Identifiability of Population Size from Capture-Recapture Data with Heterogeneous Detection Probabilities

ALESSIO FARCOMENI
Sapienza University of Rome
alessio.farcomeni@uniroma1.it

LUCA TARDELLA
Sapienza University of Rome
luca.tardella@uniroma1.it

February 4, 2009

Abstract

In this note we show how it is possible to overcome identifiability issues related to the capture-recapture model $\mathcal{M}_h$, in which capture probabilities are allowed to be heterogeneous. Link (2003) highlighted the non-identifiability of conditional likelihood parameterization and concluded that one can not draw valid inference on the unknown population size unless strong untestable assumptions are imposed on the specific form of F. We show that the complete likelihood based on the moments of F and the unknown population size leads in fact to an identifiable model. The seeming contradiction with the often invoked equivalence of those two likelihood approaches is thus resolved.

Key Words: Binomial Mixture; Capture-Recapture; Conditional Likelihood; Identifiability; Complete Likelihood.

1 Introduction

Many authors have been investigating the inferential difficulties of one of the simplest, yet hardest, statistical models for capture recapture data, traditionally denoted as $\mathcal{M}_h$. In model $\mathcal{M}_h$ there are S capture occasions and the capture probability of each animal is constant over all occasions but is allowed to vary from animal to animal. The individual probability is assumed to be distributed according to some unknown F. When no parametric assumption is made on F there are different available methodologies ranging from the more traditional Burnham & Overton (1978) and Chao (1989) to more recent likelihood-based solutions such as Norris & Pollock (1996) or the Bayesian approaches in Basu & Ebrahimi (2001) and Tardella (2002). However, none of the above can be considered a conclusive solution of the problem. In fact, in some recent papers different cautionary notes have been written to discuss on troublesome effects of the presence of heterogeneity (Huggins & Yip, 2001; Hwang & Huggins, 2005) and warn against the identifiability issues of the model (Huggins, 2001; Link, 2003; Holzmann et al., 2006; Mao, 2008). In this short note we point out that those identifiability issues are in fact related to the particular choice of likelihood adopted. In fact while Huggins (2001), Link (2003) and Holzmann et al. (2006) use the conditional likelihood we will show that the use of the complete likelihood overcomes the main issue. It is usually argued that those likelihoods are asymptotically equivalent (Sanathanan, 1972), hence working with either one should not matter eventually. However, the famous result of Sanathanan (1972, Theorem 2) is in fact valid under a technical condition which we will show is not met in the case of $\mathcal{M}_h$ model if no restriction is made on F. The paper is organized as follows: in Section 2 we fix the notation and model setup, in Section 3 we review the nonidentifiability arguments of Link (2003) for conditional likelihood and moment based parameterization of F and prove the identifiability of the same model under parameterization with the complete likelihood. In Section 4 we give some brief concluding remarks.

2 Setup and notation

We start by writing down the complete probability structure of the distribution of the underlying sequence of binary capture histories $\boldsymbol{x}_i = (x_{i1}, ..., x_{iS})$ $(i = 1, 2, ..., N)$ which can be observed in a capture-recapture experiment consisting of S trapping occasions for inferring the unknown population size N. Some of the capture histories (let us say the first n) are actually observed, while the remaining ones are not and they correspond to the $N-n$ animals for which S consecutive zeros are recorded. All the capture histories are summarized in the binary $N \times S$ matrix $\boldsymbol{X}$ up to the row index permutation. Model $\mathcal{M}_h$ assumes independent and identically distributed (i.i.d) binary capture outcomes for each animal

$$Pr(X_{ik}=1|p_i) = p_i \quad \forall k = 1, ..., S$$

with subject-specific capture probability p_i that is in turn assumed i.i.d. from an unknown distribution F with support in $(0,1)$. Hence,

$$Pr\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_N | N, F\right) = \frac{N!}{\left(N - \sum_{k=1}^S n_k\right)! \prod_{k=1}^S n_k!} \prod_{k=0}^S P_k(F)^{n_k} \tag{1}$$

where n_k denotes the number of subjects which have been captured exactly k times and

$$P_k(F) = \binom{S}{k} \int_0^1 p^k (1-p)^{(S-k)} F(dp), \qquad k = 0, 1, ..., S. \tag{2}$$

The distribution in (1) of the observable as well as unobservable sequences conditionally on N and F is equivalent to that of the capture counts n_k ($k = 1, ..., S$), which are in fact sufficient statistics for the model parameters (N, F). Notice that $n = \sum_{k=1}^S n_k$ and $n_0 = N - n = N - \sum_{k=1}^S n_k$. Also, from (1) we can recombine the *complete likelihood* $L(N, F)$ as follows

$$\begin{aligned}
Pr\left(\boldsymbol{X}|N, F\right) &= \frac{\left(\sum_{k=1}^S n_k\right)!}{\prod_{k=1}^S n_k!} \frac{\prod_{k=1}^S P_k(F)^{n_k}}{(1-P_0(F))^{\sum_{k=1}^S n_k}} \times & (3)\\
&\times \frac{N!}{\left(N - \sum_{k=1}^S n_k\right)! \left(\sum_{k=1}^S n_k\right)!} P_0(F)^{N - \sum_{k=1}^S n_k} (1 - P_0(F))^{\sum_{k=1}^S n_k} & (4)\\
&= \frac{n!}{\prod_{k=1}^S n_k!} \prod_{k=1}^S P_{(c),k}(F) \times \frac{N!}{(N-n)!n!} P_0^{(N-n)} (1-P_0)^n \\
&= L_{(c)}(F) \times L_{(r)}(N, F) & (5)\\
&= L(N, F) & (6)
\end{aligned}$$

where the so called *conditional likelihood* $L_{(c)}(F)$ is factored out and only conveys information about the unknown F while the remaining part of the expression $L_{(r)}(N,F) = g(N, P_0(F))$ is the only one which is functionally dependent on the parameter of interest N and depends on F only through $P_0(F)$. Moreover the conditional likelihood has been reparameterized in terms of the conditional probabilities

$$P_{(c),k}(F) = \frac{P_k(F)}{1 - P_0(F)} = Pr\left\{ \boldsymbol{x} : \sum_{j=1}^{S} x_j = k \,\middle|\, \sum_{j=1}^{S} x_j > 0 \right\}, \quad k = 1, 2, ..., S$$

From the seminal work of Sanathanan (1972), the classical inferential approach, followed also in Huggins (2001), Link (2003) and Holzmann et al. (2006), consists of breaking down the estimation process into two consecutive steps: in the first step one estimates F, or its corresponding parameters, through the multinomial conditional likelihood $L_{(c)}(F)$ obtaining $\hat{F}$ and then plugs it in $P_0(F)$ which is the only feature of F involved in $L_{(r)}(N,F)$. This allows us to estimate the parameter of interest N through a simple binomial structure with an unknown size parameter yielding the so-called Horvitz-Thompson estimator which can be written in the explicit form $\hat{N}_{P_0(\hat{F})} = \frac{n}{1-P_0(\hat{F})}$. In Link (2003) it was neatly pointed out that if no assumption about F is made then this procedure has the drawback of relying on a likelihood function, $L_{(c)}(F)$, for which $P_0(F)$ and hence N are strongly unidentifiable.

There is a convenient alternative parameterization which we now discuss. The multinomial structure of the likelihood has been written down in terms of the $S+1$ binomial mixture probabilities $\boldsymbol{P}(F) = (P_0(F), P_1(F), ..., P_S(F))$ which live in an S-dimensional subspace of the simplex constrained by (2).

It has long been recognized in binomial mixture literature (Rolph, 1968; Lindsay, 1995) and more recently in the capture-recapture context (Yoshida et al., 1999; Tardella, 2002; Link, 2003) that the probabilities $\boldsymbol{P}(F)$ are in one-to-one correspondence with the first S moments of F through the relation

$$P_k(F) = \sum_{r=k}^{S} \frac{S!}{k!(r-k)!(S-r)!} (-1)^{(r-k)} m_r(F) \qquad k = 0, 1, ..., S$$

where $m_r(F) = \int_{[0,1]} p^r F(dp)$ is the ordinary r-th moment of F. The different parametrization in (5), involving alternative $S+1$ parameters

$(P_0(F), \boldsymbol{P}_{(c)}(F))$ and lying in a S-dimensional space, is also in one-to-one correspondence with the original $\boldsymbol{P}(F)$.

We will now review the main arguments leading to the non-identifiability of $L_{(c)}(F)$ and see how the problem can be overcome by relying on the original complete likelihood $L(N, F)$ as in (6), with the parameterization based on the first S moments of F.

3 Non identifiability of the conditional and identifiability of the marginal likelihood

First of all we recall that the first source of nonidentifiability of the nonparametric mixture of binomials has already been clarified by Rolph (1968), where it is clearly stated that the identifiable features of F must be in one-to-one correspondence with the first S moments. This means that when no parametric constraints are considered there is a class of (likelihood) equivalence related to the so called moment class i.e. the class of all distributions F which have the same first S moments. The function F itself is not wholly identifiable.

Link (2003) showed that the use of the conditional likelihood for estimating $P_0(F)$ suffers from a different identifiability issue that results in an invalid inference for the unknown population size of the $\mathcal{M}_h$ model. He showed how, for any mixing distribution F with corresponding $P_0(F) < 1$, there is an infinite collection of other mixing distributions $\{G_\gamma = \gamma F + (1 - \gamma)\delta_0; \gamma \in (0, 1)\}$ for which $L_{(c)}(F) = L_{(c)}(G_\gamma)$, where δ_0 is the degenerate distribution at 0. The proof was easily carried out by noting that the first S moments of G_γ were linearly related to those of F by the relation

$$m_k(G_\gamma) = \gamma m_k(F) \qquad k = 1, ..., S$$

and hence

$$P_k(G_\gamma) = \gamma P_k(F) \qquad k = 1, ..., S$$

so that

$$1 - P_0(G_\gamma) = \sum_{k=1}^{S} P_k(G_\gamma) = \sum_{k=1}^{S} \gamma P_k(F) = \gamma(1 - P_0(F))$$

and hence the ratios defining the corresponding conditional probabilities are identical:

$$P_{(c),k}(G_\gamma) = \frac{P_k(G_\gamma)}{1 - P_0(G_\gamma)} = \frac{P_k(F)}{1 - P_0(F)} = P_{(c),k}(F).$$

Link (2003) concluded that there is no way of making inference on model $\mathcal{M}_h$ unless one is willing to overcome the above indeterminacy of $P_0(F)$ due to the flatness of the conditional likelihood corresponding to the equivalence class by admitting untestable model assumptions on the F distribution. We will argue now that his conclusion is somehow overstated and a fairer conclusion would be simply that one cannot rely on the usual two-step procedure based on the conditional likelihood for estimating $\hat{P}_0$ and on the Horvitz-Thompson rule $\hat{N}_{\hat{P}_0} = \frac{n}{\hat{P}_0}$ to make inference on N.

First of all let us show in Figure 1 a simple numerical check on the fact that the complete likelihood does not appear to have flat regions for $P_0 \to 1$. We take the same numerical example in Link (2003) and show that when γ runs in $(0, 1)$ while the conditional likelihood is flat the marginal likelihood is not. Note also the fact that when γ approaches 1 the complete likelihood starts decreasing eventually.

Figure 1 is in seeming contradiction with the results of Sanathanan (1972), who showed that both approaches to inference are asymptotically equivalent. The fundamental relationship between the conditional probabilities of positive counts and the zero-count unconditional probability, both depending on a common parameter vector, then suggested Sanathanan (1972) to solve the inferential problem via the conditional likelihood for simplicity reasons. In this recapture context the asymptotics is understood as the behaviour of the model as $N \to \infty$. We now resolve this seeming contradiction by noting that the regularity conditions invoked by Sanathanan (1972) to prove the asymptotic equivalence of conditional likelihood and complete likelihood are in fact not met in the general case of model $\mathcal{M}_h$ when no assumption is made on F. We report the original result (Sanathanan, 1972, Theorem 2, p. 147) with a slightly different notation.

Theorem 1. *[Sanathanan (1972)] Suppose that N_0 and λ_0 are the true parameter values and both the following conditions hold:*

- *(A1) At every admissible value of λ the functions $\lambda \to P_k(\lambda)$ admit continuous first-order partial derivatives*

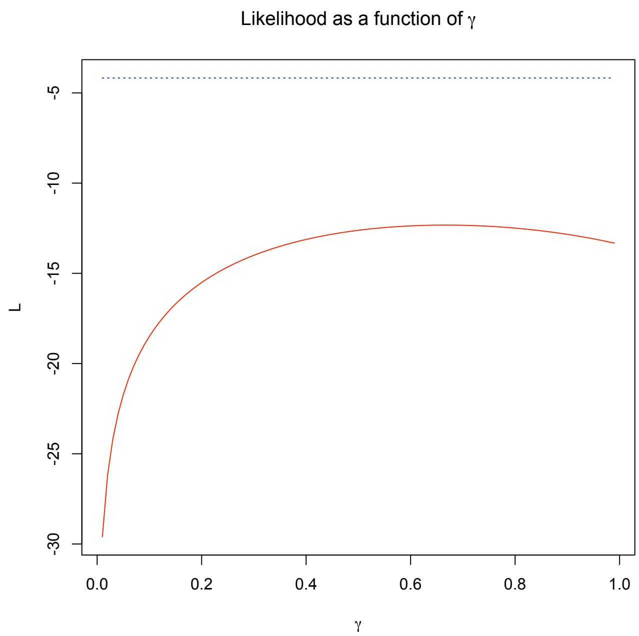

Figure 1: Complete (red solid line) and conditional (blue dashed) likelihood mapped as a function of γ.

- *(A2) Given a $\delta > 0$ it is possible to find an $\varepsilon > 0$ such that*

$$\inf_{|\lambda^* - \lambda_0| > \delta} \sum_{k=1}^{S} P_{(c),k}(\lambda_0) \log \frac{P_{(c),k}(\lambda_0)}{P_{(c),k}(\lambda^*)} > \varepsilon.$$

Then

1. $\left(\frac{\hat{N}_c}{N_0}, \hat{\lambda}_c, P_0(\hat{\lambda}_c)\right) \overset{N_0 \to \infty}{\longrightarrow} (1, \lambda_0, P_0(\lambda_0))$
2. $\left(\frac{\hat{N}}{N_0}, \hat{\lambda}, P_0(\hat{\lambda})\right) \overset{N_0 \to \infty}{\longrightarrow} (1, \lambda_0, P_0(\lambda_0))$
3. *when* $N_0 \to \infty$, $\frac{(\hat{\lambda}_c - \lambda_0)}{\sqrt{N_0}} \approx \frac{(\hat{\lambda} - \lambda_0)}{\sqrt{N_0}} \approx N(0, \Sigma)$

with $\Sigma = \begin{bmatrix} A & a_0^T \\ a_0 & a_{00} \end{bmatrix}$, *with* $A = (a_{ij})$ *defined through* $a_{ij} =$

$\sum_{k=1}^{S} P_k^{-1} \frac{\partial P_k}{\partial \lambda_{0,i}} \frac{\partial P_k}{\partial \lambda_{0,j}}$, $a_0 = (a_{10}, ..., a_{S0})$ *defined through* $a_{k0} = -P_0^{-1} \frac{\partial P_0}{\partial \lambda_{0,k}}$ *and* $a_{00} = \frac{1-P_0(\lambda_0)}{P_0(\lambda_0)}$.

The reason why this theorem can not be used in model $\mathcal{M}_h$ when no assumption is made on F is that $(A2)$ does not hold and it can be just shown with the arguments in Link (2003). Indeed, take $\lambda_0 = (m_1(F), ..., m_S(F))$ and $\lambda^* = (m_1(G_\gamma), ..., m_S(G_\gamma))$. For all $\delta > 0$ one can always take $\gamma > 0$ small enough such that from (7) it follows that

$$\inf_{|\lambda^* - \lambda_0| > \delta} \sum_{k=1}^{S} P_{(c),k}(\lambda_0) \log \frac{P_{(c),k}(\lambda_0)}{P_{(c),k}(\lambda^*)} = 0.$$

Notice that Sanathanan (1972) borrowed the condition (A2) from Rao (1965) which in fact refers to it as *strong identifiability condition*.

We now formally prove that using the complete likelihood overcomes the identifiability problem and can then allow a valid inference on N.

Theorem 2. *Assume $N > 0$ and $(m_1, \ldots, m_S)$ live in the interior of the space of moments of the class of F distributions with support on $[0, 1]$. The complete likelihood is identifiable within the class of F distributions with unique first S moments.*

Proof: Suppose (N, F) and (N', F') are such that $(N, m_1, \ldots, m_S)$ are not all equal to $(N', m_1', \ldots, m_S')$. Then we will show that there exists at least one binary data configuration which yields the likelihood ratio to be different from 1.

When $(m_1, \ldots, m_S) \neq (m_1', \ldots, m_S')$ it can be shown that $(N, P_0(F), \ldots, P_S(F))$ are not all equal to $(N', P_0(F'), \ldots, P_S(F'))$ since $P_0(F), \ldots, P_S(F)$ is one-to-one with $m_1, \ldots, m_S$. Moreover, we have

$$\frac{L(N, F)}{L(N', F')} = \frac{N!(N' - \sum_{k=1}^{s} n_k)!}{N'!(N - \sum_{k=1}^{s} n_k)!} \prod_{k=0}^{S} \left[\frac{P_k(F)}{P_k(F')} \right]^{n_k}.$$

Without loss of generality assume $N \geq N'$. If $P_0(F) < P_0(F')$, then let us consider the data corresponding to no capture, i.e. $n = n_1 = ... = n_k = ... = n_S = 0$. We have

$$\frac{L(N, F)}{L(N', F')} = \frac{P_0(F)^N}{P_0(F')^{N'}} < 1.$$

If $P_0(F) = P_0(F')$ and $N > N'$, the same inequality holds taking $n = 1$.

If $P_0(F) = P_0(F')$ and $N = N'$, then there must be at least one $k > 0$ such that $P_k(F) \neq P_k(F')$. Let $n_k = N = N'$, we have

$$\frac{L(N, F)}{L(N', F')} = \frac{P_k(F)^N}{P_k(F')^N} \neq 1.$$

If $P_0(F) > P_0(F')$ and $N = N'$ we have as above that, letting $n_0 = N = N'$

$$\frac{L(N, F)}{L(N', F')} = \frac{P_0(F)^N}{P_0(F')^N} \neq 1.$$

It remains for us to show what happens when $P_0(F) > P_0(F')$ and $N > N'$. Since the cell probabilities sum to 1, there must also be at least one $k > 0$ such that $P_k(F) < P_k(F')$. Let $f(n)$ be the likelihood ratio obtained when $n_k = n$. We have that when there is no capture $n = n_k = 0$

$$f(0) = \frac{P_0(F)^N}{P_0(F')^{N'}} = P_0(F)^{N-N'} \left(\frac{P_0(F)}{P_0(F')}\right)^{N'}$$

while, if we observe just one unit at the k-th occasion, i.e. $n = n_k = 1$ then

$$f(1) = \frac{N}{N'} \frac{P_0(F)^{N-1}}{P_0(F')^{N'-1}} \frac{P_k(F)}{P_k(F')} = \frac{N}{N'} P_0(F)^{N-N'} \left(\frac{P_0(F)}{P_0(F')}\right)^{N'-1} \frac{P_k(F)}{P_k(F')}.$$

If either $f(0) \neq 1$ or $f(1) \neq 1$ we have the thesis. Assume $f(0) = f(1) = 1$.

We have that when $n = n_k = 2$ (two animals captured, both twice)

$$f(2) = \frac{N(N-1)}{N'(N'-1)} P_0(F)^{N-N'} \left(\frac{P_0(F)}{P_0(F')}\right)^{N'-2} (P_k(F)/P_k(F'))^2$$

By using the equation $f(0) = 1$ the above expressions can be conveniently simplified. Since $f(0) = 1$, we have $P_0(F)^{N-N'} = \left(\frac{P_0(F')}{P_0(F)}\right)^{N'}$, so that

$$\begin{aligned} f(1) &= \frac{N}{N'} \left(\frac{P_0(F')}{P_0(F)}\right) \frac{P_k(F)}{P_k(F')} \\ f(2) &= \frac{N(N-1)}{N'(N'-1)} \left(\frac{P_0(F')}{P_0(F)} \frac{P_k(F)}{P_k(F')}\right)^2. \end{aligned}$$

Now, since $f(1) = 1$ and $(N/N')^2 \neq (N(N-1))/(N'(N'-1))$ we have that $f(2) \neq 1$, and hence the thesis.

4 Concluding remarks

In this note we have just investigated from a theoretical perspective the problem of the identifiability of the parameters involved in model $\mathcal{M}_h$ when no restrictive assumption is made on the distribution of the heterogeneous probabilities. We have shown how model $\mathcal{M}_h$ can be safely dealt with through the complete likelihood, thus overcoming the nonidentifiability issues raised by Link (2003). In fact, in a very recent note Mao (2008) highlighted that somehow identifiability can also be bypassed through the conditional likelihood itself. Indeed he shows that by solving some extremal problem one can derive a useful lower bound estimate of $\hat{P}_0(F)$ and hence $\hat{N}$ within the class of distributions sharing those moments which match with the estimated conditional probabilities $\hat{P}_{(c),k}(F)$ $(1 \leq k \leq S)$.

We are currently investigating in more depth the relationship between the advocated solution of maximizing the complete likelihood and the approach of Mao (2008). Notice that in both cases we end up with a consistent estimate of the true N_0 and a comparison in terms of efficiency would also be very interesting. Another competitor to be compared with is the Bayesian solution already available from the reference Bayesian approach in Tardella (2002) which has already proved itself competitive in terms of relative mean square error when compared to other classical, parametric and nonparametric approaches.

References

Basu, S. & Ebrahimi, N. (2001). Bayesian capture-recapture methods for error detection and estimation of population size: heterogeneity and dependence. *Biometrika* **88**, 269–279.

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

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

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

Huggins, R. (2001). A note on the difficulties associated with the analysis of capture-recapture experiments with heterogeneous capture probabilities. *Statistics & Probability Letters* **54**, 147–152.

Huggins, R. M. & Yip, P. S. F. (2001). A note on nonparametric inference for capture-recapture experiments with heterogeneous capture probabilities. *Statistica Sinica* **11**, 843–853.

Hwang, W.-H. & Huggins, R. (2005). An examination of the effect of heterogeneity on the estimation of population size using capture-recapture data. *Biometrika* **92**, 229–233.

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

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

Mao, C. X. (2008). On the nonidentifiability of population sizes. *Biometrics* **64**, 977–979.

Norris, J. L. I. & Pollock, K. H. (1996). Nonparametric MLE under two closed capture-recapture models with heterogeneity. *Biometrics* **52**, 639–649.

Rao, C. R. (1965). *Linear Statistical Inference*. Wiley, New York.

Rolph, J. E. (1968). Bayes estimation of mixing distributions. *The Annals of Mathematical Statistics* **39**, 1289–1302.

Sanathanan, L. (1972). Estimating the size of a multinomial population. *The Annals of Mathematical Statistics* **43**, 142–152.

Tardella, L. (2002). A new Bayesian method for nonparametric capture-recapture models in presence of heterogeneity. *Biometrika* **89**, 807–817.

Yoshida, O., Leite, J. G. & Bolfarine, H. (1999). Stochastic monotonicity properties of Bayes estimation of the population size for capture-recapture data. *Statistics and Probability Letters* **42**, 257–266.