BIOMETRICS 47, 725-732
June 1991

Some Practical Aspects of a Conditional Likelihood Approach to Capture Experiments

R. M. Huggins

Department of Statistics, La Trobe University, Bundoora 3083, Australia

SUMMARY

The use of conditional likelihood methods in the analysis of capture data allows the modeling of capture probabilities in terms of observable characteristics of the captured individuals and the trapping occasions. The resulting models may then be used to estimate the size of the population. Here the use of conditional likelihood procedures to construct models for capture probabilities is discussed and illustrated by an example.

1. Introduction

The recent work of Huggins (1989) introduces a procedure for estimating the size of a closed population when the capture probabilities are heterogeneous by modeling the capture probabilities in terms of observable covariates such as age, sex, weight, trapping history, rainfall, etc. The modeling is done by constructing a likelihood conditional on the captured individuals to estimate the associated parameters. The use of observable covariates in capture-recapture experiments has been previously discussed in Pollock, Hines, and Nichols (1984). They examined inferences based on the full likelihood, which necessitated the construction of categories of individuals according to the values of the covariates, using the midpoint of the category as the covariate, and estimating the number of individuals in each category. The construction of the categories is necessary to overcome the problem that the covariates for uncaptured individuals are not known. The determination of these categories can be difficult if several continuous covariates are involved and the determination of category boundaries may not be clear. The procedure initially presented in Huggins (1989) avoids these problems by making inferences conditional on the captured individuals so that characteristics of uncaptured individuals are not required.

The basic theory of this approach was presented in Huggins (1989) and here this modeling is discussed in more detail than in Huggins (1989) with particular attention being paid to the more practical problem of determining an appropriate model.

To illustrate these results a data set collected by V. Reid and distributed with the CAPTURE program of Otis et al. (1978) is considered. The data in a form suitable for this purpose are reproduced in Appendix 1. In a previous discussion of this data set, Otis et al. (1978, p. 32), using their techniques, determined that the most suitable model of those available to them was model M_b, which allows the capture probabilities to vary in response to prior capture. Accompanying this data set are several covariates on the captured individuals—sex, age, and weight. The models considered here are described in Section 2 and they are fitted to the data in Section 3. In Section 4 a test is given to determine whether all the variability in the capture probabilities is explained by the model and this test is applied to the data sets. Finally, in Section 5 the estimation of the population size is examined.

Key words: Capture experiment; Population size estimation; Variable capture probability.

2. The Conditional Likelihood and the Linear Logistic Model

Let p_{ij} denote the probability that individual i is captured on occasion j, where $i = 1, \ldots, N$ are the individuals in the population and $j = 1, \ldots, t$ are the capture occasions. Here N denotes the population size.

Briefly, the methods of Huggins (1989), to which the reader is referred for details, construct a likelihood conditional on the captured individuals that may be written in terms of

$$\gamma_{ij} = \frac{p_{ij}}{1 - (1 - z_{ij})\Pi_{l=j}^{t}(1 - p_{il}^{*})},$$

where p_{ij}^{*} is p_{ij} evaluated when $z_{ij} = 0$, with z_{ij} being the indicator of the past capture of individual i, i.e.,

$$z_{ij} = \begin{cases} 1, & \text{if individual } i \text{ has been captured before } j, \\ 0, & \text{elsewhere.} \end{cases}$$

Then γ_{ij} is the probability individual i is captured on occasion j given its past capture history and given it is captured at least once in the course of the experiment.

Now let $x_{ij} = 1$ if individual i is captured on occasion j and $x_{ij} = 0$ otherwise, and label the captured individuals $1, \ldots, n$ with the uncaptured individuals being labelled $n + 1, \ldots, N$. The conditional likelihood is then proportional to

$$L = \prod_{i=1}^{n} \prod_{j=1}^{t} \gamma_{ij}^{x_{ij}}(1 - \gamma_{ij})^{(1-x_{ij})}, \tag{2.1}$$

which depends only on the captured individuals. An equivalent form of this conditional likelihood is given in Appendix 2.

Various models for the capture probabilities p_{ij}, such as the linear logistic models discussed below, may be used and associated parameters can be estimated from the conditional likelihood.

Here it is supposed that $\ln[p_{ij}/(1 - p_{ij})]$ is a linear function of covariates corresponding both to characteristics of the individuals and to differences in the environment on the different trapping occasions. This linear logistic model has been previously used in Pollock et al. (1984) and has been extensively used in the statistical literature; see Cox (1970). Asymptotic properties of the corresponding maximum conditional likelihood estimators of the parameters in linear logistic models are derived in Huggins (1989); in particular, the estimators are (unconditionally) asymptotically normal and their variances can be recovered from the matrix of second derivatives of the log-likelihood as usual. From a computational point of view it is usually most convenient to maximize the log-likelihood and recover the variances from the resulting Hessian matrix. Due to the nonstandard form of the conditional probabilities γ_{ij} appearing in the conditional likelihood (2.1) it is necessary to use a general-purpose optimisation routine to find the estimates rather than a standard logistic regression package.

The following models for the data of Appendix 1 are considered with the closest model of Otis et al. (1978) given in parentheses where appropriate. In some cases these models are a simple reparameterisation of those of Otis et al. whilst others are of greater complexity.

Model 0. (M_0)

$$\ln\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_0, \quad j = 1, \ldots, t, \quad i = 1, \ldots, N.$$

This model assumes all the individuals have the same catchability that does not vary with time, the individual's prior capture history, or any covariates. The model is a reparameterisation of model M_0 of Otis et al. (1978).

Model 1. (M_b)

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_b z_{ij}, \quad j = 1, \ldots, t, \quad i = 1, \ldots, N.$$

This parameterisation allows the capture probabilities to vary only according to an individual's capture history and is a reparameterisation of model M_b of Otis et al. (1978).

Model 2. (M_t)

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_j, \quad j = 1, \ldots, t, \quad i = 1, \ldots, N,$$

where $\beta_t = 0$. This model allows the capture probabilities to vary only by time.

Model 3. (M_{tb})

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_j + \beta_b z_{ij}, \quad j = 1, \ldots, t, \quad i = 1, \ldots, N,$$

where again $\beta_t = 0$ and β_b is the effect of previous capture. This model allows the capture probabilities to vary according to time and past behaviour.

Model 4. (special case of M_h)

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_{\text{sex}} + \beta_{\text{age}} + \beta_{\text{wt}} \times \text{weight}(i), \quad j = 1, \ldots, t, \quad i = 1, \ldots, N,$$

where β_{sex} is the effect of sex, β_{age} the effect of age, β_{wt} the effect of one unit of weight, and weight(i) the weight of individual i. This model accounts for heterogeneity resulting only from differences in sex, age, and weight. There is no equivalent model of Otis et al. (1978) although this model is a restricted version of their model M_h.

Model 5. (special case of M_{bh})

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_{\text{sex}} + \beta_{\text{age}} + \beta_{\text{wt}} \times \text{weight}(i) + \beta_b z_{ij},$$

$$j = 1, \ldots, t, \quad i = 1, \ldots, N.$$

Here the capture probabilities depend on the covariates of Model 4 and the individual's past trapping history. There is no equivalent model of Otis et al. (1978) although this model is a restricted version of their model M_{bh}.

Model 6. (special case of M_{tbh})

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_{\text{sex}} + \beta_{\text{age}} + \beta_{\text{wt}} \times \text{weight}(i) + \beta_j + \beta_b z_{ij},$$

$$j = 1, \ldots, t, \quad i = 1, \ldots, N.$$

This is Model 5 with probabilities allowed to vary over time. Again this is a special case of the model M_{tbh} of Otis et al. (1978) but is not equivalent to their model.

Model 7. (variant of M_{bh})

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_{\text{sex}} + \beta_{\text{age}} + \beta_{\text{wt}} \times \text{weight}(i) + \theta_i, \quad j = 1, \ldots, t, \quad i = 1, \ldots, N.$$

This model allows some of the variability between individuals to be explained by the observed covariates but still allows variation due to unobserved characteristics.

3. Model Testing

The first six models above are computationally straightforward and one may construct a conditional likelihood in the manner of Huggins (1989) to estimate the model parameters. Here we are interested in the value of the logarithm of the conditional likelihood upon which we base our model selection techniques. The conditional likelihoods were computed on an IBM-compatible personal computer using the general purpose statistical computing package Gauss. As noted above, standard logistic regression packages are not suited to this analysis. For the analysis the three semi-adult individuals were recoded as adult.

For the Models 1–6 tests may be based on the likelihood ratio test or on Akaike's information criterion (AIC), given by

$$\mathrm{AIC} = -2\ln L(\hat{\beta}) + 2s,$$

where $\ln L(\hat{\beta})$ is the log-likelihood for the model evaluated at the maximum conditional likelihood estimator $\hat{\beta}$ and s is the number of parameters estimated by $\hat{\beta}$; see Read and Cressie (1988, §8.3).

For the above models the corresponding maximised conditional log-likelihoods and AIC for the data set of Appendix 1 are given in Table 1. The use of the AIC would result in the choice of Model 5, a special case of Model M_{bh} of Otis et al. (1978).

A more traditional method of comparing two nested models is to compute a likelihood ratio statistic that is twice the difference in the log likelihoods of the two models. This statistic then has asymptotically a chi-square distribution whose degrees of freedom is the difference in the number of parameters between the two models. One can then conduct a sequence of tests to determine the "best" model. The main problem with this approach is that the order in which the tests are performed could lead to different models being selected as "best." The choice of procedure for selecting an appropriate model is dependent on one's personal preference and the desired degree of control over the modeling process.

The approach taken here, summarised in Table 2, is to initially test Model 0 separately against the model involving only the past capture history, Model 1, the model involving only a time effect, Model 2, and a model involving only the covariates, Model 4. The results of these tests are given in the first columns of Table 2 where it is seen that Model 0 is rejected in favour of Model 1 or Model 4. The next tests involve comparing Model 1 with Models 3 and 5 and Model 4 with Model 5, and it is seen from Table 2 that there is no reason to reject Model 1 in favour of Model 3 but both of Models 1 and 4 are rejected in favour of Model 5, which involves the covariates and the previous capture history. Model 5 is then compared with Model 6 in the final columns of Table 2 and there is no reason to reject Model 5. Thus Model 5 seems most suitable.

Other tests not reported here confirmed that all the covariates should be in the model. To complete the modeling process we need to determine whether our model suitably explains the variation in the capture probabilities or if Model 7 is more appropriate. In Section 4 below we

Table 1
Log-likelihoods and AIC for the Models 1–6 *of Section* 2

Model	0	1	2	3	4	5	6
Number of parameters	1	2	6	7	4	5	10
Log-likelihood	157.3	150.4	152.4	148.2	144.9	139.5	137.3
AIC	316.6	304.8	316.8	310.4	297.8	289	294.6

Table 2
Major testing sequence for fitting the models of Section 2

Test	χ^2	d.f.	P	Test	χ^2	d.f.	P	Test	χ^2	d.f.	P
0 vs 1	13.68	1	.0002	1 vs 5	21.8	3	.0001	5 vs 6	4.45	5	.486
0 vs 2	9.8	5	.081	1 vs 3	4.5	5	.479				
0 vs 4	24.8	3	.0001	4 vs 5	10.8	1	.001				

The tests in the first column were conducted first, then those in the second column, and finally that in the third. In each case the null model is stated first, then the alternative (e.g., 0 vs 1 is a test of the null hypothesis that Model 0 is the correct model versus the alternative that Model 1 is the correct model. The likelihood ratio statistic (χ^2), its degrees of freedom, and associated P-value are given for each test.

Table 3
Summary data and capture probabilities for the data of Appendix 1. The stated capture probabilities are computed at the average weights for each age and sex.

	Young	Adult	Young	Adult
	Number caught		Average weights	
Females	9	8	9.89	20.5
Males	15	6	12.53	18.5
	First capture probability		Recapture probability	
Females	.21	.18	.46	.41
Males	.50	.28	.76	.56

propose a procedure that does not involve computing the estimates of the θ_i of Model 7 to test if a model explains all the variations in the capture probabilities; based on the results of that test, Model 5 is quite reasonable. The parameter estimates for this model, with estimated standard errors in parentheses, are $\beta_0 = -2.907$ (.873), $\beta_{\text{sex}} = .916$ (.354), $\beta_{\text{age}} = -1.884$ (.634), $\beta_{\text{wt}} = .159$ (.0628), $\beta_b = 1.176$ (.405).

Some summary data for the individuals in the data set of Appendix 1 are given in Table 3.

4. Testing for Individual Variation

In this section we develop some theory that allows us to determine whether a proposed model suitably explains the variation in the capture probabilities. The major problems are the large number of parameters to be estimated and the small number of observations on each individual, which does not allow the use of the usual asymptotic arguments.

We are concerned first with testing H_0: $\theta_1 = \cdots = \theta_n = 0$ without having to estimate the θ and an obvious candidate for such a test is the score test of Rao (1973, §6e.3), which one could hope to have a chi-square distribution with n degrees of freedom. Unfortunately, standard asymptotic arguments are not available to prove this result and the score test is ad hoc at best. Note that the number of parameters increases as the population size increases and one has only a finite number of observations on each individual, which clearly precludes asymptotic arguments unless one allows the number of trapping occasions to increase to infinity. Here it is assumed that the covariates associated with individual i are independently and identically distributed so that the asymptotic arguments of Huggins (1989) may be used.

Denote by $\boldsymbol{\beta}$ the vector of parameters, apart from the θ_i, associated with the model, and define, where $\boldsymbol{\theta}$ is the vector of the θ_i,

$$v_i(\boldsymbol{\beta}, \theta_i) = \frac{d \ln L(\boldsymbol{\beta}, \boldsymbol{\theta})}{d\theta_i} I(C_i),$$

where $I(C_i) = 1$ if individual i is captured in the course of the experiment and $I(C_i) = 0$ otherwise, and $L(\boldsymbol{\beta}, \boldsymbol{\theta})$ is the conditional likelihood.

Further let

$$a_i(\boldsymbol{\beta}, \theta_i) = \frac{-dv_i(\boldsymbol{\beta}, \theta_i)}{d\theta_i}.$$

Then $\mathrm{E}[v_i(\boldsymbol{\beta}, \theta_i)] = 0$ and $\mathrm{E}[v_i(\boldsymbol{\beta}, \theta_i)^2] = \mathrm{E}[a_i(\boldsymbol{\beta}, \theta_i)]$.

We base our test on the statistic

$$\frac{\sum_{i=1}^{n}\left[v_i(\hat{\boldsymbol{\beta}}, 0)^2 - a_i(\hat{\boldsymbol{\beta}}, 0)\right]}{\left(\sum_{i=1}^{n}\left[v_i(\hat{\boldsymbol{\beta}}, 0)^2 - a_i(\hat{\boldsymbol{\beta}}, 0)\right]^2\right)^{1/2}},$$

which, under our assumptions that the individual covariates z_i are independently and identically distributed and the θ_i are identically zero, will be asymptotically normally distributed. To see this note that the $v_i(\boldsymbol{\beta}, \theta_i)^2 - a_i(\boldsymbol{\beta}, \theta_i)$ are independently and identically distributed when θ_i is at its true value. Hence $N^{-1/2} \sum_1^N [v_i(\boldsymbol{\beta}, \theta_i)^2 - a_i(\boldsymbol{\beta}, \theta_i)]$ will be asymptotically standard normal. Further, under H_0, $\hat{\boldsymbol{\beta}}$ has been shown in Huggins (1989) to be asymptotically normal so that the usual Taylor series arguments show that $N^{-1/2} \sum_1^N [v_i(\hat{\boldsymbol{\beta}}, 0)^2 - a_i(\hat{\boldsymbol{\beta}}, 0)]$ is asymptotically normal as well. Under H_0 a consistent estimate of the variance of a single term is $N^{-1} \sum_1^N [v_i(\hat{\boldsymbol{\beta}}, 0)^2 - a_i(\hat{\boldsymbol{\beta}}, 0)]^2$, which gives the above test.

When applied to the data of Appendix 1 using Model 5, the value of the test statistic was 1.44, which gives a P-value of .1505, as our test is two-sided. By comparison, when the test is applied to the same data using Model 0 the resulting statistic is 3.12, a P-value of .0017; and when applied using Model 1, that selected by Otis et al. (1978), the resulting statistic is 2.434, a P-value of .014. This is a result quite different from that of Otis et al. (1978). However, in our terms their test is testing Model 1 against Model 3 and our P-value of .479 is in agreement with the value of .43 obtained using their test. The increased power of the test presented here is due to the more general alternative.

5. Estimating the Population Size

The population size may be estimated by a direct application of the methods of Huggins (1989). Having used a conditional likelihood to select a model and estimate the associated parameters, if this model does not include separate parameters for each individual, the methods of Huggins (1989) are directly applicable to estimate the population size and obtain confidence intervals.

Let

$$p_i(\boldsymbol{\beta}) = 1 - \prod_{j=1}^{t}\left(1 - p_{ij}^*\right),$$

where $\boldsymbol{\beta}$ is the vector of parameters associated with the model. Then $p_i(\boldsymbol{\beta})$ is the probability that individual i is captured at least once in the course of the trapping experiment. An unbiased estimate of the population size n is

$$\hat{N}(\boldsymbol{\beta}) = \sum_{i=1}^{n} p_i(\boldsymbol{\beta})^{-1},$$

and an estimate of the variance of $\hat{N}(\boldsymbol{\beta})$ is

$$s^2(\boldsymbol{\beta}) = \sum_{i=1}^{n} p_i(\boldsymbol{\beta})^{-2}\left[1 - p_i(\boldsymbol{\beta})\right].$$

When $\boldsymbol{\beta}$ is estimated from the data by $\hat{\boldsymbol{\beta}}$ we use $\hat{N}(\hat{\boldsymbol{\beta}})$ to estimate N and if the model does not include the θ_i of Section 4 we may take $\hat{N}(\hat{\boldsymbol{\beta}})$ to be asymptotically normally distributed with mean N and variance $s^2(\hat{\boldsymbol{\beta}}) + \hat{\mathbf{D}}\hat{\mathbf{I}}^{-1}\hat{\mathbf{D}}$, where $\hat{\mathbf{I}}$ is the matrix of second derivatives (i.e., the Hessian) of the conditional log-likelihood evaluated at $\hat{\boldsymbol{\beta}}$, and $\mathbf{D}$ is the vector

$$\left.\frac{dN(\boldsymbol{\beta})}{d\boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} = \sum_{i=1}^{n} p_i(\hat{\boldsymbol{\beta}})^{-2} \left.\frac{dp_i(\boldsymbol{\beta})}{d\boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}}$$

$$= \sum_{i=1}^{n} p_i(\hat{\boldsymbol{\beta}})^{-2} \sum_{j=1}^{t} \left(1 - p_{ij}^*\right)^2 \prod_{k=1,\, k\neq j}^{t} \left(1 - p_{ij}^*\right).$$

Note that both $\mathbf{I}$ and $\mathbf{D}$ can be obtained numerically on many computer packages. For the data set of Appendix 1 the population size was estimated according to Model 5 to be 47.144 with an estimated standard error of 7.18. Most of the variance in this estimate was due to the variation in estimating the parameters associated with the model. The corresponding estimate of Otis et al. (1978) for this data set was 41 with an estimated standard error of 3.0518. Our Model 1 is essentially equivalent to that fitted by Otis et al. and for this model our methods gave an estimate of the population size as 42.26 with a standard error of 3.75.

Acknowledgements

This work was motivated by several fruitful conversations with David Anderson, Ken Burnham, and Ken Wilson of the Colorado Cooperative Fish and Wildlife Research Unit. The use of Akaike's information criterion was suggested independently by Ken Burnham and an anonymous referee.

Résumé

L'emploi des méthodes de vraisemblance conditionnelle dans l'analyse des données de capture permet la modélisation des probabilités de capture en termes de caractéristiques observables des individus capturés et des occasions de piégeage. Les modèles résultants peuvent alors être utilisés pour estimer l'effectif de la population. Ici, l'emploi des procédures de vraisemblance conditionnelle pour construire des modèles de probabilité de capture est discuté et illustré par un exemple.

References

Cox, D. R. (1970). *Analysis of Binary Data.* London: Methuen.

Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* **76**, 133–140.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). *Statistical Inference from Capture Data on Closed Animal Populations.* Wildlife Monographs, No. 62.

Pollock, K. H., Hines, J. E., and Nichols, J. D. (1984). The use of auxiliary variables in capture–recapture and removal experiments. *Biometrics* **40**, 329–340.

Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data.* New York: Springer-Verlag.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition. New York: Wiley.

Received May 1989; *revised October* 1989; *accepted March* 1990.

Appendix 1

Captures of peromyscus maniculatus *collected by V. Reid at East Stuart Gulch, Colorado. The columns represent the sex* (m *or* f), *the ages* (y: *young*, sa: *semi-adult*, a: *adult*), *the weights in grams, and the capture histories of* 38 *individuals over* 6 *trapping occasions* (1: *captured*, 0: *not captured*).

m	y	12	1	1	1	1	1	1	m	y	13	0	1	1	0	1	0
f	y	15	1	0	0	1	1	1	f	y	5	0	1	0	1	0	1
m	y	15	1	1	0	0	1	1	f	a	20	0	1	0	0	0	1
m	y	15	1	1	0	1	1	1	m	y	12	0	1	0	0	1	1
m	y	13	1	1	1	1	1	1	f	y	6	0	0	1	0	0	0

APPENDIX 1—*Continued*

m	a	21	1	1	0	1	1	1	f	a	22	0	0	1	1	1	1
m	y	11	1	1	1	1	1	0	f	y	10	0	0	1	0	1	1
m	sa	15	1	1	1	0	0	1	f	y	14	0	0	1	1	1	1
m	y	14	1	1	1	1	1	1	f	a	19	0	0	1	0	0	0
m	y	13	1	1	0	1	1	1	f	a	20	0	0	0	1	0	0
f	a	22	1	1	1	0	1	1	m	sa	16	0	0	0	1	1	1
m	y	14	1	1	1	1	1	1	f	y	11	0	0	0	1	1	0
m	y	11	1	0	1	1	1	0	m	y	14	0	0	0	0	1	0
f	y	10	1	0	0	1	0	0	f	y	11	0	0	0	0	1	0
f	a	23	0	1	0	0	1	0	m	a	24	0	0	0	0	1	0
f	y	7	0	1	1	0	0	1	m	y	9	0	0	0	0	0	1
m	y	8	0	1	0	0	0	1	m	sa	16	0	0	0	0	0	1
m	a	19	0	1	0	1	0	1	f	a	19	0	0	0	0	0	1

APPENDIX 2

The conditional likelihood given in Section 2 is the formulation of Huggins (1989), which may be expressed in an equivalent form as follows. If individual i has not been captured before occasion j then it is easily shown that

$$(1-\gamma_{ij}) = \frac{(1-p_{ij})\left[1-\Pi_{l=j+1}^{t}(1-p_{il}^{*})\right]}{1-\Pi_{l=j}^{t}(1-p_{il}^{*})},$$

so that for an individual captured for the first time on occasion k,

$$\prod_{j=1}^{t}\gamma^{x_{ij}}(1-\gamma_{ij})^{(1-x_{ij})} = \prod_{j=1}^{k-1}\frac{(1-p_{ij})\left[1-\Pi_{l=j+1}^{t}(1-p_{il}^{*})\right]p_{ik}}{\left[1-\Pi_{l=j}^{t}(1-p_{il}^{*})\right]\left[1-\Pi_{l=k}^{t}(1-p_{il}^{*})\right]}$$

$$\times \prod_{j=k+1}^{t} p_{ij}^{x_{ij}}(1-p_{ij})^{(1-x_{ij})}$$

$$= \prod_{j=1}^{t}\frac{p_{ij}^{x_{ij}}(1-p_{ij})^{(1-x_{ij})}}{1-\Pi_{l=1}^{t}(1-p_{il}^{*})}.$$

All individuals appearing in the conditional likelihood of Section 2 must have been captured at such a k and thus the conditional likelihood of Section 2 is equivalent to

$$\prod_{i=1}^{n}\prod_{j=1}^{t}\frac{p_{ij}^{x_{ij}}(1-p_{ij})^{(1-x_{ij})}}{1-\Pi_{l=1}^{t}(1-p_{il}^{*})}.$$