Biometrics 64, 977–981
September 2008
DOI: 10.1111/j.1541-0420.2008.01078.x

READER REACTION

On the Nonidentifiability of Population Sizes

Chang Xuan Mao
Department of Statistics, University of California, Riverside, CA 92521, U.S.A.
email: cmao@stat.ucr.edu

Summary. When a nonparametric mixture model is adopted to deal with the heterogeneity among individual capture probabilities, the population size is nonidentifiable (Link, 2003, *Biometrics* **59,** 1123–1130). Holzmann, Munk, and Zucchini (2006, *Biometrics* **62,** 934–936) discussed the conditions under which a subfamily of mixing distributions is identifiable. Link (2006, *Biometrics* **92**, 936–939) found that the nonidentifiability occurs across identifiable subfamilies. It is shown that there is a subfamily in which each mixing distribution is determined by its mixture, and the population size admits estimable lower bounds that can be used to construct lower confidence limits.

Key words: Binomial mixtures; Lower bounds; Species richness.

1. Introduction

There is a huge literature on estimating population sizes because of its enormous applications. The heterogeneity among individual capture probabilities usually should not be ignored so that mixture models are necessary. The population size can be nonidentifiable in a mixture model (Link, 2003). Holzmann, Munk, and Zucchini (2006; hereafter HMZ) and Link (2006) investigated the issue of nonidentifiability. As a follow-up to these important studies, we will provide a deep investigation on the nonidentifiability, determine what statements are over-optimistic or over-pessimistic, and define inferential tasks useful for the practical purpose.

Suppose that a population has N individuals. Let f_x denote the number of individuals captured x times in a study of T captures. These counts are modeled by binomial mixtures when individual capture probabilities are assumed to follow a mixing distribution. Although the binomial mixture model is our focus, similar investigation can be done for some other mixture models.

It is over-optimistic if one tries to estimate the population size N nonparametrically, which unfortunately, is a common practice in the literature. An exception is Link's (2003) analysis, which established the issue of nonidentifiability. HMZ provided the conditions under which a subfamily of mixing distributions is identifiable. Link (2006) raised the issue of nonidentifiability across subfamilies. There is a subfamily in which each mixing distribution satisfies a condition slightly stronger than that in HMZ and is determined by the truncated mixture.

Let $n = \sum_{x=1}^{T} f_x$ denote the number of observed individuals. The problem of estimating the population size N can be reduced to estimating the odds $\theta = E(f_0)/E(n)$ that an individual is unseen (Mao, 2007, 2008). A more pertinent issue is when the odds θ are identifiable. The odds are nonidentifiable nonparametrically (Mao, 2007). There is a subfamily of mixing distributions in which the odds are determined by the truncated mixture.

Link (2006) concluded that "*without strong assumptions*, f_1, $f_2, \ldots, f_T$ *tell us essentially nothing about* f_0," which is over-pessimistic. In fact, $E(f_0)$ or equivalently θ admits estimable lower bounds. A lower bound to θ leads to a lower bound to the population size N. Lower confidence limits for these lower bounds can be constructed and also serve as lower confidence limits for N.

2. Results

Let X denote the number of captures of an individual in a study of T captures, which is binomial given the capture probability p. If p is assumed to follow a mixing distribution G, then

$$X \sim \pi_G(x) = \int \binom{T}{x} p^x (1-p)^{T-x}\, dG(p),\ x = 0, 1, \ldots, T.$$

Conditioning on n, $(f_1, f_2, \ldots, f_T)'$ is multinomial with probabilities $\pi^c_G(x) = \pi_G(x)/\{1 - \pi_G(0)\}$, $x = 1, 2, \ldots, T$ (Link, 2003). Reformulate π^c_G as a mixture h_Q by reparameterizing G (Mao and Lindsay, 2002), where

$$\begin{aligned} h_Q(x) &= \int \binom{T}{x} \frac{p^x(1-p)^{T-x}}{1-(1-p)^T}\, dQ(p), x = 1, 2, \ldots, T, \\ dQ(p) &= \frac{\{1-(1-p)^T\}\, dG(p)}{\int \{1-(1-q)^T\}\, dG(q)}. \end{aligned} \qquad (1)$$

Because (1) is invertible for $p \in (0, 1]$ and Q and G have the same support points, relevant statements in HMZ and Link (2003, 2006) in terms of G will be cited here in terms of Q.

Let $\mathcal{A}$ be the family of all mixing distributions Q over (0, 1]. The family $\mathcal{A}$ is nonidentifiable in the sense that there are Q_1 and Q_2 in $\mathcal{A}$ with $h_{Q_1} = h_{Q_2}$ but $Q_1 \neq Q_2$ (Mao, 2007). For example, if $G = 3/4\delta(1/4) + 1/4\delta(3/4)$ in (1), then

$$Q = 35/52\delta(1/4) + 17/52\delta(3/4), \tag{2}$$

$$h_Q = (28, 18, 12, 7)'/65, \tag{3}$$

where $\delta(p)$ is a distribution degenerate at p. If $G = \mathrm{Beta}(1/2, 1/3)$ in (1), then a mixture identical to h_Q in (3) can be produced (Link, 2006).

A subfamily $\mathcal{S} \subset \mathcal{A}$ is said to be identifiable if, given Q_1 and Q_2 in $\mathcal{S}, h_{Q_1} = h_{Q_2}$ implies that $Q_1 = Q_2$. HMZ considered conditions under which a subfamily is identifiable. Link (2006) found that there are two identifiable subfamilies $\mathcal{S}_1$ and $\mathcal{S}_2$ and two mixing distributions $Q_1 \in \mathcal{S}_1$ and $Q_2 \in \mathcal{S}_2$ such that $Q_1 \neq Q_2$ but $h_{Q_1} = h_{Q_2}$. The observation in Link (2006) invites us to address whether there is a subfamily in which Q is uniquely determined by its mixture h_Q.

Let index (Q) represent the number of support points of Q with a support point $p = 1$ counted as 1/2 (Lindsay, 1995, p. 48). For example, index (Q_1) = 1.5 and index (Q_2) = 2, where $Q_1 = 0.2\delta(0.5) + 0.8\delta(1)$ and $Q_2 = 0.2\delta(0.5) + 0.8\delta(0.9)$.

Proposition 1: *For $Q \in \mathcal{U}, h_Q$ determines Q in the sense that there is no other mixing distribution P such that $h_P = h_Q$, where*

$$\mathcal{U} = \{Q \in \mathcal{A} : \mathrm{index}(Q) \leqslant (T-1)/2\}. \tag{4}$$

Proposition 1 is an application of Proposition 6 in Lindsay (1995, p. 48). For $Q \in (\mathcal{A} - \mathcal{U}), h_Q$ is produced by many mixing distributions. For example, $Q \in (\mathcal{A} - \mathcal{U})$ for Q in (2) as index (Q) = 2 and $T = 4$ but $(T-1)/2 = 1.5$.

From HMZ, the subfamily $\mathcal{D}$ is identifiable, where

$$\mathcal{D} = \{Q \in \mathcal{A} : \text{the number of support points of } Q \leqslant T/2\}. \tag{5}$$

Proposition 2: *If T is odd, then $\mathcal{U} = \mathcal{D}$. If T is even, then*

$$\mathcal{D} - \mathcal{U} = \{Q \in \mathcal{A} : \mathrm{index}(Q) = T/2\}. \tag{6}$$

For an even T, HMZ's condition is necessary for $\mathcal{D}$ to be identifiable but not sufficient for Q to be determined by h_Q. For example, $Q \in (\mathcal{D} - \mathcal{U})$ for Q in (2).

Next we turn to the odds $\theta = E(f_0)/E(n)$. It depends on Q as follows:

$$\theta = \theta(Q) = \int \frac{(1-p)^T}{1-(1-p)^T}\, dQ(p). \tag{7}$$

The odds $\theta(Q)$ is nonparametrically nonidentifiable in the sense that there are Q_1 and Q_2 in $\mathcal{A}$ with $h_{Q_1} = h_{Q_2}$ but $\theta(Q_1) \neq \theta(Q_2)$ (Mao, 2007).

The nonidentifiability of $\theta(Q)$ implies that it is impossible to estimate $\theta(Q)$ unbiasedly and precisely. For example, given a real application with $f_1 = 187$, $f_2 = 56$, and $f_3 = 28$ (Chao et al., 2001), Figure 1 presents $\hat{f}_0, \hat{f}_1, \hat{f}_2$, and $\hat{f}_3$, where

$$\hat{f}_0 = n\theta(\widehat{Q}),\ \hat{f}_x = nh_{\widehat{Q}}(x), x = 1, 2, 3,$$
$$\widehat{Q} = \alpha\delta(p) + (1-\alpha)\delta(\eta),\ \alpha \in (0, 1),\ 0 < p \leqslant \eta \leqslant 1.$$

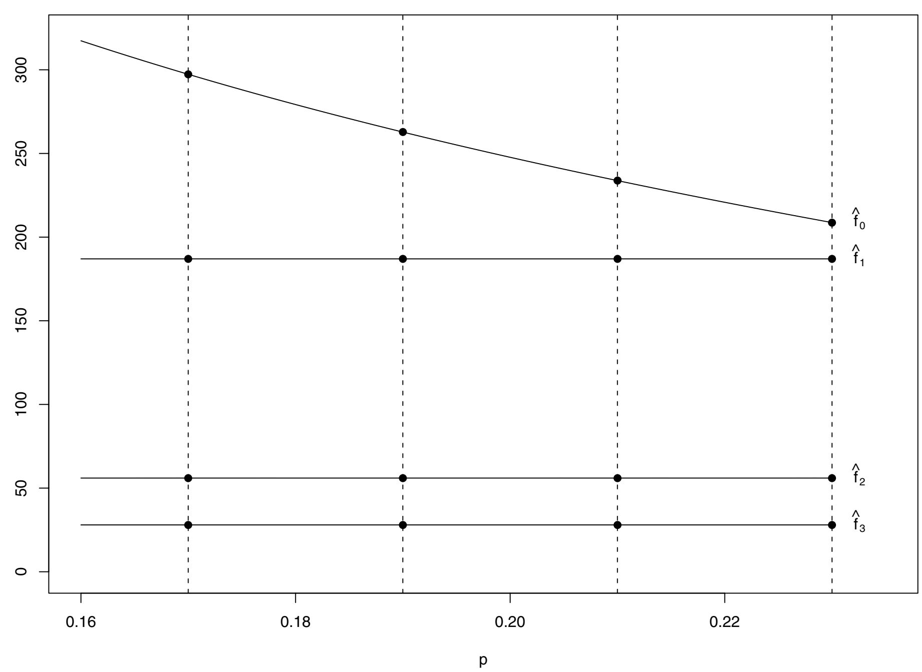

Figure 1. Given $p, \widehat{Q} = \alpha\delta(p) + (1-\alpha)\delta(\eta)$ satisfies $\hat{f}_x = f_x$ for x = 1, 2, 3 with $\hat{f}_0$ varying over p. For example, given p = 0.17, 0.19, 0.21, and 0.23 (dashed lines), $\hat{f}_1 = 187, \hat{f}_2 = 56, \hat{f}_3 = 28, \hat{f}_0$ = 297, 263, 234, and 209, respectively (solid dots).

For each fixed p, both η and α are found by solving the equation system,

$$\left.\begin{aligned} h_{\widehat{Q}}(1) = f_1/n \\ h_{\widehat{Q}}(2) = f_2/n \end{aligned}\right\} \Longleftrightarrow \alpha = \frac{f_1/n - h_{\delta(\eta)}(1)}{h_{\delta(p)}(1) - h_{\delta(\eta)}(1)} = \frac{f_2/n - h_{\delta(\eta)}(2)}{h_{\delta(p)}(2) - h_{\delta(\eta)}(2)}.$$

It is clear that $\hat{f}_x \equiv f_x$ for $x = 1, 2, 3$, and $\hat{f}_0$ varies over p.

There is no upper bound to the odds $\theta(Q)$ in a neighborhood of h_Q (Mao and Lindsay, 2007). It makes sense to consider lower bounds to $\theta(Q)$. There exist lower bounds to $\theta(Q)$. For example, from the estimator $n + (T-1)f_1^2/(2Tf_2)$ (Chao, 1989), $(T-1)h_Q^2(1)\{2Th_Q(2)\}$ is a lower bound to $\theta(Q)$ provided that $Q \neq \delta(1)$. Among various lower bounds, the sharpest one is

$$\phi(h_Q) = \inf\{\theta(P) : h_Q(x) = h_P(x), x = 1, 2, \ldots, T, \forall P \in \mathcal{A}\}. \tag{8}$$

A linear program can be used to calculate $\phi(h_Q)$ numerically (Mao, 2007).

To determine when $\phi(h_Q)$ is achieved theoretically, define

$$\mathcal{K} = \{Q \in \mathcal{A} : \text{index}(Q) \leqslant T/2\}. \tag{9}$$

PROPOSITION 3: *If $Q \in \mathcal{K}$, then $\phi(h_Q) = \theta(Q)$ and $\phi(h_Q) < \theta(Q)$ otherwise.*

Proposition 3 is an application of Theorem 1 in Mao (2008). If $Q \in (\mathcal{A} - \mathcal{K})$, then $\theta(Q)$ consists of an identifiable component $\phi(h_Q)$ and a nonidentifiable component $\theta(Q) - \phi(h_Q)$ by writing

$$\theta(Q) = \phi(h_Q) + \{\theta(Q) - \phi(h_Q)\}. \tag{10}$$

The mixture h_Q tells nothing about $\theta(Q) - \phi(h_Q)$ for $Q \in (\mathcal{A} - \mathcal{K})$. For example, $\phi(h_Q) = 61/195$ and $\theta(Q) = 61/195 = \phi(h_Q)$ for Q in (2), and for Q obtained from $G = \text{Beta}(1/2, 3/2), Q \in (\mathcal{A} - \mathcal{K})$ and $\theta(Q) = 189/195 > \phi(h_Q)$.

A lower bound to $\theta(Q)$ yields a lower bound to the population size N, e.g.,

$$N_\phi = N\{1 + \phi(h_Q)\}/\{1 + \theta(Q)\},$$

which can be estimated by $n + n\phi(h_{\widehat{Q}})$. Lower confidence limits for a lower bound to N can be constructed. They are also lower confidence limits for N and useful in practice. For example, say, $n = 20{,}000$ injection drug users (IDUs) in Los Angeles were observed from a capture–recapture study. The true size N of the IDU population is between 20,000 and the size of the whole population of Los Angeles. It is helpful to public health decision making if one concludes that there are at least, say, 30,000 IDUs at the 95% confidence level.

References

Chao, A. (1989). Estimating population size for sparse data in capture-recapture experiments. *Biometrics* **45,** 427–438.

Chao, A., Tsay, P. K., Lin, S.-H., Shau, W.-Y., and Chao, D.-Y. (2001). The applications of capture-recapture models to epidemiological data. *Statistics in Medicine* **20,** 3123–3157.

Holzmann, H., Munk, A., and Zucchini, W. (2006). On identifiability in capture-recapture models. *Biometrics* **62,** 934–936.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications*. Hayward, California: Institute of Mathematical Statistics.

Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics* **59,** 1123–1130.

Link, W. A. (2006). Reply to "On identifiability in capture-recapture models." *Biometrics* **92,** 936–939.

Mao, C. X. (2007). Estimating population sizes for capture-recapture sampling with binomial mixtures. *Computational Statistics and Data Analysis* **51,** 5211–5219.

Mao, C. X. (2008). Lower bounds to the population size when capture probabilities vary over individuals. *Australian and New Zealand Journal of Statistics* **50,** 1–10.

Mao, C. X. and Lindsay, B. G. (2002). Diagnostics for the heterogeneity of inclusion probabilities in a Bernoulli census. *Indian Journal of Statistics* **64,** 626–634.

Mao, C. X. and Lindsay, B. G. (2007). Estimating the number of classes. *Annals of Statistics* **35,** 917–930.

Received October 2006. *Revised February* 2007.
Accepted December 2007.

Appendix

Proof of Proposition 2. Let $\kappa(Q)$ denote the number of support points of Q. When T is odd with $T = 2m + 1$, conclude that $\mathcal{D} = \mathcal{U}$ because, with index $(Q) \leqslant \kappa(Q) \leqslant$ index $(Q) + 1/2$,

$$\begin{aligned} \text{index}(Q) \leqslant m \Longrightarrow\ & \kappa(Q) \leqslant m + 1/2 \\ \Longleftrightarrow\ & \kappa(Q) \leqslant m \Longrightarrow \text{index}(Q) \leqslant m. \end{aligned}$$

When T is even with $T = 2m, \mathcal{U} \subset \mathcal{D}$ because index $(Q) \leqslant (T-1)/2 = m - 1/2$ implies that $\kappa(Q) \leqslant \text{index}(Q) + 1/2 \leqslant m$. For any Q with $\kappa(Q) \leqslant m - 1$, $\text{index}(Q) \leqslant \kappa(Q) \leqslant m - 1 < m - 1/2$ so that $Q \in \mathcal{U}$. For any Q with $\kappa(Q) = m$, either $\text{index}(Q) = m - 1/2$ so that $Q \in \mathcal{U}$ or $\text{index}(Q) = m$ so that $Q \in (\mathcal{D} - \mathcal{U})$. Conclude that $\mathcal{D} - \mathcal{U}$ contains all Q with $\kappa(Q) =$ index $(Q) = m$.

The authors replied as follows:

The problem of nonidentifiability of the mixing distribution and hence of the population size in closed-population capture–recapture experiments with heterogeneous individual capture probabilities was observed and discussed in a seminal paper by Link (2003). In Holzmann, Munk, and Zucchini (2006; hereafter HMZ), we put this into the general context of identifiability of finite mixtures and gave conditions under which a parametric subfamily of mixing distributions is identifiable. In his response, Link (2006) gave examples to illustrate that identifiability does not extend across (identifiable) parametric subfamilies. Thus, if one is not willing to assume some specific parametric model for the mixing distribution (perhaps based on some additional information), the results in HMZ are not helpful any more. In the present article, Mao

(2008b) gives conditions under which mixing distributions can be identified within the family of all mixing distributions. For such mixing distributions, the problems raised by Link (2003, 2006) cannot occur.

In Section 1 we clarify the distinction between the different notions of identifiability (as defined in HMZ and in Mao, 2008b). In Section 2 this is linked to the notion of estimable lower bounds to the population size, which has been popularized by Mao (2007, 2008a, 2008b) and Mao and Lindsay (2007). In particular, we point out that under some additional restrictions, these may also be estimated in certain parametric models by (conditional) parametric maximum likelihood.

1. Notions of Identifiability

We briefly recall the statistical model considered in HMZ and in Mao (2008b). Let X denote the number of captures of an individual in T captures. Then

$$X \sim \pi_G(x) = \int \binom{T}{x} p^x (1-p)^{T-x}\, dG(x), \quad x = 0, \ldots, T,$$

where G is a mixing distribution on (0, 1]. Let f_x be the number of individuals observed at exactly x times, and let $n = \sum_x f_x$ be the total number of observed individuals. Conditional on n, the observations $(f_1, \ldots, f_T)$ are multinomial with probabilities $\pi_G^c(x) = \pi_G(x)/(1 - \pi_G(0))$, $x = 1, \ldots, T$.

Let $\mathcal{A}$ denote the family of all mixing distributions on (0, 1]. In HMZ, we introduced the notion of *identifiability of a subfamily* $\mathcal{S} \subset \mathcal{A}$ from the observations f_x as follows. If $\pi_G^c = \pi_H^c$ for $G, H \in \mathcal{S}$, then $G = H$. This notion of identifiability corresponds to the classical notion introduced by Teicher (1963) for ordinary (nontruncated) binomial mixtures. It is what is needed to make, together with additional regularity conditions such as compactness of the parameter space, the standard maximum likelihood estimation theory work (cf. e.g., Leroux, 1992). This means that if the true $G_0 \in \mathcal{S}$, it can be consistently estimated by maximum likelihood, and hence also $\pi_{G_0}(0)$. An important example is the family of finite mixtures $\mathcal{S}_{\text{fin}}$ with at most $T/2$ support points.

However, as already indicated in HMZ and explicitly pointed out by Link (2006), identifiability in this notion does not extend across families. Indeed, Link constructs examples of identifiable families $\mathcal{S}_1$ and $\mathcal{S}_2$, for which there are $G \in \mathcal{S}_1$ and $H \in \mathcal{S}_2$ with $\pi_G^c = \pi_H^c$. Thus, based on the data one cannot distinguish between the mixing distributions G and H. This problem is serious because the quantity of interest, the proportion of unobserved individuals, $\pi_G(0)$ and $\pi_H(0)$, can differ significantly. Therefore, if there are no proper reasons (based on additional information and not on the observations f_x) to choose a specific family of mixing distributions $\mathcal{S}$, this direct parametric approach may be unsatisfactory.

Of course, a similar problem in principle also occurs for ordinary binomial mixtures. In order to overcome it, Lindsay and Roeder (1993) and Lindsay (1995) gave a stronger condition than identifiability in a subfamily, and in the present article, Mao (2008b) (see also Mao, Colwell, and Chang, 2005) extends it to the truncated binomial mixtures arising in capture-recapture studies. Call a mixing distribution $G \in \mathcal{A}$ *strongly identifiable* if π_G^c cannot be reproduced by any other mixing distribution $H \in \mathcal{A}$: If $\pi_G^c = \pi_H^c$ for any $H \in \mathcal{A}$, then $G = H$. Using results by Lindsay and Roeder (1993), Mao (2008b) describes the family $\mathcal{U}$ of strongly identifiable mixing distributions as follows. For $G \in \mathcal{A}$ let $\#G$ be the number of support points, and let index G be the number of support points, where a support point at 1 is only counted as 1/2. Then $\mathcal{U}$ is the set of all G with index $G \leq (T-1)/2$. Evidently, the family $\mathcal{U}$ is identifiable, but more is true, the problem with other subfamilies with elements giving rise to the same mixing distributions cannot occur. For odd $T, \mathcal{U}$ is simply the family $\mathcal{S}_{\text{fin}}$ of identifiable mixtures with at most $T/2$ support points, for even T the family $\mathcal{U}$ does not contain all finite mixtures with $\#G \leq T/2$, and is thus smaller then $\mathcal{S}_{\text{fin}}$.

2. Identifiability and Estimable Lower Bounds

In a series of papers, Mao (2007, 2008a) and Mao and Lindsay (2007) discussed the construction and estimation of lower bounds to the odds and hence to the population size. Due to the rather obvious problem of possible subjects that may be very hard to capture, this seems to be a very appealing and practically relevant approach. In the present article, Mao (2008b) introduces a parametric family $\mathcal{K}$ for which these lower bounds are achieved. Here we show that (a) $\mathcal{K}$ is identifiable, (b) it generates all possible conditional mixtures π_G^c.

To start, for the population size N one has $N = E_G n/(1 - \pi_G(0)) = En(1 + \pi_G(0)/(1 - \pi_G(0)))$, thus, in order to obtain lower bounds for N one needs to find lower bounds for the odds $\theta(G) = \pi_G(0)/(1 - \pi_G(0))$. Following Mao (2007), define

$$\phi(\pi_G^c) = \inf\{\theta(H) : \pi_G^c = \pi_H^c, H \in \mathcal{A}\}. \tag{1}$$

As in Mao (2008b) consider $\mathcal{K} = \{G \in \mathcal{A} : \text{index}\, G \leq T/2\}$. Mao (2008b) shows that $\mathcal{K}$ is the set of mixing distributions for which the inf in (1) is obtained: $\phi(\pi_G^c) = \theta(G)$ if and only if $G \in \mathcal{K}$. Further, from Mao (2007) the infimum in (1) is always obtained.

Together with the results on strong identifiability in Section 1, one has for every $G \in \mathcal{A}$: Either index $G \leq (T-1)/2$, in which case π_G^c is unique and $\phi(\pi_G^c) = \theta(G)$, or index $G > (T-1)/2$, in which case there is a (unique, see below) $H \in \mathcal{K}$ with index $H = T/2$, for which $\pi_G^c = \pi_H^c$ and $\phi(\pi_G^c) = \theta(H)$. Thus, $\mathcal{K}$ satisfies b.

To show (a), let us further describe the set $\mathcal{K}$. For even $T, \mathcal{K} = \mathcal{S}_{\text{fin}}$ are simply the finite mixtures with at most $T/2$ support points. For odd T, these are all G with $\#G \leq (T-1)/2$ support points as well as those with $\#G = (T+1)/2$ and for which one of the support points is 1. Therefore, for even T, identifiability of $\mathcal{K}$ follows from the results in HMZ, and for odd T we give the proof in the Appendix.

It is now tempting to use $\mathcal{K}$ as a "universal" parametric model. Under a small additional regularity condition (compactness of the parameter space, i.e. a lower bound $\delta > 0$ on the possible capture probabilities), the (conditional) MLE $\hat{G} \in \mathcal{K}$ converges to the unique $G \in \mathcal{K}$ which gives $\pi_G^c = \pi_{G_0}^c$, where $G_0 \in \mathcal{A}$ is the true mixing distribution, and $\theta(\hat{G})$ converges to the lower bound $\theta(G) = \phi(\pi_{G_0}^c)$. Thus, one directly obtains an estimate of the lower bound in (1) and of a corresponding mixing distribution. Further, it offers a direct way to construct confidence intervals to the lower bound.

References

Holzmann, H., Munk, A., and Zucchini, W. (2006). On identifiability in capture-recapture models. *Biometrics* **62,** 934–936.

Leroux, B. G. (1992). Consistent estimation of a mixing distribution. *Annals of Statistics* **20,** 1350–1360.

Lindsay, B. (1995). *Mixture Models: Theory, Geometry, and Applications*. Hayward, California: Institute for Mathematical Statistics.

Lindsay, B. G. and Roeder, K. (1993). Uniqueness of estimation and identifiability in mixture models. *The Canadian Journal of Statistics* **21,** 139–147.

Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics* **59,** 1123–1130.

Link, W. A. (2006). Reply to on identifiability in capture-recapture models. *Biometrics* **62,** 936–939.

Mao, C. X. (2007). Estimating population sizes for capture-recapture sampling with binomial mixtures. *Computational Statistics and Data Analysis* **51,** 5211–5219.

Mao, C. X. (2008a). Lower bounds to the population size when capture probabilities vary over individuals. *Australian and New Zealand Journal of Statistics* **50,** 1–10.

Mao, C. X. (2008b). On the nonidentifiability of population sizes. *Biometrics* **64,** 977–981.

Mao, C. X. and Lindsay, B. G. (2007). Estimating the number of classes. *Annals of Statistics* **35,** 917–930.

Mao, C. X., Colwell, R. K., and Chang, J. (2005). Estimating the species accumulation curve using mixtures. *Biometrics* **61,** 433–441.

Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics* **34,** 1265–1269.

Appendix

Proof of Identifiability of $\mathcal{K}$. We only have to give the proof for odd T, and start with the following lemma.

Lemma 1: *If*

$$\sum_{k=1}^{T-1} t_k\, p_k^x (1-p_k)^{T-x} = 0, \quad x = 1, \ldots, T-1,$$

for some $t_k \in \mathbb{R}$ and distinct $p_k \in (0,\,1)$, then it follows that $t_1 = \cdots = t_T = 0$.

Proof. The polynomials $P_x(p) = p^x(1-p)^{T-x}$, $x = 1, \ldots, T-1$, are linearly independent because, except for the normalization, these are the Bernstein polynomials, which are known to be linearly independent. Therefore any nontrivial linear combination, which thus is a nonzero polynomial of degree at most T, has at most T roots. Because, evidently, two of these always equal 0 and 1, there are at most $T-2$ roots within the interval (0, 1). Hence, for different $p_1, \ldots, p_T \in (0,\,1)$, if

$$\sum_{x=1}^{T-1} s_x p_k^x (1-p_k)^{T-x} = 0, \quad k = 1, \ldots, T-1,$$

it follows that the coefficients $s_1 = \cdots = s_T = 0$, all vanish. Introducing matrix notation $P = (P_{k,x}) = (p_k^x(1-p_k)^{T-x})_{k,x=1,\ldots,T-1}$ and $s = (s_1, \ldots, s_{T-1})'$, this is just

$$P \cdot s = 0 \Rightarrow s = 0. \tag{A.1}$$

Relation (A.1) implies that P has full rank, hence so does its transpose P', and we get

$$P' \cdot t = 0 \Rightarrow t = 0 \quad \text{for } t \in \mathbb{R}^{T-1}, \tag{A.2}$$

which is the claim of the lemma.

Suppose that $G, H \in \mathcal{K}$ with $\pi_G^c = \pi_H^c$. Due to strong identifiability of mixing distributions with $(T-1)/2$ support points, we can assume that index $G =$ index $H = T/2$, so that there are $(T+1)/2$ support points for both G and H, one of which is 1 (we assume that after relabeling this is the $(T+1)/2$ support point).

From Lemma 1 in HMZ, there exists an $A > 0$ with

$$\begin{aligned}&\sum_{k=1}^{(T-1)/2} \lambda_{k,G}\, p_{k,G}^x\, (1-p_{k,G})^{T-x} \\ &\quad = A \sum_{k=1}^{(T-1)/2} \lambda_{k,H}\, p_{k,H}^x\, (1-p_{k,H})^{T-x}, \quad x = 1, \ldots, T-1,\end{aligned} \tag{A.3}$$

and

$$\begin{aligned}&\sum_{k=1}^{(T-1)/2} \lambda_{k,G}\, p_{k,G}^T + \lambda_{(T+1)/2,G} \\ &\quad = A\left(\sum_{k=1}^{T-1} \lambda_{k,H}\, p_{k,H}^T + \lambda_{(T+1)/2,H}\right),\end{aligned} \tag{A.4}$$

where $\lambda_{k,G}$, $\lambda_{k,H} \geq 0$ and

$$\sum_k \lambda_{k,G} = \sum_k \lambda_{k,H} = 1. \tag{A.5}$$

Subtracting the r.h.s. from the l.h.s. in (A.3), following the argument in HMZ and using Lemma 1 we find that after relabeling, $p_{k,G} = p_{k,H}$, $k = 1, \ldots, (T-1)/2$, and

$$\lambda_{k,G} = A\, \lambda_{k,H}, \qquad k = 1, \ldots, (T-1)/2.$$

Inserting this into (A.4) and using (A.5) yields the statement of the proposition.

Hajo Holzmann
Institut für Stochastik,
University of Karlsruhe,
Englerstr. 2, D-76128 Karlsruhe, Germany

and

Axel Munk
Institut für Mathematische Stochastik,
Georg-August-University Göttingen,
Maschmühlenweg 8-10, D–37073 Göttingen, Germany
Email: *munk@math.uni-goettingen.de*