Two-stage approaches to the analysis of occupancy data I: The homogeneous case

(Analysis of occupancy data)

N. Karavarsamis* and R. M. Huggins
School of Mathematics and Statistics
The University of Melbourne, Victoria 3010, Australia.

Abstract: Occupancy models are used in statistical ecology to estimate species dispersion. The two components of an occupancy model are the detection and occupancy probabilities, with the main interest being in the occupancy probabilities. We show that for the homogeneous occupancy model there is an orthogonal transformation of the parameters that gives a natural two-stage inference procedure based on a conditional likelihood. We then extend this to a partial likelihood that gives explicit estimators of the model parameters. By allowing the separate modelling of the detection and occupancy probabilities, the extension of the two-stage approach to more general models has the potential to simplify the computational routines used there.

Key Words: Imperfect detection; Occupancy models; Orthogonal parameterisation; Partial likelihood.

*Corresponding author: N. Karavarsamis, email: nkarav@unimelb.edu.au
1 Introduction

The importance of occupancy studies in conservation planning, biodiversity monitoring, and invasive species biology, is well known (Hoffman et al., 2010) and a complete description of hierarchical occupancy models is contained in Royle and Dorazio (2008). These data consist of observations on whether a site is occupied and are collected over repeated visits to a number of sites. They require less survey effort and are cheaper than studies that estimate abundance (Pollock, 2002; MacKenzie et al., 2002; Wintle et al., 2005). The full likelihood (MacKenzie et al., 2002) involves both the probability a site is occupied and the probability of detection of a species at a site if it is occupied.

As the full likelihood involves both the detection and occupancy probabilities, in general inference requires the joint modelling of these quantities and can quickly become quite complicated. For example if there are 10 covariates then there are \(2^{20} = 1048576\) possible joint models for detection and occupancy. However, if the modelling can be done separately for each of detection and occupancy there are \(2^{10} + 2^{10} = 2048\) models. Even with a smaller number of covariates, if one fits non-parametric functions of the covariates rather than simple linear models the number of parameters to be estimated can become large. Moreover, computer algorithms to fit models with large numbers of parameters can be unstable and finding global rather than local maxima can be difficult. As a first step towards simplifying the likelihood with the aim of eventually alleviating these computational difficulties we consider the simple homogeneous model for both occupancy and detection probabilities.

In statistics there are several ways of simplifying complex likelihoods. One is through parameter orthogonalization (e.g. Cox and Reid, 1987). We consider a simple transformation of the parameters that yields two orthogonal parameters. We show that the resulting estimate of the detection probability arises from a conditional likelihood and that of the constructed parameter arises from a simple binomial likelihood, giving a natural two stage procedure
to compute the maximum likelihood estimates. Partial likelihood is also used to simplify complex likelihoods and its theory is well developed (e.g. Cox, 1975; Wong, 1986; Gill, 1992). We consider a partial likelihood approach that yields the equivalent of two binomial likelihoods and simple analytic estimates of both parameters.

In Section 2 we give our notation and the likelihood of MacKenzie et al. (2002). In Section 3 we give the orthogonal transformation and show this yields a conditional likelihood to estimate the detection probabilities. In Section 4 we give a partial likelihood approach that further simplifies estimation. In Section 5 we conduct some simulations to examine the efficiency of the resulting occupancy estimator. In Section 6 we apply the estimators to several data sets. Some technical results are given in the Appendices.

2 Notation and Likelihood

Consider $S$ sites labelled $s = 1, \ldots, S$ where each site is visited on $\tau$ occasions. We suppose the occupancy status of each site is constant over the visits. Let $\psi$ be the probability a site is occupied and $p$ be the probability the species is observed at a site on a given occasion given it is present. Then $\theta = 1 - (1 - p)^\tau$ is the probability of at least one detection at a site given the site is occupied. Let $y_s$ denote the number of occasions upon which the species was detected at site $s$ and let $y = \sum_{s=1}^{S} y_s$ be the total number of detections, let $f_0$ be the number of sites where none of the species was detected and let $O = S - f_0$ be the number of sites where they were. Reorder the $S$ sites $s = 1, \ldots, O, O + 1, \ldots S$, where $1, \ldots, O$ denote the sites at which at least one detection occurred and $O + 1, \ldots S$ the remaining sites at
which no sightings occurred. The full likelihood (MacKenzie et al., 2006) is then

\[
L(\psi, p) \propto (1 - \psi \theta)^{f_0} \psi S - f_0 \prod_{s=1}^{O} \left( \frac{\tau}{y_s} \right) p^{y_s} (1 - p)^{\tau - y_s}
\]

\[
\propto (1 - \psi \theta)^{f_0} \psi S - f_0 p^{y} (1 - p)^{O \tau - y}.
\]

This likelihood may be maximised numerically and the maximum likelihood estimates \( \hat{\psi} \) and \( \hat{p} \) found, for example using the \( \text{R} \) package \texttt{unmarked} (Fiske and Chandler, 2015).

3 Orthogonal Transformation and Conditional Likelihood

For the homogeneous occupancy model there is a natural transformation that yields orthogonal parameters. Let \( \eta = \psi \theta \) be the probability that occupancy is detected at a site and now consider the parameters \( \eta, p \). The full likelihood (1) is then proportional to

\[
L(\eta, p) = (1 - \eta)^{f_0} \eta^{S - f_0} \prod_{s=1}^{O} \frac{p^{y_s} (1 - p)^{\tau - y_s}}{\theta^{S - f_0}}.
\]

\[
= (1 - \eta)^{f_0} \eta^{S - f_0} p^{y} (1 - p)^{O \tau - y} \theta^{O}.
\]

To determine orthogonality of \( \eta \) and \( p \) we need to examine the resulting information matrix (Cox and Reid, 1987). The log-likelihood is

\[
\ell(\eta, p) = f_0 \log(1 - \eta) + (S - f_0) \log(\eta)
\]

\[
+ y \log(p) + (O \tau - y) \log(1 - p) - O \log(\theta).
\]
Hence

\[
\frac{\partial \ell(\eta, p)}{\partial \eta} = - \frac{f_0}{1 - \eta} + \frac{(S - f_0)}{\eta}, \tag{5}
\]

\[
\frac{\partial \ell(\eta, p)}{\partial p} = y - \frac{(O\tau - y)}{1 - p} - \frac{O\tau(1 - \theta)}{(1 - p)\theta}, \tag{6}
\]

so that \(\frac{\partial^2 \ell(\eta, p)}{\partial \eta \partial p} = \frac{\partial^2 \ell(\eta, p)}{\partial p \partial \eta} = 0\), and the transformed parameters \(\eta\) and \(p\) are orthogonal.

From (5) it is seen that the mle of \(\eta\) is \(\hat{\eta} = (S - f_0)/S = O/S\). The maximum likelihood estimator of \(p\) maximises \(\ell\), which is the log-likelihood conditional on at least one detection at each site. It can be computed using the VGAM package in R (Yee 2010). Let \(\hat{p}\) be the resulting estimator of \(p\) and let \(\hat{\theta} = 1 - (1 - \hat{p})\tau\). The invariance property of maximum likelihood estimates yields that the mle of \(\psi\) is \(\hat{\psi} = \hat{\eta}/\hat{\theta}\). As these are the maximum likelihood estimates the usual estimates of their variances may be used. Thus for the homogeneous model maximum likelihood estimation for the occupancy model may be naturally conducted as a two-stage procedure.

**Remark** A direct verification of the equivalence of the maximum likelihood and conditional likelihood estimates is possible. Note that \(\log L(\psi, p) = f_0 \log(1 - \psi\theta) + (S - f_0) \log(\psi) + y \log(p) + (O\tau - y) \log(1 - p)\) and

\[
\frac{\partial L(\psi, p)}{\partial \psi} = \frac{S(1 - \psi\theta) - f_0}{\psi(1 - \psi\theta)}, \tag{7}
\]

\[
\frac{\partial L(\psi, p)}{\partial p} = \frac{y}{p} - \frac{O\tau - y}{1 - p} - \frac{f_0\psi\tau(1 - \theta)}{(1 - p)(1 - \psi\theta)} = \frac{y}{p} - \frac{O\tau - y}{1 - p} - \frac{f_0\psi\theta}{(1 - \psi\theta)O(1 - p)\theta}. \tag{8}
\]

Now, setting (7) equal to 0 and solving for \(\psi\) yields \(\psi = O/S\theta\) and substituting this into \(f_0\psi\theta/\{(1 - \psi\theta)O\}\) yields \(f_0O\theta/\{\theta S(1 - O/S)O\} = f_0/(S - O) = \frac{f_0}{f_0} = 1\). Thus, setting (7)
and (8) to zero are equivalent to setting (5) and (6) to zero.

4 Partial Likelihood

We have seen the conditional likelihood can be fitted in the VGAM package in R. However, this package is not yet as sophisticated as the glm function in R. Hence, to further simplify estimation we exploit that for occupancy models there are repeated observations at each site so that there is more information on the detection probabilities than on the occupancy probabilities. Let $b_s$ be the number of occasions remaining after the first detection at site $s$ and let $b = \sum_{s=1}^{S} b_s$. The number of re-detections at site $s$ is just $y_s - 1$ so that the total number of re-detections is $y - O$. Let $a = O\tau - O - b$ be the total over the sites where occupancy was detected of the number of occasions before the first detection. Then (2) may be written as

$$L(\eta, p) = (1 - \eta)^{f_0} \eta^{S - f_0} 
\times \frac{p^O (1 - p)^a}{\theta^O} 
\times p^{y - O} (1 - p)^{b - (y - O)}.$$  

In this decomposition (9) and (11) are proportional to simple binomial likelihoods and we base inference on these two components. The component (10) is proportional to the product of the probabilities of an individual’s first detection time given they are detected at least once. In our partial likelihood approach we ignore this component. That is, this partial likelihood approach ignores the information up to and including the first occasion at which a sighting was made at each site.

Using the partial likelihood (11) to estimate $p$ yields that the partial likelihood estimator of $p$ is $\hat{p} = (y - O)/b$ and the usual Binomial variance $\text{Var}(\hat{p}) = \hat{p}(1 - \hat{p})/b$. To estimate $\psi$ we
then estimate \( \eta \) from (9) as in Section 3 then back transform to yield \( \tilde{\psi} = \tilde{\eta}/\tilde{\theta} = (S - f_0)/S\tilde{\theta} \), where \( \tilde{\theta} = 1 - (1 - \tilde{p})^\tau \). In Appendix B we show that

\[
\text{Var} \left( \tilde{\psi} \right) \approx \left( \frac{\psi(1 - \psi\theta)}{S\theta} + \psi^2 \right) \frac{\tau^2(1 - p)^2(\tau - 1)}{\theta^2} \frac{p(1 - p)}{b} + \frac{\psi(1 - \psi\theta)}{S\theta}. \tag{12}
\]

| \( S = 1000, \tau = 5 \) | Partial | Full | Partial | Full |
|--------------------------|---------|------|---------|------|
| \( \hat{p} \) | 0.100 | 0.100 | 0.050 | 0.050 |
| \( \hat{\psi} \) | 0.400 | 0.400 | 0.400 | 0.400 |
| Median estimate | 0.100 | 0.100 | 0.049 | 0.049 |
| Median SE | 0.016 | 0.015 | 0.016 | 0.015 |
| MAD | 0.015 | 0.015 | 0.016 | 0.015 |
| Efficiency | 0.915 | 0.925 | 1.021 | 0.988 |
| \( S = 100, \tau = 5 \) | Partial | Full | Partial | Full |
| \( \hat{p} \) | 0.200 | 0.200 | 0.200 | 0.200 |
| \( \hat{\psi} \) | 0.400 | 0.400 | 0.600 | 0.600 |
| Median estimate | 0.199 | 0.197 | 0.198 | 0.197 |
| Median SE | 0.049 | 0.045 | 0.040 | 0.037 |
| MAD | 0.051 | 0.045 | 0.039 | 0.036 |
| Efficiency | 0.801 | 0.843 | 0.909 | 0.909 |

Table 1: Simulation results to compare our partial likelihood and full likelihood approaches. We report the median of the estimated values, the median of the estimated standard errors, the median absolute deviation of the estimates and efficiency computed from the robust measures as ratio of variances.

5 Simulations

As the conditional likelihood approach to estimating the detection probabilities that arose from the orthogonal transformation yields the mle’s there is no need to evaluate it in simulations. To examine the efficiency of the partial likelihood approach, we simulate an experiment and compute the full and partial likelihood estimates. The standard errors of the full maximum likelihood estimates were computed using a numerically computed observed information matrix. We conducted 1000 simulations for each parameter combination and considered \( S = 1000 \) and 100, \( \psi = 0.4 \) and 0.6, and \( p = 0.05, 0.1 \) and 0.2. In Table 1 we
report the medians of the estimates, the associated standard errors and the median absolute deviation (MAD) of the partial and full likelihood estimators for $\tau = 5$. We report the median, rather than the mean, so our results were not unduly influenced by the occasional outlying estimate. However, in line with convention the efficiency was computed using the variances of the estimates. There is no indication of bias in either method, both estimated standard errors appear reliable and the efficiency of our method for $\psi$ is above 90%. Further, simulations to examine the effects of small probabilities, and different numbers of occasions, and the performance of the partial likelihood estimator, in a setting similar to that in our example, are in Appendix A. In the first instance, there was evidence of some bias for small values of $p$ and small $\tau$ and the estimated standard errors appear slightly too large. In the latter setting, the partial likelihood estimator performs well, the estimated standard errors appear quite reasonable and the estimator of $\psi$, is quite efficient.

6 Application

The application consists of detections of the Growling Grass Frog (*Litoria raniformis*) that were collected at $S = 27$ sites with $\tau = 4$ occasions during the 2002–2003 season, as part of a larger study (Heard et al., 2006). For these data, $f_0 = 12$ and $y = 47$. Here the method of Section 4 yielded $\hat{p} = 0.889$ with estimated standard error 0.052 and $\tilde{\psi} = 0.556$ with estimated standard error 0.096. The full likelihood, with these estimates used as starting values, yielded $\hat{p} = 0.780$ with estimated standard error 0.054 and $\hat{\psi} = 0.557$ with estimated standard error 0.096. The method of Section 3 yielded the same values. The difference between the two estimates of the detection probability may be of interest. One interpretation is that there is some difference between the initial detection probability and that on subsequent occasions, as the full likelihood includes information from the first captures whereas

---

1The MAD is corrected by $R$ for asymptotically normal consistency by a factor of 1.4826.
the partial likelihood does not.

7 Discussion

Two stage procedures are not new in statistical inference. For example in other contexts Sanathanan (1972, 1977) developed two-stage procedures that she termed conditional likelihood and Huggins (1989) has employed a two-stage procedure to estimate population size using capture-recapture data where the first stage employed a conditional likelihood to estimate the capture probabilities. We have focused on estimating the occupancy and detection probabilities in the simple homogeneous case. Firstly, after using an orthogonal transformation we have a natural two-stage method to compute the maximum likelihood estimates. This also provides an interesting example where the maximum likelihood estimator of the detection probability \( p \) is also a conditional maximum likelihood estimator. This may also be useful to illustrate parameter orthogonality and conditional and partial likelihood approaches to inference. The orthogonality of \( \eta \) and \( p \) implies that: the mle’s \( \hat{\eta} \) and \( \hat{p} \) are asymptotically independent, the asymptotic standard error of \( \hat{\eta} \) is the same whether \( p \) is known or unknown, there may be simplifications in the numerical derivation of the estimates and the mle \( \hat{\eta}(p) \) of \( \psi \) when \( p \) is given varies slowly as a function of \( p \) (Cox and Reid, 1987). In our case, the last three points are emphasised as \( \hat{\eta} \) does not depend on \( p \). The resulting sensitivity of \( \hat{\psi} \) to changes in \( \hat{p} \) is examined in Appendix C. Both two-stage procedures offer some simplifications. The conditional likelihood approach allows separate estimation of the detection probability and the occupancy probability. However, it requires a relatively non-standard conditional likelihood to estimate the detection probabilities. This can be implemented in the \texttt{R} package \texttt{VGAM}. The two-stage partial likelihood approach further simplifies the derivation of the estimates giving explicit estimators of both the occupancy and detection probabilities. The cost is a small loss of efficiency. The difference between
the partial and conditional likelihood approaches is in how the detection probability \( p \) is estimated. We have restricted ourselves to the homogeneous case, but the extension to more complex models where occupancy and detection probabilities both depend on covariates has potential to reduce the computational burden and increase the computational efficiency of maximum likelihood estimation there. This will be examined elsewhere.

Acknowledgements

We would like to thank Dr Geoffrey Heard (School of Botany, The University of Melbourne) for supplying us with the data for the Growling Grass Frog.

References

Cox, D. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society Series B*, 49:1–39.

Fiske, I. J. and Chandler, R. B. (2015). *Overview of Unmarked: An R Package for the Analysis of Data from Unmarked Animals*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Gill, R. (1992). Marginal partial likelihood. *Scandinavian Journal of Statistics*, 19(2):133–137.

Heard, G., Robertson, P., and Scroggie, M. (2006). Assessing detection probabilities for the endangered growling grass frog in Southern Victoria. *Wildlife Research*, 33:557–564.

Hoffman, J., Aguilar-Amuchastegui, N., and Tyre, A. (2010). Use of simulated data from a
orices-based habitat model to evaluate methods for predicting species occurrence. *Ecography*, 33:656–666.

Huggins, R. M. (1989). On the Statistical Analysis of Capture Experiments. *Biometrika*, 76(1):133–140.

MacKenzie, D. I., Nichols, J., Royle, J., Pollock, K., Bailey, L., and Hines, J. (2006). *Occupancy Estimation and Modeling Inferring Patterns and Dynamics of Species Occurrence*. Elsevier, San Diego, CA.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J., and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8):2248–2255.

Pollock, K. H. (2002). The use of auxiliary variables in capture-recapture modeling: an overview. *Journal of Applied Statistics*, 29:85–102.

Royle, J. A. and Dorazio, R. (2008). *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations, and Communities*. Academic Press, San Diego, CA.

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

Sanathanan, L. (1977). Estimating the Size of a Truncated Sample. *Journal of the American Statistical Association*, 72(359):669–672.

Wintle, B., Kavanagh, R. P., Mccarthy, M. A., and Burgman, M. (2005). Estimating and dealing with detectability in occupancy surveys for forest owls and arboreal marsupials. *The Journal of Wildlife Management*, 69(3):905–917.
Wong, W. (1986). Theory of partial likelihood. *The Annals of Statistics*, 14(1):88–123.

Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, 32:1–34.
A Simulation Results

To examine the effects of small probabilities and different numbers of occasions in our method, we took $S = 100$, $p = 0.05$, $\psi = 0.6$ and considered both $\tau = 5$ and $\tau = 10$. The results are summarized in Table 2 where we only report simulations with $\hat{\psi} < 1$. Next, we took $S = 27$, $\tau = 4$, $\psi = 0.6$ and $p = 0.6$, which is similar to the values in the Growling Grass Frog application. The results are reported in Table 3 along with those for the full likelihood.

|          | $\tau = 5$ | $\tau = 10$ |
|----------|------------|------------|
|          | $p$  | $\psi$  | $p$  | $\psi$  |
| True value | 0.050  | 0.600  | 0.050  | 0.600  |
| Median estimate | 0.067  | 0.475  | 0.052  | 0.587  |
| Median SE | 0.046  | 0.326  | 0.020  | 0.208  |
| MAD | 0.031  | 0.227  | 0.017  | 0.182  |

Table 2: Simulations in the homogeneous case for the two-stage estimator with a moderate number $(S = 100)$ of sites, small $(\tau = 4)$ and large $(\tau = 10)$ numbers of occasions and small values of $p$. We give the true values of the parameters, the median of the estimated values, the median of the estimated standard errors and the median absolute deviation (MAD) of the estimates.

|          | Partial | Full |
|----------|---------|------|
|          | $p$  | $\psi$  | $p$  | $\psi$  |
| True Value | 0.600  | 0.600  | 0.600  | 0.600  |
| Median estimate | 0.600  | 0.604  | 0.600  | 0.604  |
| Median SE | 0.078  | 0.097  | 0.066  | 0.097  |
| MAD | 0.078  | 0.104  | 0.065  | 0.105  |
| Efficiency | 0.709  | 0.991  |

Table 3: Comparisons of the two-stage (partial likelihood) and the full maximum likelihood estimates in the homogeneous case for a small number of sites $(S = 27)$, small number of occasions $(\tau = 4)$ and large values of $p$. We give the true values of the parameters, the median of the estimated values, the median of the estimated standard errors and the median absolute deviation (MAD) of the estimates.
B Variance of $\tilde{\psi}$

Firstly the variance of $\hat{p}$ is relatively straightforward. Given $b$, $y - O \sim \text{Bin}(b, p)$ so that $E(\hat{p}|b) = p$ and $\text{Var}(\hat{p}|b) = p(1 - p)/b$. Hence $\text{Var}(\hat{p}) = p(1 - p)E(1/b)$, which we estimate by $S_p^2 = \hat{p}(1 - \hat{p})/b$. To determine the variance of $\tilde{\psi}$ let $\overline{\psi} = (S - f_0)/(S\theta)$ be the mle for known $p$. Now, $f_0 \sim \text{Bin}(S, 1 - \psi\theta)$ so that $\overline{\psi}$ is unbiased and has variance $V = \psi(1 - \psi\theta)/(S\theta)$. To find the variance of $\tilde{\psi}$, write $\overline{\psi} = \psi\theta/\tilde{\theta}$ and a Taylor expansion yields $\tilde{\theta} \approx \theta + \tau(1 - p)^{\tau - 1}(\tilde{p} - p)$. Then $\tilde{\psi} \approx \overline{\psi}\{1 - \tau(1 - p)^{\tau - 1}(\tilde{p} - p)/\theta\}$. Thus, $E\left(\tilde{\psi}|b, f_0\right) \approx \overline{\psi}$, $E\left(\tilde{\psi}^2\right) = \psi(1 - \psi\theta)/(S\theta) + \psi^2$, so that $\text{Var}\left(\tilde{\psi}|b, f_0\right) \approx \overline{\psi}^2\tau^2(1 - p)^{2(\tau - 1)}\theta^{-2}\text{Var}\left(\hat{p}|b, n_0\right)$. This yields (12).

![Figure 1: Values of $\hat{\psi}_p$ plotted against $p$ from the partial likelihood score equations for an example with $S = 77$ and $\tau = 3$. The vertical dashed line indicates $\hat{p} = 0.54$. The horizontal dashed line shows minimum value for $\hat{\psi}_p$, at 0.415.](image)
C Sensitivity

Small differences in the estimated detection probabilities had little effect on the estimated occupancies. Let $\psi_p$ be the estimated value of $\psi$ for given $p$. Then, $\partial \psi_p / \partial p = (S - f_0)(1 - p)^{(1 - \tau - 1)\tau}/((S(1 - (1 - p)^\tau))$ which is usually small and when multiplied by the difference in the estimated probabilities is even smaller still. To illustrate this suppose $S = 77$ and $\tau = 3$, $f_0 = 45$ and $y = 57$. This gave $\hat{p} = 0.54$. In Figure 1 we give a plot of $\psi_p$ against $p \in (0, 1)$. It is clear that when $p$ is small (i.e. $p \leq 0.15$) values for $\psi_p$ are greater than 1. For values of $p \geq 0.3$ there is little change in $\psi_p$ and for $p \geq 0.5$ practically no change in $\hat{\psi}_p$. Hence, when estimates for $p$, $\hat{p}$, are obtained from our method that differ to estimates given by the full likelihood, there is little to no change in $\psi_p$. 

14