constitute a lower bound sequence for $\hat{\mu}(0) = n_0/N$. Thus the $N\hat{\mu}_p$'s can be regarded as a predictor sequence for $N\hat{\mu}(0) = n_0$ and the $n + N\hat{\mu}_p$'s can be regarded as an estimator sequence for N . The estimator $n + N\hat{\mu}_1$ is the linear extrapolation estimator. Note that in the $N\hat{\mu}_p$'s, N is cancelled out. For detailed description, readers are referred to Mao and Lindsay (2001).

5.2 Subsample combination

We start from a method only for the multiple Poisson sample, which is based on the well-known reproductivity of Poisson distributions. Recall that conditionally, each $x_i^{(k)}$ is a Poisson random variable with mean $\lambda_i^{(k)}$, $k = 1, 2, \dots, K$. Let

$$z_i = \sum_{k=1}^K x_i^{(k)} \text{ and } \lambda_{i,S} = \sum_{k=1}^K \lambda_i^{(k)}, \quad i = 1, 2, \dots, N. \quad (17)$$

We present the reproductivity property as a proposition to emphasize its importance.

Proposition 3 *Conditionally, the random variable z_i is a Poisson random variable with mean $\lambda_{i,S}$.*

Thus the z_i 's constitute a Poisson sample. Regardless of dependence of the $\lambda_i^{(k)}$'s, the $\lambda_{i,S}$'s can be regarded as a random sample arising from a latent distribution Q_S , which can be derived from the multivariate latent distribution \mathbf{Q} . The estimation method for one population can be used for the combined sample, see the review subsection or Mao and Lindsay (2001).

Unfortunately, the binomial distributions do not have the same reproducibility property as the Poisson distributions. However, we can take a rather abstract approach which results in estimators both for the multiple Poisson sample and the multiple binomial sample. Additionally, such an approach is equivalent to the subsample combination for the multiple Poisson sample as we will show later.

5.3 Moment sequence construction

There are two key ideas used in the one-population models (Mao and Lindsay 2001): first, the total mass of a finite measure on \mathcal{R}^+ can be characterized by its higher moments, and secondly, the mixture densities can be used to construct a moment sequence of a finite measure on \mathcal{R}^+ . There is only one natural choice of the moment sequence in the one-population models. In the multiple-population model, we first present a simple approach to construct moment sequences of a measure on \mathcal{R}^+ based on the multivariate mixture densities.

Let

$$\Omega = \begin{cases} \{0, 1, \dots\} & \text{(multiple Poisson)} \\ \{0, 1, \dots, M\} & \text{(multiple binomial)} \end{cases},$$

where $M = \min(M^{(1)}, M^{(2)}, \dots, M^{(K)})$. Let

$$C_{\boldsymbol{\eta}}^{|\boldsymbol{\eta}|} = \frac{|\boldsymbol{\eta}|!}{\prod_{k=1}^K \eta_k!} \text{ and } \lambda_S = \sum_{k=1}^K \lambda^{(k)}.$$

For each j in Ω , define

$$g(j) = \sum_{|\boldsymbol{\eta}|=j} C_{\boldsymbol{\eta}}^{|\boldsymbol{\eta}|} \frac{f(\boldsymbol{\eta}; \mathbf{Q})}{\mathbf{h}(\boldsymbol{\eta})} \text{ and } \mathcal{M} = \{g(j)\}_{j \in \Omega}. \quad (18)$$

The following proposition is about the sequence \mathcal{M} .

Proposition 4 *The sequence \mathcal{M} is the moment sequence for a measure on \mathcal{R}^+ with T (possibly infinite) support points.*

The proof is given in the appendix.

In particular, for a multiple Poisson sample, for all j in Ω , we have

$$g(j) = j! \sum_{|\boldsymbol{\eta}|=j} f(\boldsymbol{\eta}; \mathbf{Q}),$$

because

$$C_{\boldsymbol{\eta}}^{|\boldsymbol{\eta}|} = |\boldsymbol{\eta}|! \mathbf{h}(\boldsymbol{\eta}).$$

Define

$$d(j) = g(j)/j! = \sum_{|\boldsymbol{\eta}|=j} f(\boldsymbol{\eta}; \mathbf{Q}) \quad \text{and} \quad \mathcal{D} = \{d(j)\}_{j \in \Omega}.$$

The following proposition is a corollary of **Proposition 4**.

Proposition 5 *The sequence \mathcal{D} represents a Poisson mixture density.*

Because of **Proposition 4**, we can construct a lower bound sequence for $g(0)$. For each natural number p , define the p th order lower bound for $g(0)$ to be

$$g_p = \mathbf{b}'_p(\mathcal{M}) [A_p(\mathcal{M})]^{-1} \mathbf{b}_p(\mathcal{M}),$$

where $p \leq T$ for a multiple Poisson sample when T is finite, and for a multiple binomial sample, $p \leq \min(T, [M/2])$.

Proposition 6 *If T is finite and for a multiple binomial sample $T \leq [M/2]$, then $g_T = g(0)$. If for a multiple Poisson sample, T is infinite, then $\{g_p\}_{p \in \mathcal{N}}$ is strictly increasing sequence with $g_p < g(0)$. Let*

$$g_{\infty} = \lim_{p \rightarrow \infty} g_p.$$

Then $g(0) = g_{\infty}$. So we have $g_T = g(0)$ in both cases.

The proof is omitted. See Mao and Lindsay (2001).

5.4 The estimator sequence

If $f(\boldsymbol{\eta}; \mathbf{Q})$ in (18) is “estimated” by $n_{\boldsymbol{\eta}}/N$, then we obtain an empirical version of the sequence \mathcal{M} , denoted by $\widehat{\mathcal{M}} = \{\hat{g}(j)\}_{j \in \Omega}$, where $\hat{g}(j)$ is an “estimator” for $g(j)$, and is given by

$$\hat{g}(j) = \sum_{|\boldsymbol{\eta}|=j} C_{\boldsymbol{\eta}}^{|\boldsymbol{\eta}|} \frac{n_{\boldsymbol{\eta}}}{N \mathbf{h}(\boldsymbol{\eta})}, \quad j \in \Omega.$$

In particular, for a multiple Poisson sample, we have

$$\hat{d}(j) = \sum_{|\boldsymbol{\eta}|=j} \frac{n_{\boldsymbol{\eta}}}{N}, \quad j \in \Omega.$$

Such an empirical sequence $\widehat{\mathcal{M}}$ can be used to construct a predictor sequence for n_0 and an estimator sequence for N . For each natural number p , if g_p is defined, then $A_p(\mathcal{M})$ is positive definite. Thus we have

$$P(A_p(\widehat{\mathcal{M}}) \text{ is positive definite}) \xrightarrow{P} 1 \quad \text{as } N \rightarrow +\infty.$$

When $A_p(\widehat{\mathcal{M}})$ is positive definite, let

$$\hat{g}_p = \mathbf{b}'_p(\widehat{\mathcal{M}})[A_p(\widehat{\mathcal{M}})]^{-1} \mathbf{b}_p(\widehat{\mathcal{M}}).$$

Thus \hat{g}_p can be regarded as an estimator for g_p and a prediction lower bound estimator for $g(0)$, and $N\hat{g}_p$ is a lower bound predictor for n_0 . Note that N is cancelled out in $N\hat{g}_p$. Let

$$\widehat{N}_p = n + N\hat{g}_p. \quad (19)$$

This will be called the p th order estimator for N . Note that \widehat{N}_p is a consistent estimator for N_p , where

$$N_p = N[1 - f(\mathbf{0}; \mathbf{Q}) + g_p].$$

Additionally, it is clear that the first order estimator \widehat{N}_1 is a linear extrapolation estimator.

For more discussions about such estimators, see Mao and Lindsay (2001). Mao and Lindsay (2001) also provided two selection rules about choosing a specific estimator from the \widehat{N}_p 's.

We next point out that for the multiple Poisson sample, the estimators developed in this subsection are identical to the estimators from subsample combination. In fact, we can obtain the sequence $\hat{d}(j)$ by combining all subsamples. Let n_j be the number of z_i 's that occur with frequency j .

We have

$$N\hat{d}(j) = n_j = \sum_{i=1}^N I(z_i = j), \quad j \in \Omega,$$

because

$$N\hat{d}(j) = \sum_{|\boldsymbol{\eta}|=j} n_{\boldsymbol{\eta}} = \sum_{|\boldsymbol{\eta}|=j} \sum_{i=1}^N I(\mathbf{x}_i = \boldsymbol{\eta}) = \sum_{i=1}^N I\left(\sum_{k=1}^K x_i^{(k)} = j\right) = \sum_{i=1}^N I(z_i = j).$$

Finally, we explain that the moment-based estimators are also pseudo-maximum likelihood estimators. When the p th order estimator \widehat{N}_p is used, we can think that the nuisance parameter ϑ is estimated by $N\widehat{g}_p/n$, which is clear from (5) and (19).

5.5 Confidence inference

We first present the asymptotic distribution for the estimators developed in last subsection. Let \mathbf{f}_p be a vector composed of elements in the set $\{f(\boldsymbol{\eta}; \mathbf{Q}) : |\boldsymbol{\eta}| \leq 2p\}$ with each element used once and only once. Let

$$\xi_p = \xi_p(\mathbf{f}_p) = N_p/N = 1 - f(\mathbf{0}; \mathbf{Q}) + g_p,$$

and

$$\sigma_p^2 = \frac{\partial \xi_p}{\partial \mathbf{f}_p'} (\text{diag} \mathbf{f}_p - \mathbf{f}_p \mathbf{f}_p') \frac{\partial \xi_p}{\partial \mathbf{f}_p}.$$

Then

$$(N\sigma_p^2)^{-1/2}(\widehat{N}_p - N_p) \xrightarrow{d} N(0, 1),$$

and

$$(N\xi_p^2/\sigma_p^2)^{1/2}(\log \widehat{N}_p - \log N_p) \xrightarrow{d} N(0, 1),$$

as N goes to infinity.

Again, the width of the confidence interval for N_p is proportional to \sqrt{N} while only the width of the confidence interval for $\log N_p$ shrinks at standard rate $1/\sqrt{N}$.

For the multiple Poisson sample, we give an explicit formula for the linear extrapolation estimator \widehat{N}_1 .

$$\xi_1 = 1 - f(\mathbf{0}; \mathbf{Q}) + [d(1)]^2/[2d(2)],$$

and

$$\sigma_1^2 = f(\mathbf{0}; \mathbf{Q}) + [d(1)]^3/[d(2)]^2 + [d(1)]^4/\{4[d(2)]^3\} - (\xi_1 - 1)^2.$$

Although we can define the σ_p^2 's for any p , it is difficult to express them explicitly generally. Moreover, since $f(\mathbf{0}; \mathbf{Q})$ is contained in σ_p^2 , if $f(\mathbf{0}; \mathbf{Q})$ cannot be consistently estimated, we cannot estimate σ_p^2 consistently.

As an alternative feasible approach, we present a bootstrap procedure to obtain confidence statements. We can estimate the multinomial distribution for all frequency-profiles counts. Let \hat{N} be an estimator for N . Then the cell probabilities are estimated by

$$\hat{f}(\mathbf{0}; \mathbf{Q}) = 1 - n/\hat{N} \quad \text{and} \quad \hat{f}(\boldsymbol{\eta}; \mathbf{Q}) = n_{\boldsymbol{\eta}}/\hat{N}, \quad \boldsymbol{\eta} \in \Omega_0.$$

We take the bootstrap re-samples from the estimated multinomial distribution and use the re-samples to get variance estimators.

Additionally, for a multiple Poisson sample, **Proposition 5** provides an easier way to perform the bootstrap procedure. Set

$$\hat{D}(j) = \sum_{|\boldsymbol{\eta}|=j} n_{\boldsymbol{\eta}}.$$

Since we can regard $\{\hat{D}_j\}_{j \in \Omega}$ as the frequency counts of a Poisson sample, which means they arise from a multinomial distribution with index N and cell probabilities $d(j)$'s. Thus we can estimate this multinomial distribution. The cell probabilities are estimated by

$$\tilde{d}(0) = 1 - n/\hat{N} \quad \text{and} \quad \tilde{d}(j) = \hat{D}(j)/\hat{N} \quad \text{for } j > 0.$$

For the construction of bootstrap confidence intervals, see Mao and Lindsay (2001).

5.6 A family of moment sequences

There are many alternative approaches to construct moment sequences of finite measures on \mathcal{R}^+ based on the multivariate mixture densities. We present some results as generalizations of **Propositions 4, 5, and 6**.

Let $\mathbf{w} = (w_1, w_2, \dots, w_K)'$ be a *weight vector*, where $w_k > 0$ for each k . Let

$$\lambda_{S, \mathbf{w}} = \sum_{k=1}^K w_k \lambda^{(k)}.$$

For each j in Ω , let

$$g_{\mathbf{w}}(j) = \sum_{|\boldsymbol{\eta}|=j} C_{\boldsymbol{\eta}}^{|\boldsymbol{\eta}|} \prod_{k=1}^K w_k^{\eta_k} \frac{f(\boldsymbol{\eta}; \mathbf{Q})}{\mathbf{h}(\boldsymbol{\eta})}.$$

In particular, for a multiple Poisson sample, we have

$$g_{\mathbf{w}}(j) = j! \sum_{|\boldsymbol{\eta}|=j} \prod_{k=1}^K w_k^{\eta_k} f(\boldsymbol{\eta}; \mathbf{Q}), \quad j \in \Omega.$$

Let

$$d_{\mathbf{w}}(j) = g_{\mathbf{w}}(j)/j! = \sum_{|\boldsymbol{\eta}|=j} \prod_{k=1}^K w_k^{\eta_k} f(\boldsymbol{\eta}; \mathbf{Q}), \quad j \in \Omega,$$

and

$$\mathcal{M}_{\mathbf{w}} = \{g_{\mathbf{w}}(j)\}_{j \in \Omega} \quad \text{and} \quad \mathcal{D}_{\mathbf{w}} = \{d_{\mathbf{w}}(j)\}_{j \in \Omega}.$$

Proposition 7 *The sequence $\mathcal{M}_{\mathbf{w}}$ is the moment sequence for variable $\lambda_{S, \mathbf{w}}$ of a measure on \mathcal{R} .*

The proof follows that of **Proposition 7** in a similar way.

Proposition 8 *The sequence $\mathcal{D}_{\mathbf{w}}$ represents a Poisson mixture density.*

For each p , when $A_p(\mathcal{M}_{\mathbf{w}})$ is positive finite, let

$$g_{p, \mathbf{w}} = \mathbf{b}'_p(\mathcal{M}_{\mathbf{w}}) [A_p(\mathcal{M}_{\mathbf{w}})]^{-1} \mathbf{b}_p(\mathcal{M}_{\mathbf{w}}).$$

Then the $g_{p, \mathbf{w}}$'s constitute a lower bound sequence for $g_{\mathbf{w}}(0) = g(0)$.

Proposition 9 *If T is finite and $T \leq [M/2]$ for a multiple binomial sample, then $g_{T, \mathbf{w}} = g(0)$. If T is infinite, for a multiple Poisson sample, $\{g_{p, \mathbf{w}}\}_{p \in \mathcal{N}}$ is strictly increasing sequence with $g_{p, \mathbf{w}} < g(0)$. Let*

$$g_{\infty, \mathbf{w}} = \lim_{p \rightarrow \infty} g_{p, \mathbf{w}}.$$

Then $g(0) = g_{\infty, \mathbf{w}}$. So we have $g_{T, \mathbf{w}} = g(0)$ in both cases.

Note that $\lambda_S = \lambda_{S, \mathbf{1}}$. Thus **Propositions 4, 5, 6** can be regarded as a special case of **Propositions 7, 8 and 9** respectively.

If $|\mathbf{w}| = 1$, then \mathbf{w} is called a *standardized weight vector*. Given \mathbf{w} , let $w_k^* = |\mathbf{w}|^{-1} w_k$. We can get a standardized weight vector \mathbf{w}^* , where

$$\mathbf{w}^* = (w_1^*, w_2^*, \dots, w_K^*)'.$$

It is clear that

$$\lambda_{S, \mathbf{w}} = |\mathbf{w}| \lambda_{S, \mathbf{w}^*} \quad \text{and} \quad g_{\mathbf{w}}(j) = |\mathbf{w}|^j g_{\mathbf{w}^*}(j).$$

The sequence $\mathcal{M}_{\mathbf{w}^*} = \{g_{\mathbf{w}^*}(j)\}_{j \in \Omega}$ is called a *standardized sequence*. The following proposition shows that any standardized weight vector uniquely determines a lower bound sequence.

Proposition 10 *The original sequence $\mathcal{M}_{\mathbf{w}}$ and the standardized sequence $\mathcal{M}_{\mathbf{w}^*}$ yield the same lower bound sequence. That is, for each p .*

$$g_{p,\mathbf{w}} = g_{p,\mathbf{w}^*} \quad \text{where} \quad g_{p,\mathbf{w}^*} = \mathbf{b}'_p(\mathcal{M}_{\mathbf{w}^*})[A_p(\mathcal{M}_{\mathbf{w}^*})]^{-1}\mathbf{b}_p(\mathcal{M}_{\mathbf{w}^*}).$$

The proof is given in the appendix.

Remark 2 *We recommend taking the inverse of the sample size of subsamples as the weights for a multiple Poisson sample when the sample size is an indicator of sampling effort. Let s_k be the sample size of the k th subsample. Then set*

$$w_k = s_k^{-1}, \quad k = 1, 2 \dots K.$$

For a multiple binomial sample, we suggest the weights as follows.

$$w_k = [M^{(k)}]^{-1}, \quad k = 1, 2 \dots K.$$

When all these s_k 's or $M^{(k)}$'s are close to each other, the simple sequence with weight vector $\mathbf{1}$ is recommended.

6 EXAMPLES

6.1 Hepatitis A virus data

From April to July 1995, an outbreak of the Hepatitis A virus infection occurred mainly in and around a technical college with about 5000 students in Taiwan. The dataset is about the infected students in that college, which was analyzed in Tsay and Chao (2001). There are three lists: records based on a serum test taken by the Institute of Preventive Medicine, Department of Health of Taiwan, records by doctors in local hospitals, and records based on questionnaires conducted by epidemiologists, in which cases were either confirmed by serum test or identified by symptom combinations.

The frequency profiles are: $n_{(1,0,0)} = 69$, $n_{(0,1,0)} = 55$, $n_{(0,0,1)} = 63$, $n_{(1,1,0)} = 21$, $n_{(0,1,1)} = 18$, $n_{(1,0,1)} = 17$, and $n_{(1,1,1)} = 28$. We have $n = 271$.

In the model M_t , we assume that all affected students have the same probability to be in a list but the probabilities in different lists may be different. We have

$$\hat{N}_t = n + \left[\frac{[n_{(1,0,0)} n_{(0,1,0)} n_{(0,0,1)}]^2}{n_{(1,1,0)} n_{(0,1,1)} n_{(1,0,1)}} \right]^{1/3} = 271 + 207 = 478.$$

Suppose the model M_t is appropriate, then we have the two-sided asymptotic 95% confidence interval for N_t is (397, 559). The one-sided asymptotic 95% confidence interval for N_t is (410, $+\infty$). The two-sided bootstrap 95% confidence interval for N_t is (415, 589) (500 re-samples) The one-sided bootstrap 95% confidence interval for N_t is (428, $+\infty$). Note that the lower limit are valid for N regardless of the heterogeneity structure.

In December 1995, the National Quarantine Service of Taiwan conducted a screen serum test and concluded that the number of infected students N was about 545. It seems that \hat{N}_t underestimated the true value although the confidence intervals cover N . The underestimation could also be due to the negative bias of linear extrapolation if there was heterogeneity of the probabilities that students were infected.

6.2 Bird census data

The bird data were collected in two river estuaries in Taiwan (Ke-yar river and Chung-kang river) every Sunday morning from April 1994 to March 1995. The two bird communities share many species because they are only 20 kilometers apart, and the environmental factors are similar. The dataset was first analyzed by Chen et al. (1995). Chao et al. (2000) re-analyzed the dataset. In Chao et al. (2000), the estimates for the number of bird species in Ke-yar River, the number of bird species in Chung-kang River and the number of shared bird species are 181, 163

and 134, associated with bootstrap 95% confidence intervals (167, 213), (150, 192) and (116, 210) respectively. The total number of distinct species in at least one of the two communities is therefore estimated to be $181 + 163 - 134 = 210$.

The first six frequency counts of based on the subsample combination are given by $n_1 = 22$, $n_2 = 13$, $n_3 = 8$, $n_4 = 5$, $n_5 = 7$, $n_6 = 5$. The remainder nonzero frequency counts are n_{10} and n_{14} equal 3, n_{18} equal 5, n_9 , n_{13} , n_{15} , n_{20} , n_{25} , n_{77} , n_{97} and n_{353} equal 2, and n_7 , n_8 , n_{12} , n_{16} , \dots , totally 95 frequency counts equal 1. The estimated lower bound sequence is $\hat{N}_1 = 203$, $\hat{N}_2 = 209$, $\hat{N}_3 = 203$, $\hat{N}_4 = 212$ and $\hat{N}_5 = 216$. Following the first selection rule from Mao and Lindsay (2001), we use \hat{N}_p with the largest p for which the sequence $\{\hat{N}_1, \hat{N}_2, \dots, \hat{N}_p\}$ is increasing. This gives $\hat{N}_2 = 209$ as the estimator for N . which is close to the estimate in Chao et al. (2000). The two-sided bootstrap 95% confidence interval is (196, 264). The one-sided bootstrap 95% confidence interval is (197, $+\infty$).

6.3 Tomato EST data

A prepared cDNA library may consist of $10^5 \sim 10^7$ clones. Each clone represents a copy of an expressed gene. Sequencing random picked clones will generate Expressed Sequence Tags (ESTs), which are the “single-pass” cDNA sequences (Cantor and Smith 1999). The ESTs from the sample are clustered into groups, each corresponding to a distinct gene. The Institute of Genomic Research (TIGR) hosts public databases of plant, animal and microbial ESTs (Quackenbush et al. 2000). Two tomato flower cDNA libraries (TIGR library identifier: T1626 and T1527) are analyzed in this paper. The total number of distinct expressed genes in at least one of two libraries is estimated.

The first ten frequency counts of based on subsample combination are given by $n_1 = 2314$, $n_2 = 503$, $n_3 = 163$, $n_4 = 71$, $n_5 = 33$, $n_6 = 24$, $n_7 = 16$, $n_8 = 6$, $n_9 = 8$ and $n_{10} = 1$. The remainder nonzero frequency counts are $n_{14} = 6$, $n_{12} = 3$, n_{13} , n_{15} , n_{17} and n_{18} equal 2, and n_{11} ,

n_{16} , n_{19} , n_{25} , n_{28} , n_{38} and n_{43} equal 1. The estimated lower bounds are $\hat{N}_1 = 8486$, $\hat{N}_2 = 10539$, $\hat{N}_3 = 10640$, $\hat{N}_4 = 9895$, $\hat{N}_5 = 12118$, $\hat{N}_6 = 12149$, $\hat{N}_5 = 9626$, $\hat{N}_8 = 11900$ and $\hat{N}_9 = 9259$. If we follow the first selection rule from Mao and Lindsay (2001), we will use $\hat{N}_3 = 10640$ as the estimator for N . The two-sided bootstrap 95% confidence interval is (8879, 38988). The upper limit is noninformative since it is known that the total number of genes in tomatoes is about 23000. The one-sided bootstrap 95% confidence interval is (9115, $+\infty$).

7 DISCUSSION

We presented a mixture model framework for the estimation of the number of classes in at least one of a collection of populations. The linear extrapolation estimators are tight for the case when the multivariate latent distribution is degenerate at a point. They are lower bounds for general situations. In particular, we provide a family of linear extrapolation estimators for population size models, which include several estimators developed in the literature. The moment-based estimators do allow a general multivariate latent distribution.

All of the estimators presented in this paper can be regarded as pseudo maximum likelihood estimators with the nuisance parameter estimated by moment methods. The authors are investigating nonparametric maximum likelihood estimation for the nuisance parameter ϑ .

APPENDIX: PROOFS

Proof of Proposition 1

The Poisson and binomial densities form exponential families. Let

$$\tau^{(k)} = \log \lambda^{(k)}, \quad k = 1, 2 \dots K, \quad \text{and} \quad \boldsymbol{\tau} = (\tau^{(1)}, \tau^{(2)}, \dots, \tau^{(K)})'.$$

Then the \mathbf{Q} -distributed $\boldsymbol{\lambda}$ induces a distribution for $\boldsymbol{\tau}$, denoted by \mathbf{U} , that is,

$$\boldsymbol{\lambda} \sim \mathbf{Q}(\boldsymbol{\lambda}) \implies \exists \mathbf{U}, \quad \text{such that} \quad \boldsymbol{\tau} \sim \mathbf{U}(\boldsymbol{\tau}).$$

Define a re-weighted measure of $d\mathbf{U}$ as follows:

$$d\mathbf{V} = \prod_{k=1}^K c(\tau^{(k)}) d\mathbf{U}.$$

Then

$$R(\boldsymbol{\eta}) = \log \frac{f(\boldsymbol{\eta}; \mathbf{Q})}{\mathbf{h}(\boldsymbol{\eta})} = \log \int \exp(\boldsymbol{\tau}' \boldsymbol{\eta}) d\mathbf{V},$$

which is the logarithm of the moment generating function for the re-weighted measure $d\mathbf{V}$. Thus the Hessian matrix of $R(\boldsymbol{\eta})$ is non-negative definite, which is defined by

$$\frac{\partial^2 R(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}$$

The Hessian matrix is identical to a zero-matrix if and only if \mathbf{Q} is degenerate at one point. Convexity comes from the nonnegative definiteness of the Hessian matrix while linearity is due to the fact that the Hessian matrix is identical to a zero-matrix.

Proof of Proposition 4

Define a re-weighted version of the latent distribution as follows:

$$d\mathbf{P}(\boldsymbol{\lambda}) = c(\mathbf{Q}) \prod_{k=1}^K c(\lambda^{(k)}) d\mathbf{Q}(\boldsymbol{\lambda})$$

where

$$[c(\mathbf{Q})]^{-1} = \int \prod_{k=1}^K c(\lambda^{(k)}) d\mathbf{Q}(\boldsymbol{\lambda}).$$

In fact, \mathbf{Q} is a re-weighted measure from \mathbf{P} since we have

$$d\mathbf{Q}(\boldsymbol{\lambda}) = c(\mathbf{P}) \left[\prod_{k=1}^K c(\lambda^{(k)}) \right]^{-1} d\mathbf{P}(\boldsymbol{\lambda})$$

where

$$[c(\mathbf{P})]^{-1} = \int \left[\prod_{k=1}^K c(\lambda^{(k)}) \right]^{-1} d\mathbf{P}(\boldsymbol{\lambda}).$$

We obtain

$$f(\boldsymbol{\eta}; \mathbf{Q}) = \mathbf{h}(\boldsymbol{\eta})c(\mathbf{P}) \int \left[\prod_{k=1}^K \lambda^{(k)} \right]^{\boldsymbol{\eta}_k} d\mathbf{P}(\boldsymbol{\lambda}).$$

For each j in Ω , we have

$$\lambda_S^j = \sum_{|\boldsymbol{\eta}|=j} C_{\boldsymbol{\eta}}^{|\boldsymbol{\eta}|} \prod_{k=1}^K [\lambda^{(k)}]^{\eta_k},$$

which yields

$$E_{\mathbf{P}} \lambda_S^j = \sum_{|\boldsymbol{\eta}|=j} C_{\boldsymbol{\eta}}^{|\boldsymbol{\eta}|} \int \prod_{k=1}^K [\lambda^{(k)}]^{\eta_k} d\mathbf{P}(\boldsymbol{\lambda}) = [c(\mathbf{P})]^{-1} \sum_{|\boldsymbol{\eta}|=j} C_{\boldsymbol{\eta}}^{|\boldsymbol{\eta}|} \frac{f(\boldsymbol{\eta}; \mathbf{Q})}{\mathbf{h}(\boldsymbol{\eta})},$$

and

$$g(j) = \sum_{|\boldsymbol{\eta}|=j} C_{\boldsymbol{\eta}}^{|\boldsymbol{\eta}|} \frac{f(\boldsymbol{\eta}; \mathbf{Q})}{\mathbf{h}(\boldsymbol{\eta})} = c(\mathbf{P}) E_{\mathbf{P}} \lambda_S^j.$$

Therefore, the proposition holds.

Proof of Proposition 10

Let

$$H_{p, \mathbf{w}}(y) = \begin{pmatrix} y & \mathbf{b}_p(\mathcal{M}_{\mathbf{w}})' \\ \mathbf{b}_p(\mathcal{M}_{\mathbf{w}}) & A_p(\mathcal{M}_{\mathbf{w}}) \end{pmatrix},$$

and

$$H_{p, \mathbf{w}^*}(y) = \begin{pmatrix} y & \mathbf{b}_p(\mathcal{M}_{\mathbf{w}^*})' \\ \mathbf{b}_p(\mathcal{M}_{\mathbf{w}^*}) & A_p(\mathcal{M}_{\mathbf{w}^*}) \end{pmatrix},$$

which yields

$$|H_{p, \mathbf{w}}(y)| = |\mathbf{w}|^{p(p+1)} |H_{p, \mathbf{w}^*}(y)|.$$

Note that that the p th order lower bounds $g_{p, \mathbf{w}}$ and g_{p, \mathbf{w}^*} are unique roots of the following equations respectively (Mao and Lindsay 2001)

$$|H_{p, \mathbf{w}}(y)| = 0 \quad \text{and} \quad |H_{p, \mathbf{w}^*}(y)| = 0.$$

Hence, we have

$$g_{p, \mathbf{w}} = g_{p, \mathbf{w}^*}.$$