

On Identifiability in Capture–Recapture Models

Hajo Holzmann,¹ Axel Munk,^{1,*} and Walter Zucchini²

¹Institut für Mathematische Stochastik, Georg-August-Universität Göttingen,
Maschmühlenweg 8-10, D-37073 Göttingen, Germany

²Institut für Statistik und Ökonometrie, Platz der Göttinger Sieben 5,
D-37073 Göttingen, Germany

*email: munk@math.uni-goettingen.de

SUMMARY. We study the issue of identifiability of mixture models in the context of capture–recapture abundance estimation for closed populations. Such models are used to take account of individual heterogeneity in capture probabilities, but their validity was recently questioned by Link (2003, *Biometrics* **59**, 1123–1130) on the basis of their nonidentifiability. We give a general criterion for identifiability of the mixing distribution, and apply it to establish identifiability within families of mixing distributions that are commonly used in this context, including finite and beta mixtures. Our analysis covers binomial and geometrically distributed outcomes. In an example we highlight the difference between the identifiability issue considered here and that in classical binomial mixture models.

KEY WORDS: Abundance estimation; Beta mixture; Capture–recapture; Finite mixture; Heterogeneity; Identifiability; Uniform mixture.

1. Introduction

Capture–recapture methods are widely used in wildlife abundance estimation and also in fields such as epidemiology and quality control. They have been developed to estimate the size of both closed and open populations, but here we restrict our attention to the former. For terminology and an overview of the methods see, for example, Seber (1982).

An important issue in this context is the fact that, in many applications, the probability of capture/recapture differs among individuals in ways that are caused by factors that are not, or cannot be, observed (see, e.g., Borchers, Buckland, and Zucchini, 2002, Section 11.3). Ignoring such heterogeneity can lead to substantial bias and to inaccurate confidence intervals. One can address this problem by regarding the capture probabilities as realizations of a random variable, from which it follows that the number of animals captured in x out of T capture occasions follows a mixture distribution (Burnham, 1972; Agresti, 1994; Norris and Pollock, 1995, 1996; Pledger, 2000, 2005; Dorazio and Royle, 2003). However, the use of mixture models raises the issue of identifiability (e.g., Huggins, 2001). Indeed Link (2003) concludes “Thus even with very large samples, the analyst will not be able to distinguish among reasonable models of heterogeneity, even though these yield quite distinct inferences about population size.” Furthermore, he gives examples to illustrate this statement, thereby casting doubt on the validity of using mixture models for estimating abundance in the presence of unobserved individual heterogeneity.

The aim of this article is to examine the identifiability issue in more detail. In particular we prove identifiability within the mixture families that are most commonly used in

this application, including finite mixtures and beta mixtures. Thus, so long as the analyst is prepared to assume that the mixture distribution belongs to a certain family then identifiability is not a problem. In this case corresponding parametric estimators such as maximum likelihood estimators can be used. Of course, if the analyst is prepared to make no assumptions about the distribution of probabilities then, as is well known in the context of mixture models in general, Link’s conclusion is correct, and nonparametric estimation methods (cf. Burnham and Overton, 1978) might be preferable.

Proofs of the theorems in this article are provided separately on the *Biometrics* website.

2. Notation and Preliminaries

Suppose that a closed population of unknown size N is sampled on T occasions. We assume that the number of captures X_i of animal i is distributed as binomial $B(T, p_i)$, where p_i is the capture probability of this animal over T independent samples. We assume that the p_i ’s are distributed according to some distribution G on $[0, 1]$. This implies that the probability that an individual is sampled x times is given by

$$\pi_G(x) = \binom{T}{x} \int_0^1 p^x (1-p)^{T-x} dG(p). \quad (1)$$

Let n be the number of animals that were captured at least once, that is, for which $X_i > 0$. Let

$$f_x = \#\{i : X_i = x\}, \quad x = 1, \dots, T.$$

As pointed out by Link (2003), the vector (f_1, \dots, f_T) is multinomially distributed with T cells, n repetitions, and cell probabilities:

$$\pi_G^c = (\pi_G^c(1), \dots, \pi_G^c(T)), \quad \pi_G^c(x) = \frac{\pi_G(x)}{1 - \pi_G(0)},$$

$$x = 1, \dots, T.$$

The probabilities π_G^c are simply the conditional probabilities of the mixture of binomial distributions given that $x \geq 1$. Note that only these conditional probabilities can be estimated from the observations f_x . Consequently the problem of establishing identifiability in this context differs from that of establishing identifiability of the mixing distribution G from the probabilities of a binomial mixture, π_G , in the classical mixture context (cf. Teicher, 1961, 1963; Lindsay, 1995). Here we need to investigate the identifiability of G from the conditional probabilities, π_G^c , given that $x \geq 1$. Note that once this identifiability is settled, G can be consistently estimated (within the given parametric family) by the maximum likelihood estimator \hat{G} , for example. Then $\pi_G(0)$ is consistently estimated by $\pi_{\hat{G}}(0)$, and N by

$$\hat{N} = \frac{n}{1 - \pi_{\hat{G}}(0)}.$$

By embedding the issue of identifiability in such capture–recapture models in the general context of identifiability of finite mixtures from the binomial distribution, it is immediately clear that G cannot be identified within the set of all distributions (for any fixed T), because this does not even hold for the complete (nonconditional) model (1) (cf. Teicher, 1961). However, in this note we show that *within* various commonly used parametric families, G is identifiable. Specifically, we show that within the class of finite mixtures with at most m components (cf. Pledger, 2000), G is identifiable if and only if $2m \leq T$. Furthermore, we give a general criterion for identifiability based on the moments of the mixing distribution. As particular cases, this yields the identifiability of the class of beta distributions if $T \geq 3$, and of the class of uniform distributions on $[0, b]$, $b \leq 1$, if $T \geq 2$.

We stress that establishing identifiability in the context of this application is a more subtle problem than identifiability of binomial mixtures with fixed T , because only the conditional probabilities are available. Finally we remark that some of our results carry over to other distributions of the X_i 's, such as the truncated geometric (cf. Norris and Pollock, 1996).

3. Theory and Examples

DEFINITION 1: In the capture–recapture model (1) we shall call a family \mathcal{G} of distributions on $[0, 1]$ *identifiable* if, for each $G \in \mathcal{G}$, the vector π_G^c uniquely determines G within the class \mathcal{G} , that is, if for $G, H \in \mathcal{G}$,

$$\pi_G^c = \pi_H^c \Rightarrow G = H. \tag{2}$$

LEMMA 1: Let $\pi = (\pi(0), \dots, \pi(T))$ and $\rho = (\rho(0), \dots, \rho(T))$ be two probability vectors on $\{0, \dots, T\}$, and let π^c and ρ^c be the conditional probability vectors on $1, \dots, T$, given that $x \geq 1$. Then

$$\pi^c = \rho^c \Leftrightarrow \text{there is an } A > 0 : \pi(x) = A\rho(x), \quad x = 1, \dots, T.$$

It is known (e.g., Link, 2003) that for identifiability to be possible, 0 has to be excluded from the support of the distributions in \mathcal{G} . Indeed, let G be any distribution on $[0, 1]$ and

consider $H = \lambda\delta_0 + (1 - \lambda)G$, $\lambda \in (0, 1)$, where δ_0 is the point mass at 0. Because $\pi_{\delta_0}(0) = 1$, we have that $\pi_H(x) = (1 - \lambda)\pi_G(x)$ for $x = 1, \dots, T$. From Lemma 1, it follows that $\pi_G^c = \pi_H^c$. Thus in the following we will concentrate on distributions G with support in $(0, 1]$. This simply means that the whole population is in principle observable.

Teicher (1961) observed that the probabilities (1) can be expressed in terms of the moments of the mixing distribution G . In fact, we have that

$$\pi_G(x) = \binom{T}{x} \sum_{k=x}^T (-1)^{k-x} \binom{T-x}{k-x} m_G(k), \quad x = 1, \dots, T, \tag{3}$$

where $m_G(k) = \int_0^1 t^k dG(t)$ is the k th moment of G . For our problem this implies

THEOREM 1: For two distributions G, H on $(0, 1]$, $\pi_G^c = \pi_H^c$ implies that there is an $A > 0$ such that

$$m_G(x) = A m_H(x), \quad x = 1, \dots, T. \tag{4}$$

Therefore if there exist no two different $G, H \in \mathcal{G}$ such that (4) holds, then \mathcal{G} is identifiable.

Example 1. The beta distribution $B(p, q)$, $p, q > 0$ was used as a mixing distribution by Dorazio and Royle (2003). We show that this family is identifiable if $T \geq 3$. The x th moment is given by

$$m_{p,q}(x) = \frac{(p+x-1) \cdot \dots \cdot p}{(p+q+x-1) \cdot \dots \cdot (p+q)}.$$

From $m_{p,q}(x) = A m_{p',q'}(x)$, $x = 1, 2, 3$, and some $A > 0$, it follows that

$$\frac{(p+i)}{(p+q+i)} = \frac{(p'+i)}{(p'+q'+i)}, \quad i = 1, 2.$$

Straightforward algebra now shows that $p = p'$ and $q = q'$. Let us briefly illustrate the difference between identifiability in the capture–recapture context and in the context of usual binomial mixtures. From the moment criterion in Teicher (1961), it follows that beta mixtures of the binomial distribution are identifiable for $T = 2$. However, it is easily seen that relation $m_{p,q}(x) = A m_{p',q'}(x)$, $x = 1, 2$ is satisfied for $p = 1, q = 2, p' = 3, q' = 4$, and $A = 7/9$, and hence identifiability in the capture–recapture context does not hold for $T = 2$.

Example 2. The uniform distribution on $(0, b]$ was considered as a mixing distribution in Pledger (2005). The first two moments are given by $m_b(1) = b/2$ and $m_b(2) = b^2/3$. From these expressions and Theorem 1 it is simple to see that for $T \geq 2$ the uniform distribution is identifiable.

Now let us consider the class of finite mixing distributions with at most m support points

$$\mathcal{G}_m = \left\{ G = \sum_{k=1}^m \lambda_k \delta_{p_k}, \quad \lambda_k \geq 0, \quad \sum_k \lambda_k = 1, \quad p_k \in (0, 1] \right\}.$$

Note that $\mathcal{G}_{m-1} \subset \mathcal{G}_m$, because the λ_k 's are allowed to be zero. Theorem 1 also applies to finite mixtures, however, it is not

helpful in this case. Therefore we use a direct method of proof for the following:

THEOREM 2: *The class \mathcal{G}_m is identifiable if and only if $2m \leq T$.*

Pledger (2000) used finite mixtures to model population heterogeneity. She observed that the condition $2m \leq T$ is necessary for identifiability. In fact, the class \mathcal{G}_m has $2m - 1$ parameters, and these have to be identified from the $T - 1$ variable probabilities. Thus $2m \leq T$ is a necessary and sufficient condition. An inspection of the proof of Theorem 2 (given in the online Supplementary Materials) shows that the same arguments apply if the outcomes X_i follow a discrete distribution which, as a function of the parameter, is a Čebyšev system (cf. Karlin and Studden, 1966) with a joint zero outside the interval $(0, 1]$. An example is the truncated geometric distribution (for which $P(X_i = x) = p_i(1 - p_i)^x$, $1 \leq x \leq T$) used in Norris and Pollock (1996), to model population heterogeneity with behavioral response to capture.

4. Supplementary Materials

Proofs can be found at the *Biometrics* website <http://www.tibs.org/biometrics>.

ACKNOWLEDGEMENTS

Hajo Holzmann and Axel Munk gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft, grant MU 1230/8–1. We would also like to thank the referees for helpful comments.

REFERENCES

- Agresti, A. (1994). Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics* **50**, 494–500.
- Borchers, D. L., Buckland, S. T., and Zucchini, W. (2002). *Estimating Animal Abundance: Closed Populations*. London: Springer.
- Burnham, K. P. (1972). Estimation of population size in multinomial capture-recapture studies when capture probabilities vary among animals. Ph.D. Thesis, Oregon State University, Corvallis.
- Burnham, K. P. and Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrics* **65**, 625–633.
- Dorazio, R. M. and Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **59**, 351–364.
- Huggins, R. (2001). A note on the difficulties associated with the analysis of capture-recapture experiments with heterogeneous capture probabilities. *Statistics and Probability Letters* **54**, 147–152.
- Karlin, S. and Studden, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. New York: Wiley.
- Lindsay, B. (1995). *Mixture Models: Theory, Geometry, and Applications*. Hayward, California: Institute for Mathematical Statistics.
- Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics* **59**, 1123–1130.
- Norris, J. L. and Pollock, K. H. (1995). A capture-recapture model with heterogeneity and behavioural response. *Environmental and Ecological Statistics* **2**, 305–313.
- Norris, J. L. and Pollock, K. H. (1996). Nonparametric MLE under two closed capture-recapture models with heterogeneity. *Biometrics* **52**, 639–649.
- Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* **56**, 434–442.
- Pledger, S. (2005). The performance of mixture models in heterogeneous closed population capture-recapture. *Biometrics* **61**, 868–876.
- Teicher, H. (1961). Identifiability of mixtures. *Annals of Mathematical Statistics* **32**, 244–248.
- Teicher, H. (1963). Identifiability of finite mixtures. *Annals of Mathematical Statistics* **34**, 1265–1269.
- Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters*, 2nd edition. London: Charles Griffin.

Received March 2005. Revised December 2005.

Accepted December 2005.

The author replied as follows:

In *Biometrics* **59**, 1123–1130, I wrote

The purpose of this note is to draw attention to a problem inherent to all attempts to model heterogeneity in detection probability. The problem has not gone entirely unnoticed previously, but its extreme consequences appear not to have been fully recognized. The problem is this: that while it may be possible to discern heterogeneity in detection probability, it is likely that an analyst will be unable to distinguish between reasonable descriptions of the heterogeneity, even when these alternative descriptions lead to vastly different inferences about population sizes.

In the first half of my article, I illustrated these points in the context of the closed population heterogeneity model M_h , based on T sampling occasions. The sufficient statistic is $\mathbf{f} = (f_1, f_2, \dots, f_T)'$, where f_x is the number of individuals captured exactly x times; $n = \sum_{x=1}^T f_x$ is the number of distinct individuals observed.

I showed that ruling out mass on or near zero in the distribution of detection probabilities was not sufficient to ensure identifiability of population size N (contrary to a suggestion of Huggins, 2001). To do so, I constructed distributions G and H on $(\delta, 1]$, $\delta > 0$, with the property that the first T moments of G are proportional to, but distinct from, the corresponding moments of H . It is easily verified that the sampling distribution of \mathbf{f} is identical for such G and H , but that the indicated proportions of individuals not detected are different, so that N is not identifiable.

Holzmann, Munk, and Zucchini (2006, subsequently HMZ) have shown in their Theorem 1 that the sufficient condition I used to construct the example is also a necessary condition. Thus if we can restrict our attention to a family of distributions \mathcal{G} in which distinct distributions G and H cannot satisfy

the moment condition, there is no problem of nonidentifiability of N .

This is an interesting observation. HMZ have used their theorem to show that N is identifiable if it is known that the distribution of detectability is in the beta family, or within the class of finite mixtures with at most m components, provided that $2m \leq T$. While these specific results can perhaps be established by other means (e.g., for the beta distribution by K. P. Burnham in his 1972 thesis), the result is nonetheless interesting as a general mathematical characterization of the problem of nonidentifiability under M_h .

I am happy to have the opportunity to respond to HMZ. Their commentary includes two common reactions to my 2003 paper. First, that given sufficiently strong assumptions about the mixture family, identifiability is not a problem. This is true, but not very reassuring. Without such strong and untestable assumptions, identifiability *is* a problem. Pledger (2005) has commented with regard to my paper that “the impact of these findings on capture–recapture analysis has yet to be determined” but concluded with the optimistic statement that “despite all the problems... heterogeneous models are working well for many data sets.” I am somewhat more confident about the implications of my findings, and less sanguine about whether “heterogeneous models are working well.” Thus in the first section of this response I will discuss practical implications of HMZ’s Theorem 1, reiterating that nonidentifiability among families of distributions means that the data cannot inform a fundamental and noninnocuous modeling decision.

Another reaction to my 2003 paper has been that if parametric models cannot be used with confidence, then perhaps we ought to use nonparametric methods instead. HMZ wrote

Of course, if the analyst is prepared to make no assumptions about the distribution of probabilities then, as is well-known in the context of mixture models in general, Link’s conclusion is correct, and nonparametric estimation methods (cf. Burnham and Overton, 1978) might be preferable.

The second part of this response addresses this suggestion through consideration of the jackknife estimator of Burnham and Overton (1978). I conclude that the availability of nonparametric methods provides no solution to the problem of identifiability.

1. Practical Implications of Holzmann et al.’s Theorem 1

Interesting as HMZ’s Theorem 1 is from a mathematical perspective, I believe it offers little comfort to analysts of capture–recapture data with heterogeneous detection probabilities. The choice of a family of distributions for heterogeneity is typically made on the basis of convenience, rather than because of compelling prior knowledge. If it is argued that we may try various models of heterogeneity, and let some model selection criterion such as Akaike’s Information Criterion (AIC) sort through them, then it is essential that identifiability of N prevails across the set of candidate models, and not simply within models.

Thus, for instance, when $T = 4$, N is identifiable with heterogeneity described by a two-point mixture or by a beta distribution. N is identifiable under either family; however, N is

Table 1
Distribution of f given n under alternative models of heterogeneity

x	$E(f_x/n)$						
	1	2	3	4	5	6	7
Logit-normal	0.319	0.199	0.143	0.112	0.092	0.077	0.059
Beta	0.320	0.196	0.144	0.113	0.092	0.076	0.060

not identifiable in the larger family consisting of the union of the two families. There exist two-point mixtures and beta distributions yielding identical sampling distributions of \mathbf{f} . An example is the two-point mixture placing mass of $w = 3/4$ on $p_1 = 1/4$ and $1 - w$ on $p_2 = 3/4$, and the beta distribution with parameters $a = 1/2$ and $b = 3/2$; the first four moments of the beta distribution are precisely $2/3$ of the corresponding moments of the two-point mixture. The frequencies f_x , $x = 1, 2, 3, 4$, occur in the exact same proportions 28:18:12:7 under both models, but the expected number of individuals detected is $0.508N$ under the beta model, and $0.762N$ under the two-point mixture model. Thus the beta model indicates a population size 50% larger than the two-point mixture model. It is nice to know that N is identifiable within the beta family and within the two-point mixture family, but the data give no basis for choosing between the models; inference regarding N can only be made on the basis of an untestable model assumption.

Consider also the frequency distribution for \mathbf{f} given in Table 1: the logit-normal frequencies are based on the assumption that $\text{logit}(p)$ is normal with mean -1.75 and variance 4; the beta frequencies arise from assuming that p has a $\beta(0.256, 1.135)$ distribution. These frequency distributions were obtained by numerical minimization of the Kullback–Leibler distance from the logit-normal-induced distribution for \mathbf{f} to the corresponding family of distributions induced by the $\beta(a, b)$ mixing distribution. Under the logit-normal model, the expected number of individuals detected is $0.6072N$, while under the beta model it is $0.4397N$. I note that the ratios of the first seven moments of the two mixing distributions (logit-normal/beta) are nearly but not precisely constant, ranging from 1.371 to 1.382. Two possibilities exist: (1) that the joint family (consisting of logit-normal and beta distributions) is not identifiable, and that the nonconstant ratio I obtained was the result of round-off error; or (2) that the union of logit-normal and beta families is identifiable, but very nearly not so. I would guess that the second explanation is the correct one; regardless, the example shows that unless one has compelling grounds for believing that the mixing distribution is of the beta family, and not of the logit-normal family (or vice versa), then the problem of nonidentifiability of N must be taken seriously, this even when the number of sampling occasions is large relative to the number of model parameters.

2. Nonparametric Estimation Methods

There is current among some analysts the notion that nonparametric methods can be applied for free, with no

assumptions entailed. D. H. Johnson, in a paper aptly titled “Statistical sirens: The allure of nonparametrics” (1995), cited examples of scientific papers in which this misapprehension is explicitly stated:

The main advantage of nonparametric methods over their parametric counterparts is the *absence of assumptions regarding the distribution underlying the observations.*

(Potvin and Roff, loc. cit., Johnson’s emphasis). HMZ’s comment that should an analyst be unwilling to make parametric assumptions, “nonparametric methods might be preferable,” could nurture this misconception. The fact is that nonparametric methods do make assumptions, but that the assumptions tend to be less restrictive, or at least appear so, with the consequence that they may be less vigorously scrutinized.

Nonparametric methods are typically applicable to larger families of distributions than their parametric counterparts. If the problem of nonidentifiability can only be resolved by restrictions to specific parametric families (as HMZ have demonstrated) then it is hard to see how omnibus nonparametric methods might be preferable, except to those who would sweep assumptions under the rug.

Consider the popular jackknife estimators of Burnham and Overton (1978, 1979) referenced by HMZ. The estimators were developed using the bias reduction properties of the jackknife procedure, beginning with the biased estimator n of N . Burnham and Overton’s k th order estimator \hat{N}_{jk} is a linear combination of jackknife estimators having order $j = 1, 2, \dots, k$, and has bias of order $1/T^k$. The greater bias reduction for larger k is understood as coming at the expense of greater sampling variability; this observation stands as the basis for a rather ad hoc procedure for selecting k , described in Burnham and Overton (1978) and implemented in the widely used program CAPTURE (White et al., 1978).

Using HMZ and my notation, Burnham and Overton’s (1978) development of \hat{N}_{jk} assumes that

$$E(n) = N + a_1/T + a_2/T^2 + \dots, \tag{1}$$

where a_1, a_2, \dots are constants, functionally independent of T , but not of N . This assumption defines a family of distributions for which the jackknife estimator is appropriate: it is the family \mathcal{G}_J of distributions G on $[0, 1]$ with the property that we may write

$$\pi(0) = \frac{N - E(n)}{N} = \int_0^1 (1 - p)^T dG(p) = \sum_{j=1}^{\infty} (b_j/T^j) \tag{2}$$

for some sequence of constants $b_j = -a_j/N$, with (2) holding for all positive integers T (or perhaps, for all positive integers $T \geq T_0 > 0$).

I would guess that few analysts have much of an intuition for the family \mathcal{G}_J , or could name a distribution in the family, or even state with confidence that such distributions exist. While an investigation of the family \mathcal{G}_J seems beyond the scope of this response, it seems worthwhile to point out that the family is limited, and perhaps even rather severely (proofs of subsequent comments are available from the author). For instance, if $\pi(0)$ approaches zero too rapidly as $T \rightarrow \infty$, condition (2) cannot be satisfied. This is the case

if $\lim_{T \rightarrow \infty} T^k \pi(0) = 0$, for all integers $k > 0$. The finite mixture models have this property, hence are not members of the family \mathcal{G}_J .

Suppose that $G(p)$ corresponds to a $\beta(a, b)$ distribution. It can be shown that G is not in the family \mathcal{G}_J unless a is an integer. If a is an integer, (2) is demonstrable for values $T \geq a + b$. Why, one might reasonably ask, would one choose to allow integer values of a , but to rule out all others? Indeed, why would one choose the family \mathcal{G}_J as the basis of inference?

The apparently adequate performance of Burnham and Overton’s jackknife in empirical evaluations (Greenwood et al., 1985; Manning et al., 1995) is likely due to its relation, for large T , to an estimator based on extrapolation of an interpolating polynomial. Burnham and Overton (1979) state that

$$\lim_{T \rightarrow \infty} \hat{N}_{J,k} = n + \sum_{i=1}^k (-1)^{i+1} \binom{k}{i} f_i. \tag{3}$$

Now consider the unique polynomial of degree k , $g_k(x)$, defined as passing through the points (i, f_i) for $i = 1, 2, \dots, k$. For each of these values i , the polynomial tells how many individuals were captured precisely i times: $g_k(i) = f_i$. We may hazard a guess that the polynomial will do an adequate job of predicting f_0 , the number of animals captured zero times. We might then estimate N by $n + g_k(0)$. It turns out that $g_k(0)$ is the same as the second term on the right-hand side of (3).

It might be argued that this sort of curve-fitting and extrapolation is a reasonable thing to do. The question remains, “under what conditions?” Are analysts motivated by the belief that the assumption defining family \mathcal{G}_J is appropriate for their specific problem? Or is there an unstated assumption that the jackknife is useful in vaguely defined broader circumstances? In either case, are we to be content with an estimator of N merely on the basis that it has smaller bias than the naïve estimator n ?

The message of my 2003 paper was that without strong assumptions, f_1, f_2, \dots, f_T tell us essentially nothing about f_0 . HMZ have clarified how strong the assumptions need to be, and for this I applaud them. Given the message “strong assumptions are needed,” the response “fewer/vaguer assumptions is the solution” seems unwarranted. Easily assailable but clearly articulated assumptions ought always to be preferable. Nonparametric estimation of N is not the solution to the identifiability problem.

3. Conclusions

Where does this leave us? Unmodeled heterogeneity is a shipwreck. Indeed, without specific parametric models for the heterogeneity in p , we find ourselves in circumstances quite similar to those considered by Kiefer and Wolfowitz (1956), who demonstrated that maximum likelihood estimates of parameters of interest may be asymptotically biased, and badly so, if the number of nuisance parameters (in this case, individual p ’s) is allowed to increase without bound. In fact, the situation is worse: no estimate of N is available under model M_h without recourse to untestable model assumptions.

HMZ's Theorem 1 implies that we may have as much confidence in our analyses of model M_h as we have confidence in the selected distribution of detectability. Rather than relying on untestable assumptions, it seems advisable that we attempt to identify and control for sources of heterogeneity through appropriate covariate analyses.

REFERENCES

- Burnham, K. P. (1972). Estimation of population size in multiple capture-recapture studies when capture probabilities vary among animals. Ph.D. Thesis, Oregon State University, Corvallis.
- Burnham, K. P. and Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**, 625–633.
- Burnham, K. P. and Overton, W. S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology* **60**, 927–936.
- Greenwood, R. J., Sargeant, A. B., and Johnson, D. H. (1985). Evaluation of mark-recapture for estimating striped skunk abundance. *Journal of Wildlife Management* **49**, 332–340.
- Holzmann, H., Munk, A., and Zucchini, W. (2006). On identifiability in capture–recapture models. *Biometrics*, this volume.
- Huggins, R. (2001). A note on the difficulties associated with the analysis of capture-recapture experiments with heterogeneous capture probabilities. *Statistics and Probability Letters* **54**, 147–152.
- Johnson, D. H. (1995). Statistical sirens: The allure of non-parametrics. *Ecology* **76**, 1998–2000.
- Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Annals of Mathematical Statistics* **27**, 887–906.
- Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics* **59**, 1123–1130.
- Manning, T., Edge, W. D., and Wolff, J. O. (1995). Evaluating population-size estimators: An empirical approach. *Journal of Mammalogy* **76**, 1149–1158.
- Pledger, S. (2005). The performance of mixture models in heterogeneous closed population capture-recapture. *Biometrics* **61**, 868–876.
- White, G. C., Burnham, K. P., Otis, D. L., and Anderson, D. R. (1978). *Users' Manual for Program CAPTURE*. Logan: Utah State University Press.

William A. Link
USGS Patuxent Wildlife Research Center
Laurel, Maryland 20708
U.S.A.