

Heterogeneity for a general class of recapture models based on equality constraints on the conditional capture probabilities

Eterogeneità in cattura-ricattura per una classe generale di modelli basata su vincoli di uguaglianza per le probabilità di cattura condizionate

Alessio Farcomeni

Abstract We define a class of discrete-time capture-recapture models which takes into account time (t), behavioural (b), observed heterogeneity (o) and unobserved heterogeneity (h) effects. The resulting M_{hotb} model is completely general in the tb part. We specify a model in the class through equality constraints on the conditional capture probabilities. A general EM algorithm is described to fit the chosen model.

Abstract Definiamo una classe di modelli per cattura-ricattura in tempo discreto. La classe è in grado di modellare effetti tempo dipendenti, effetti comportamentali, covariate ed eterogeneità non osservata. Il modello risultante è del tutto generale per le prime due parti, e viene specificato attraverso vincoli di uguaglianza per le probabilità di cattura condizionate.

Key words: Capture history, conditional likelihood, equality constraints, heterogeneity, population size.

1 Introduction

Capture-recapture is concerned with estimation of the size of a closed population based on the capture history over repeated occasions. Capture probabilities may depend on four elements: the specific capture occasion (M_t models), the previous capture occasions (M_b models), observed heterogeneity (that is, covariates, called M_o models in Farcomeni and Scacciatelli (2013)) and unobserved heterogeneity (M_h models). These sources of heterogeneity may be combined, to obtain the most general possible model, tagged M_{hotb} in Farcomeni and Scacciatelli (2013). See also Otis et al (1978). Traditional behavioural models let the capture probability depend on the past occasions by updating it after the first capture event. There are generalizations to this approach, see for instance Ramsey and Usner (2003) and

Alessio Farcomeni
Sapienza - University of Rome, e-mail: alessio.farcomeni@uniroma1.it

Yang and Chao (2005). A completely general M_{tb} model is proposed in Farcomeni (2011), where a completely arbitrary dependence among capture occasions is obtained. These do not even need to be ordered along a time horizon. The method is based on equality constraints for the conditional capture probabilities, and surprisingly enough there are closed form expressions for the MLE of nuisance parameters and a simple estimating equation for the MLE of the population size. In the approach of Farcomeni (2011) there is no way of including observed or unobserved heterogeneity, though. We do this in this paper. The inferential approach is not straightforward anymore, but can anyway be based on a general and very efficient algorithm for constrained inference in categorical data analysis. The rest of the paper is as follows: in the next section we revisit Farcomeni (2011) approach. In Section 3 we extend the approach to observed and unobserved heterogeneity. In Section 4 we describe a constrained expectation-maximization (EM) algorithm to obtain the MLE.

2 Set up

Let $y_i = (y_{i1}, \dots, y_{iS})$, $i = 1, \dots, N$, denote the binary capture history for the i -th subject, where we have $S > 1$ capture occasions and an unknown population size N . We observe capture histories of the n subjects for which $\sum_j y_{ij} > 0$. Let $p(y)$ denote the probability of a capture history y . A fully general parameterization is given by the chain rule, so that

$$p(y) = p(Y_{i1} = y_{i1})p(Y_{i2} = y_{i2}|y_{i1}) \cdots p(Y_{iS} = y_{iS}|y_{i,S-1}, \dots, y_{i1}).$$

Call now $p_1 = \Pr(Y_{i1} = 1)$, $p_j(a_{j-1}, \dots, a_1) = p(Y_{ij} = 1|Y_{i,j-1} = a_{j-1}, \dots, Y_{i1} = a_1)$; for $a_j = \{0, 1\}$. These conditional capture probabilities are arranged in lexicographical order in a vector $p = (p_1, p_2(0), p_2(1), p_3(00), \dots, p_S(1, \dots, 1))$. Let further C denote a binary matrix of contrasts, so that equality constraints can be expressed as $C'p = 0$. It is shown in Farcomeni (2011) that all possible M_{tb} models are obtained by varying C , and that for fixed C the MLE can be found with two simple estimating equations. Note that the maximum number of parameters is equal to the number of cells of the contingency table, and therefore the model is always identifiable regardless of C . Further, given usual regularity conditions the MLE is asymptotically consistent and asymptotically normal under the model (Sanathanan (1972)).

3 Observed and unobserved heterogeneity

In this section we obtain a fully general M_{hotb} model. We begin by introducing a mixed-effects logit reparameterization (Coull and Agresti (1999, 2000)), which will be convenient for inclusion of categorical and continuous predictors. It is straight-

forward to check assumptions in the previous section are equivalent to

$$\log \left(\frac{p_j(a_{j-1}, \dots, a_1)}{1 - p_j(a_{j-1}, \dots, a_1)} \right) = \beta_{ja_1, \dots, a_{j-1}}. \quad (1)$$

The logit transformation simply maps the $2^S - 1$ parameters p to the $2^S - 1$ parameters $\beta \in \mathcal{R}^{2^S - 1}$. Equality constraints are imposed similarly, as $C'p = 0$ if and only if $C'\beta = 0$.

Let now X_{ik} denote a subject- and time-specific vector of covariates for the i -th subject. Furthermore, assume that the capture-history probability depends also on a subject-specific parameter θ_i , summarizing subject-specific unobserved heterogeneity. Our final M_{hotb} model is expressed as

$$\log \left(\frac{p_j(a_{j-1}, \dots, a_1, \theta, X)}{1 - p_j(a_{j-1}, \dots, a_1, \theta, X)} \right) = \beta_{ja_1, \dots, a_{j-1}} + \theta + \gamma'X. \quad (2)$$

We make the following assumptions in the model above: (i) *additivity*, that is, that the effects of observed and unobserved heterogeneity is additive on the log-scale (also known as proportionality of the odds) and (ii) *slope homogeneity*, that is, that the effects of the covariates γ do not depend on the specific capture occasion, past capture history, and unobserved heterogeneity. More explicitly, the h , o , and tb effects do not interact. A consequence of slope homogeneity, which can be easily relaxed to some extent if needed, is that all occasion-specific and behavioural effects are captured by the parameter β . The model is completed by equality constraints on the *adjusted* conditional capture probabilities, in the form of an opportune constraint $C'\beta = 0$ and with an assumption for the random effect of the kind $\theta \sim F(\alpha)$, where α is a parameter for the random effects model. The constraint $C'\beta = 0$ guarantees equality of capture-history probabilities for any two subjects with the same covariate configuration X and the same subject-specific parameter θ . By varying the matrix C and the distributional assumption on θ a huge number of previously known and completely new M_{hotb} models can be obtained. See Farcomeni (2011) for some examples when θ and $\gamma'X$ are omitted. There are many parametric choices for the mixing distribution F , including Gaussian, Student's T , univariate symmetric Laplace, logit-Beta, latent class models. Non-parametric assumptions may not be pursued due to identifiability issues given that we work with the conditional likelihood (Link (2003); Farcomeni and Tardella (2012)). It is then recommended that as usual few options are specified and compared through information criteria like Akaike Information Criterion (AIC).

4 Inference through the EM algorithm

The covariate configuration of the unobserved subjects, whose capture history is summarized by $Y_{ij} = 0$ for $j = 1, \dots, S$, is not known. As in Huggins (1989) and

Alho (1990) we condition the likelihood on the event that the subject is captured at least once. The resulting expression does not depend on N or on the unobserved covariates. Additionally, we *complete* the conditional likelihood, by conditioning also on the random effects. The complete conditional log-likelihood is defined when $C'\beta = 0$ and is given by the following expression:

$$\begin{aligned} l_c(\beta, \theta, \gamma, \alpha) = & \sum_{i=1}^n \sum_{j=1}^S Y_{ij} \log(p_j(Y_{i,j-1}, \dots, Y_{i1}, \theta_i, X_i)) + \\ & + (1 - Y_{ij}) \log(1 - p_j(Y_{i,j-1}, \dots, Y_{i1}, \theta_i, X_i)) + \\ & - \sum_{i=1}^n \log(1 - \prod_{j=1}^S (1 - p_j(0, \dots, 0, \theta_i, X_i))) + \sum_{i=1}^n \log(f(\theta_i, \alpha)), \end{aligned} \quad (3)$$

with the conventions that $p_1(0, \dots, 0, \theta_i, X_i) = p_1(\theta_i, X_i)$. For ease of notation we illustrate assuming a continuous distribution is used for θ , where f denotes the density. Latent class models can be accomodated with few adjustments, which are given at the end of the section (see also Coull and Agresti (1999)). It is now a matter of setting up a constrained EM algorithm to obtain the MLE. At the E step we must obtain the expected value of (3) with respect to the current posterior distribution of the random effects

$$f(\theta_i | \alpha, Y, X, \beta, \gamma) \propto f(\theta_i | \alpha) p(Y | X, \theta_i, \beta, \gamma), \quad (4)$$

where (4) is evaluated conditionally on the current parameter estimates. The resulting integral is of the form

$$\begin{aligned} Q(\beta, \gamma | \beta', \gamma', \alpha) = & \sum_{i=1}^n \int \left[\sum_{j=1}^S Y_{ij} (\beta_{jY_{i1}, \dots, Y_{i,j-1}} + \theta_i + \gamma' X_i) + \right. \\ & - \log(1 + \exp(\beta_{jY_{i1}, \dots, Y_{i,j-1}} + \theta_i + \gamma' X_i)) + \\ & - \log(1 - \prod_{j=1}^S (1 + \exp(\beta_{j0, \dots, 0} + \theta_i + \gamma' X_i))^{-1}) \\ & \left. + \log(f(\theta_i | \alpha)) \right] f(\theta_i | \alpha, Y, X, \beta', \gamma') d\theta_i. \end{aligned} \quad (5)$$

Given that θ_i is one-dimensional, an accurate and computationally efficient way of computing (5), and similarly the normalizing constant in (4), is given by use of quadrature. A Gauss-Hermite quadrature would be best for Gaussian random effects, while a Gauss-Laguerre may be better under other assumptions. In all cases, integrals reduce to a sum.

At the M step we maximize (5) under the constraints $C'\beta = 0$. To this end, we implement an instance of the Aitchison and Silvey (1958) (AS) algorithm, see also Bergsma (1997) and Lang (2004). The AS algorithm is an iterative quadratic programming algorithm based on indefinite Lagrange multipliers. Optimization of (5) is equivalent to the system of non-linear equations

$$\begin{cases} s(\eta|\eta') + C\lambda = 0 \\ C'\beta = 0, \end{cases}$$

where $\eta = (\beta \ \gamma)$, $s(\eta|\eta')$ denotes the gradient of (5) with respect to η and λ is a vector of Lagrange multipliers. The AS algorithm proceeds by substituting $s(\eta|\eta')$ with a first order linear approximation based on the Hessian $-I(\eta|\eta')$, that is,

$$s(\eta|\eta') \approx s(\eta_t|\eta') - (\eta - \eta_t)'I(\eta_t|\eta'),$$

where η_t is the value of η at the current iteration of the AS algorithm. The expression for $s(\eta|\eta')$ is available in closed form, while $I(\eta_t|\eta')$ is obtained as minus the numerical first derivative of $s(\eta|\eta')$.

We now augment C with blocks of zeros to obtain D , so that the constraint $C'\beta = 0$ is equivalent to $D'\eta = 0$. Once we do this, the approximated system of non-linear equations is exactly solved by the updating rule

$$\eta_{t+1} = \eta_t + I(\eta_t|\eta')^{-1}s(\eta_t|\eta') - I(\eta_t|\eta')^{-1}D(D'I(\eta_t|\eta')^{-1}D)^{-1}(D'I(\eta_t|\eta')^{-1}s(\eta_t|\eta')). \quad (6)$$

The updating rule, which is similar to the Newton-Raphson one, is then iterated until convergence to obtain the current update of β and γ . The M step is terminated by an update of α obtained through the maximization of

$$\int \log(f(\theta_i|\alpha)) f(\theta_i|\alpha, Y, X, \beta', \gamma') d\theta_i.$$

Under the assumption of Gaussian random effects with variance α^2 , this is accomplished by the updating rule

$$\alpha^2 = \frac{1}{n} \sum_i \int \theta_i^2 f(\theta_i|\alpha, Y, X, \beta', \gamma') d\theta_i - \left(\int \theta_i f(\theta_i|\alpha, Y, X, \beta', \gamma') d\theta_i \right)^2,$$

where both integrals are approximated using the usual Gauss-Hermite quadrature rule. Similar updating rules are obtained under T and Laplace assumptions, while under logit-Beta assumptions one can use the method of moments. With latent class models, $F(\alpha)$ is expressed as $\Pr(\theta_i = \alpha_{j1}) = \alpha_{j2}$ for $j = 1, \dots, k$ and some $k \in \mathcal{N}$ (which can be chosen using AIC). One should therefore estimate two sets of α parameters, the latent locations α_1 and the latent masses α_2 . The first shall be updated within the AS algorithm, with the same expression as (6) after augmenting $s(\eta|\eta')$ and $I(\eta|\eta')$ with the opportune derivatives. The latent masses are instead obtained as in (4), based on the updates for the other parameters.

An Horvitz-Thompson estimator for the population size is given by

$$\hat{N} = \sum_{i=1}^n \left(1 - \prod_{j=1}^S (1 - p_j(0, \dots, 0, \hat{\theta}_i, X_i)) \right)^{-1},$$

where $p_j(\cdot)$ is obtained after plug-in the MLE for all parameters and $\hat{\theta}_i$ is the posterior expectation for θ_i , that is

$$\hat{\theta}_i = \int \theta_i f(\theta_i | \alpha, Y, X, \beta, \gamma) d\theta_i,$$

where the posterior for θ_i is conditional on the MLE and can be evaluated through the same quadrature rule used at the E step when F is continuous.

References

- Aitchison J, Silvey SD (1958) Maximum-likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics* 29:813–828
- Alho JM (1990) Logistic regression in capture-recapture models. *Biometrics* 46:623–635
- Bergsma WP (1997) Marginal models for categorical data. Tilburg University Press, Tilburg
- Coull BA, Agresti A (1999) The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics* 55:294–301
- Coull BA, Agresti A (2000) Random effects modeling of multiple binomial responses using the multivariate binomial logit-normal distribution. *Biometrics* 56(1):73–80
- Farcomeni A (2011) Recapture models under equality constraints for the conditional capture probabilities. *Biometrika* 98:237–242
- Farcomeni A, Scacciarelli D (2013) Heterogeneity and behavioural response in continuous time capture-recapture, with application to street cannabis use in Italy. *Annals of Applied Statistics* 7:2293–2314
- Farcomeni A, Tardella L (2012) Identifiability and inferential issues in capture-recapture experiments with heterogeneous detection probabilities. *Electronic Journal of Statistics* 6:2602–2626
- Huggins R (1989) On the statistical analysis of capture experiments. *Biometrika* 76:133–140
- Lang J (2004) Multinomial Poisson homogeneous models for contingency tables. *Annals of Statistics* 32:340–383
- Link WA (2003) Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics* 59(4):1123–1130
- Otis DL, Burnham KP, White GC, Anderson DR (1978) Statistical Inference From Capture Data on Closed Animal Populations. *Wildlife Monographs*
- Ramsey F, Usner D (2003) Persistence and heterogeneity in habitat association studies using radio tracking. *Biometrics* 59:331–339
- Sanathanan L (1972) Estimating the size of a multinomial population. *Annals of Mathematical Statistics* 43:142–152
- Yang HC, Chao A (2005) Modeling animals' behavioral response by Markov chain models for capture-recapture experiments. *Biometrics* 61:1010–1017