A Review of the Use of Conditional Likelihood in Capture-Recapture Experiments

Author(s): Richard Huggins and Wen-Han Hwang

Source: *International Statistical Review / Revue Internationale de Statistique*, December 2011, Vol. 79, No. 3 (December 2011), pp. 385-400

Published by: International Statistical Institute (ISI)

Stable URL: https://www.jstor.org/stable/41305057

A Review of the Use of Conditional Likelihood in Capture-Recapture Experiments

Richard Huggins[1] and Wen-Han Hwang[2]

[1]*Department of Mathematics and Statistics, The University of Melbourne, Victoria 3010, Australia*
[2]*Institute of Statistics, National Chung Hsing University, Taichung 40724, Taiwan*
E-mail: huggins@ms.unimelb.edu.au

Summary

We present a modern perspective of the conditional likelihood approach to the analysis of capture-recapture experiments, which shows the conditional likelihood to be a member of generalized linear model (GLM). Hence, there is the potential to apply the full range of GLM methodologies. To put this method in context, we first review some approaches to capture-recapture experiments with heterogeneous capture probabilities in closed populations, covering parametric and non-parametric mixture models and the use of covariates. We then review in more detail the analysis of capture-recapture experiments when the capture probabilities depend on a covariate.

Key words: Capture-recapture; generalized linear model; Horvitz-Thompson.

1 Introduction

Capture-recapture experiments are a powerful tool for estimating the size of a wildlife population. They have a long history in ecological statistics and present many challenges to a statistician. We argue that they are a clever resolution of the classical statistical problem of estimating both the size and probability of success in a binomial experiment. Normally, a capture-recapture experiment consists of a number of capture occasions upon which individuals from a population are captured, marked or their existing mark noted if they have been previously marked, and then released back into the population. Thus for each captured individual a capture history is recorded. Additionally, environmental covariates (temperature, rainfall, catch effort) and individual covariates (weight, wing length, sex, body condition) may be collected as well. The marking of individuals makes this sampling scheme distinct from the more traditional independent and identically distributed observations from a binomial population. In this article, we assume that the population is closed so that its size does not change through immigration, emigration, births, or deaths over the period of experiment. Excellent reviews of the methodology for closed population models are contained in Otis *et al.* (1978); Seber (1986, 1992, 2002); Schwarz & Seber (1999); Chao (2001); Amstrup *et al.* (2005).

Perhaps reflecting their origins in the work of Graunt and Laplace in the 17th and 18th centuries, the application of capture-recapture methods has grown well beyond the estimation of animal abundance, where much of their modern development has occurred. Beyond their classical applications in estimating animal abundance, they have also been used in software

reliability (Briand *et al.*, 2000; Yip *et al.*, 2003), the estimation of census undercount (Darroch *et al.*, 1993; Chao & Tsay, 1998), estimating species richness (Boulinier *et al.*, 1998; Chao, 2005; Gotelli & Colwell, 2010), and they have become widely used in epidemiology and public health (McCarthy *et al.*, 1993; Mastro *et al.*, 1994; LaPorte, 1994; Hook & Regal, 1995; Ismail *et al.*, 1999; Simondon & Khodja, 1999; Cormack, 1999; Chao *et al.*, 2001; Böhning *et al.*, 2004; Zwane *et al.*, 2004).

Our main concern is the conditional likelihood approach to the analysis of capture-recapture experiments when the capture probabilities depend on covariates. That is, a version of the heterogeneity model M_h of Otis *et al.* (1978) when the capture probabilities depend on individual covariates. In a general setting inference based on conditional likelihood was originally proposed by Kalbfleisch & Sprott (1973) and was originally adapted to capture-recapture experiments by Sanathanan (1972a) and later by Huggins (1989) and Alho (1990) to extend classical logistic regression methods.

In Section 2, we give some background on capture-recapture methods and the conditional likelihood approach. In doing so we also examine log-linear Poisson models where conditional likelihood yields a multinomial logit model. In Section 3, we discuss parametric and non-parametric approaches to mixture models for heterogeneous capture probabilities. In Section 4, we discuss the modelling of capture probabilities as functions of covariates. We show in Section 5 that the conditional likelihood of a covariate M_h model is in fact a generalized linear model (GLM) (McCullagh & Nelder, 1989) and whilst the link function itself is intractable, its inverse and derivative are easily derived and hence the model may be fitted using iterated weighted least squares (IWLS). Some discussion is contained in Section 6.

2 The Challenge

A classical problem in statistics is to estimate N and p from observations on $X \sim$ binomial (N, p), see for example Olkin *et al.* (1981). Estimation of p for known N is straightforward, whilst the estimation of N for known p is more difficult but has been resolved (Feldman & Fox, 1968). DasGupta & Rubin (2005) and the references within discuss the general problem of unknown N and p when we have independent observations $X_1, \ldots, X_n$ from binomial (N, p). They note estimates are inherently unstable and slight perturbations of the data can lead to large fluctuations in the value of the estimate. They obtain estimators that perform well as $n \to \infty$. Capture-recapture experiments resolve these issues by adopting an alternate sampling scheme. This relationship with a classical, almost ubiquitous statistical problem, helps explain the increasing importance of capture-recapture methodologies.

2.1 Classical Models

The general mark-recapture approach dates from at least the 17th and 18th centuries where Graunt (Hald, 1990) and Laplace (Cochran, 1978) developed ratio estimators through what is equivalent to a two-sample approach. This consists of marking the first sample, returning it to the population then taking a second sample from the population and noting which individuals in the second sample had been previously marked. For Laplace, the object was to estimate the population size of France, the first "sample" was the newly born infants in a year and the second was some communities taken from the general population of France. Graunt proposed several estimators of the size of the population of London. In one the first "sample" were annual burials and the second were families where it was noted if there had been a death in the family that year.

In Laplace's example, if M is the number of newborn babies, and $\hat{p}$ is the estimated proportion of newborn babies in the second sample, then an estimate of the population size is $\hat{N} = M/\hat{p}$. That the first sample is new born infants or a burial is a convenience. One could equally take a random sample from the population, mark them in some way and release them back into the population. This is the well known Lincoln-Petersen estimator, which yields a stable estimator of N based on two independent samples, as long as the first sample is marked.

Extensions of this model to multiple capture occasions and the recording of individual capture histories gives rise to a range of urn models that have been reviewed in Otis *et al.* (1978), more recently Chapter 2 of (Amstrup *et al.*, 2005) and elsewhere. In these models all individuals in the population have the same probability of capture on a given occasion, apart from allowing a response to previous capture. This probability could change from occasion to occasion but otherwise it was assumed constant from individual to individual. These models can be analyzed using maximum likelihood as in Otis *et al.* (1978) or martingale estimating equations (Yip, 1991).

2.2 The General Multinomial Model and Conditional Likelihood

For a closed population of unknown size N, we assume that a capture experiment is conducted over τ capture occasions. A capture history is a vector $\omega = (w_1, \ldots w_\tau)$ where w_i is 1 for a capture at occasion i and 0 otherwise. Let n_ω be the number of individuals that have the same capture history ω and let Ω be the collection of $2^\tau - 1$ different possible capture histories that could be observed in the experiment. We denote by 0 the capture history $(0, \ldots, 0)$. Define $D = \sum_{\omega \in \Omega} n_\omega$ as the number of distinct individuals that have been captured at least once so that $n_0 = N - D$ represents the unknown number of individuals that were never captured.

Let $p_\omega(\theta)$ denote the probability of observing an individual with capture history ω where θ is a vector of parameters, and let $p_0(\theta)$ be the probability an individual is never captured. For example, if θ is the probability an individual is captured on a given occasion, $p_\omega(\theta) = \prod_{j=1}^{\tau} \theta^{w_j}(1-\theta)^{1-w_j}$ or if the capture probability p has density $f(p; \theta)$ then $p_\omega(\theta) = \int \prod_{j=1}^{\tau} p^{w_j}(1-p)^{1-w_j} f(p; \theta) dp$. Clearly, the vector consisting of $\{n_0, n_\omega; \omega \in \Omega\}$ follows a multinomial distribution (Darroch *et al.*, 1993) and the likelihood is

$$L(N; \theta) = \frac{N!}{\prod_{\omega \in \Omega} n_\omega!(N-D)!} \prod_{\omega \in \Omega} p_\omega(\theta) p_0(\theta)^{N-D} \tag{1}$$

$$= \frac{D!}{\prod_{\omega \in \Omega}} \prod_{\omega \in \Omega} \left\{ \frac{p_\omega(\theta)}{1 - p_0(\theta)} \right\}^{n_\omega} \times \frac{N!}{D!(N-D)!} \{1 - p_0(\theta)\}^D p_0(\theta)^{N-D} \tag{2}$$

$$= L_1(\theta) \times L_2(N; \theta).$$

The first term $L_1(\theta)$ is the conditional likelihood (Kalbfleisch & Sprott, 1973) since it is the likelihood given the observed individuals are captured at least once. To estimate θ and N, one can of course maximize the likelihood (1) and obtain the maximum likelihood estimate. Alternatively, $L_1(\theta)$ in (2), which does not involve N, can be maximized to find an estimate of θ, say $\hat{\theta}$, then maximizing $L_2(\theta)$ yields the estimator $\hat{N} = D/(1 - p_0(\hat{\theta}))$ of N. Sanathanan (1972a) shows that the resulting population size estimator using conditional likelihood is larger than that of maximum likelihood estimate, nevertheless, both estimators are consistent and they have the same asymptotic distribution. Therefore, it has become common to conduct inference using conditional likelihood (Sanathanan, 1972b; Coull & Agresti, 1999; Mao & Lindsay, 2002; Böhning *et al.*, 2005).

2.3 Log-Linear Poisson Models

Poisson models are quite popular approaches to the M_h model even when the data are from a classical capture-recapture experiment. Consider a M_h model, when the number of capture occasions τ is large and the capture probabilities p_i are small so that the distribution of the capture frequency Y_i can be approximated by a Poisson distribution. In fact, the approximation is good even for a small τ (van der Heijden *et al.*, 2003a). It also arises from a continuous time capture-recapture model where one may regard the time at capture or recapture of an animal as a Poisson process in a time period $[0, \tau]$.

This approach regards the observed counts n_ω, $\omega \in \Omega$ as an incomplete 2^τ contingency table (Fienberg, 1972; Cormack, 1989; Cormack & Jupp, 1991; Sandland & Cormack, 1984; Coull & Agresti, 1999; Rivest & Lèvesque, 2001). These counts are supposed to have independent Poisson distributions with means $\mu_\omega = E(n_\omega) = N p_\omega(\theta)$. This allows flexible modelling of the dependence between capture occasions. Consider model M_t where the capture probabilities vary with time. Then

$$\mu_w = N \prod_{j=1}^{\tau} p_j^{\omega_j} (1 - p_j)^{1-\omega_j} = N \prod_{j=1}^{\tau} (1 - p_j) \prod_{j=1}^{\tau} \left(\frac{p_j}{1 - p_j} \right)^{\omega_j}$$

so that $\ell_\omega = \log \mu_\omega = \gamma + \omega^T \beta$ where $\gamma = \log N + \sum_{j=1}^{\tau} (1 - p_j)$ and the vector β has j-th element $\beta_j = \log \{p_j/(1 - p_j)\}$. Let the vector ℓ consist of the ℓ_ω, $\omega \in \Omega$. Then with $\theta = (\gamma, \beta^T)^T$ and X a design matrix with first column being 1's and rows corresponding to the appropriate ω, we have $\ell = X\theta$. If we fit a log-linear Poisson model, we may estimate γ and β. As $\mu_0 = N \prod_{j=1}^{\tau} (1 - p_j) = \exp(\gamma)$, we can estimate the expected number of individuals not captured and hence the population size. Standard errors are discussed in Cormack & Jupp (1991) and Rivest & Lèvesque (2001). An R package is available for the implementation of this approach (Baillargeon & Rivest, 2007).

In general, let $p^*(\theta) = \sum_{\omega \in \Omega} p_\omega(\theta) = 1 - p_0(\theta)$. Then in the Poisson model (Cormack, 1989)

$$L(N, \theta) = \prod_{\omega \in \Omega} \frac{e^{-N p_\omega(\theta)} (N p_\omega(\theta))^{n_\omega}}{n_\omega!}$$

$$= \frac{D!}{\prod_{\omega \in \Omega} n_\omega} \left\{ \frac{p_\omega(\theta)}{1 - p_0(\theta)} \right\}^{n_\omega} \times \frac{e^{-N p^*(\theta)} (N p^*(\theta))^D}{D!}. \tag{3}$$

The first term is the conditional likelihood $L_1(\theta)$ which, under the log-linear model is a multinomial logit. Moreover, as noted in Sanathanan (1972a) if we treat the population size N as a continuous parameter the maximum likelihood estimators of the model parameters θ from the Poisson model are the conditional maximum likelihood estimators. This relationship has been noted and exploited by many authors (McFadden, 1973; Palmgren, 1981; Baker, 1994).

The log-linear approach is widely used for list problems, where individuals are captured through their appearance on lists. These are similar to capture-recapture experiments but there is no temporal order. The applications can be found in census undercount (Darroch *et al.*, 1993; Chao & Tsay, 1998) and in public health (Hook & Regal, 1995; Cormack, 1999). A related but different approach arises in estimating the size of a population using multiple captures and covariates, giving rise to a truncated Poisson regression model (van der Heijden *et al.*, 2003a,b), see Section 4.1 for further details.

3 Mixture Models

The most difficult models to analyse are when the capture probabilities depend on individual heterogeneity (*h*) possibly combined with time (*t*) and behaviour (*b*) effects (Otis *et al.*, 1978). The most general model is model M_{tbh}, which involves all three effects. It has been known since Otis *et al.* (1978) that ignoring these effects may result in a biased estimate of the population size. For example, typically an estimator that ignores a behavioural effect will lead to a positive bias in the trap-shy case and a negative bias in the trap happy case (Otis *et al.*, 1978). van der Heijden *et al.* (2003a), Böhning & Schön (2005) and Hwang & Huggins (2005) show that the ignorance of heterogeneity effect yields a negatively biased estimator in general. In contrast, Rivest (2008) shows that ignoring a time effect may have little impact on the population size estimate if the true model is M_t or M_{th}.

3.1 Mixing Models

In the mixing model it is supposed that some component of the capture probabilities arise from a mixing distribution. As an example, let p_{ij} denote the capture probability for the *i*-th animal at the *j*-th capture occasion. Sanathanan (1972b); Agresti (1994); Coull & Agresti (1999) consider

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \alpha_i + \beta_j, \tag{4}$$

where $\beta_j, j = 1, \ldots, \tau$, are unknown parameters of time effects, but the heterogeneity effects $\alpha_i, i = 1, \ldots, N$, are regarded as independent and identically distributed random variables. Assume that the distribution of α_i is F_γ for some parameters γ. In this case, θ consists of β_j and γ, and $p_{\omega(\theta)} = \int p_{ij}^{w_j}(1 - p_{ij})^{1-w_j} dF_\gamma(\alpha_i)$. A typical choice for F_γ is a gamma distribution (Sanathanan, 1972b) or a normal distribution (Coull & Agresti, 1999). The likelihood (1) or the conditional likelihood may then be maximized. Note that, as the probabilities $p_{\omega(\theta)}$ involve integration, implementation of a numerical integration method such as Gaussian quadrature is usually necessary. An extension to a continuous time frailty capture-recapture model that incorporated covariates has been developed in Xu *et al.* (2007).

3.1.1 Finite mixture models

When the mixing distribution is discrete with finite support, we have a finite mixture model or a latent class model. As an illustration, we consider a heterogeneity M_h model. It is supposed that the individual capture probabilities arise from a mixing distribution $P(X = p_i) = \alpha_i, p_i \in (0, 1]$ for $i = 1, \ldots, m$ and $\sum_{i=1}^{m}\alpha_i = 1$. We note that θ is a collection of p_i and α_i, so that the probability an individual has captured *k* times is

$$P_k(\theta) = \sum_{i=1}^{m} \alpha_i \binom{\tau}{k} p_i^k (1 - p_i)^{\tau-k}, k = 0, 1, \ldots, \tau.$$

This expression may be substituted in (1) or (2), where there are often simplifications as in the present case. If we observe f_k individuals exactly *k* times for $k = 1, \ldots, \tau$ then the conditional likelihood is

$$L_1(\theta) = \frac{D!}{\prod_{k=1}^{\tau} f_k!} \prod_{k=1}^{\tau} \left\{ \frac{P_k(\theta)}{1 - P_0(\theta)} \right\}^{f_k}. \tag{5}$$

Generally, parameter estimation is carried out via an expectation-maximization (EM) algorithm. The model and estimation method has received much attention in the last decade (Norris & Pollock, 1996; Mao & Lindsay, 2002; Dorazio & Royle, 2003; Böhning *et al.*, 2004; Mao, 2007; Mao & You, 2009). Incorporation of time and behavioural effect has been considered in Pledger (2000) and Pledger & Phillpot (2008). Norris & Pollock (1998), Böhning *et al.* (2005) and Böhning & Schön (2005) develop the finite mixture model in the Poisson framework; see also Wang & Lindsay (2005) and Mao & Lindsay (2007) for some theoretical results.

3.2 Identifiability

It is now well known that for mixing models in general the population size is not identifiable (Link, 2003; Holzmann *et al.*, 2006) and Link (2003) concludes *"Unmodeled heterogeneity is a shipwreck"*. To see the problems, let $\pi(j)$ denote the probability an individual is captured exactly j times during the experiment so that $\pi = (\pi(0), \ldots, \pi(\tau))$ is a probability distribution over $0, 1, \ldots, \tau$. We may specify the experiment by (π, N) or equivalently $(\pi^c, \pi(0), N)$ where $\pi^c(j) = \pi(j)/\{1 - \pi(0)\}$ is a conditional probability distribution over $1, \ldots, \tau$. Recall f_j is the number of individuals captured exactly j times and $D = \sum_{j=1}^{\tau} f_j$ is the number of distinct individuals observed. We do not observe f_0, but only $f_1, \ldots, f_\tau$. Thus, we may estimate $\pi^c(j)$, $j = 1, \ldots, \tau$. Link (2003) gives examples of parametric models where $\pi^c = \tilde{\pi}^c$ but $\pi(0) \neq \tilde{\pi}(0)$. For a simple example, suppose $\tau = 4$ and $\pi = (0.2, 0.1, 0.1, 0.1, 0.5)$. Then $\pi(0) = 0.2$ and $\pi^c = (1/8, 1/8, 1/8, 5/8)$. However, $\tilde{\pi} = (0.8, 0.025, 0.025, 0.025, 0.125)$ also gives π^c. If we observed $D = 100$, the estimators $\widehat{N}_\pi = D/\{1 - \pi(0)\}$ and $\widehat{N}_y = D/\{1 - \tilde{\pi}(0)\}$ would give estimated population sizes of 125 and 500, respectively. Often at best one can infer lower bounds (Mao, 2007) along the lines of Chao (1989).

3.3 Sample Coverage and Other Non-Parametric Methods

In addition to parametric approaches to modelling heterogeneity several non-parametric estimators that are not based on the likelihood have been proposed and despite any reservations arising from Section 3.2 they can perform well in practice.

A non-parametric approach that can work well in practice is the sample coverage estimator. The idea of sample coverage, originally due to I. J. Good and A. M. Turing (Good, 1953) has been used in species and population size estimation. It has been applied to capture-recapture experiments with heterogeneous capture probabilities (Chao *et al.*, 1992). In model M_h, the capture probabilities p_i, $i = 1, \ldots, N$ are supposed to be randomly sampled from some distribution but no further parametric assumption is required. Denote the mean of these capture probabilities by $\overline{p} = \sum_{i=1}^{N} p_i/N$ and the coefficient of variation by $\gamma = [\sum_{i=1}^{N}(p_i - \overline{p})^2/N]^{1/2}/\overline{p}$. Let Y_i be the number of times individual i has been captured. The sample coverage is defined by $C = \sum_{i=1}^{N} p_i I(Y_i > 0) / \sum_{i=1}^{N} p_i$ which is the proportion of the total probabilities that are observed from the sample. Taking the expectation, we have $E(C) \approx 1 - \sum_{i=1}^{N} \tau p_i(1 - p_i)^{\tau-1}/(\sum_{i=1}^{N} \tau p_i)$ so that an approximated moment estimator is $\hat{C} = 1 - f_1/(\sum_{k=1}^{\tau} k f_k)$. The estimator has the same form as that of Good (1953), who considers a multinomial model with $\sum_i p_i = 1$. When τ is small, the estimator may be biased and a further bias-correction of $\hat{C}$ can be found in Chao *et al.* (1992). Notice that $C = D/N$ if all p_i are equal, hence a convenient population size estimator is $\widehat{N}_0 = D/\hat{C}$ in the homogeneous case. However, it can result in a negative bias as the p_i are generally not equal. Using a Taylor expansion, the asymptotic bias is $E\widehat{N}_0 \approx N - (Ef_1/EC)\gamma^2$. The bias corrected sample coverage estimator is therefore taken to be $\widehat{N} = D/\hat{C} + (f_1/\hat{C})\hat{\gamma}^2$, where $\hat{\gamma}^2 = \max\{\{\tau \widehat{N}_0 \sum_k k(k-1)f_k\}/\{(\tau-1)(\sum_k f_k)^2\} - 1, 0\}$. There are many extensions

of this approach: Lee & Chao (1994) consider a M_{tbh} model where the ratio of time effects are given, Chao & Tsay (1998); Chao *et al.* (2001) apply it to list dependence for epidemiology data; Chao & Lee (1992) address the estimation of species richness or the number of classes; Chao *et al.* (2001); Huggins & Chao (2002) develop a martingale setting in a general M_{tbh} model. The concept of sample coverage is also useful in the Poisson framework, see Chao *et al.* (1993) and its application to determine the stopping rule in a reliability problem.

There are other methods that do not rely on the likelihood approach. Burnham & Overton (1978, 1979) propose a non-parametric approach to a heterogeneity model using the jackknife procedure. They obtain the k-th order jackknife estimator is $\widehat{N}_k = D + \sum_{i=1}^{k} (-1)^{i+1} \binom{k}{i} f_i$ and provide a sequential test to select the order k. Also in a non-parametric M_h model, Chao (1989) apply the Cauchy-Schwarz inequality and establish a simple lower bound estimator with $\widehat{N} = D + (\tau - 1)f_1^2/(2\tau f_2)$ or $\widehat{N} = D + f_1^2/(2f_2)$ when τ is not too small. The estimator has been shown to perform well in species richness estimation and is called the Chao 1 estimator in the ecology literature (Colwell & Coddington, 1994; Gotelli & Colwell, 2010). Moreover, the Chao 1 estimator can be derived from the Poisson model as well (Chao, 1987).

4 Covariate Models

An alternative approach to dealing with heterogeneity, dating from Pollock *et al.* (1984), is to model the capture probabilities as functions of covariates. For example, if the capture probabilities are related to the body weights x_i, then an analogous model to (4) is

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \alpha x_i + \beta_j, \tag{6}$$

which is a logistic regression model. The parametric regression model may also include a behavioural response, see Huggins (1989); Yang & Chao (2005). However, as we do not observe the covariates x_i of uncaptured individuals, unlike the mixing model we cannot obtain estimates from the full likelihood without further assumptions. The conditional likelihood approach (Huggins, 1989, 1991; Alho, 1990) has became the standard methodology. Let $\pi_i = 1 - \prod_{j=1}^{\tau}(1 - p_{ij})$ be the probability of being captured at least once given the covariate x_i. Let $Y_{ij} = 1$ when the i-th individual is captured at the j-th occasion and otherwise $Y_{ij} = 0$. Without loss of generality, we label $i = 1, \ldots, D$, for those individuals who have been captured at least once. Then the conditional likelihood is

$$L_1(\theta) = \prod_{i=1}^{D} \frac{\prod_{j=1}^{\tau} p_{ij}^{Y_{ij}}(1 - p_{ij})^{1-Y_{ij}}}{\pi_i}.$$

Having estimated model parameters in the conditional likelihood, the capture probabilities and hence the probability of being captured at least once may be estimated for each captured individual by $\widehat{\pi}_i$. The Horvitz-Thompson estimator (Horvitz & Thompson, 1952) allows estimation of the population size by $\widehat{N} = \sum_{i=1}^{D} \widehat{\pi}_i^{-1}$. The conditional likelihood method is incorporated in the MARK program (White & Burnham, 1999) to analyse capture-recapture data from closed populations and the function F.huggins.estim in a R package mra (McDonald, 2010) is also available.

4.1 Log-Linear Poisson Models: Heterogeneity & Covariates

For a continuous M_{th} model, Yip *et al.* (1996); Chen (2001) consider a Poisson process where the intensity function of the i-th animal is $\lambda(t) \exp(\beta x_i)$, where $\lambda(t)$ is an unknown

baseline intensity function. As a consequence, the marginal distribution of the capture frequency at time τ is a Poisson distribution. Moreover, Hwang & Huang (2007) indicate there is no information loss when estimating parameters and population size based on the (marginal) Poisson distribution.

Log-linear Poisson models also allow the incorporation of covariates (Tilling & Sterne, 1999) and when the capture probabilities are related to some covariates, van der Heijden *et al.* (2003a,b); Böhning & van der Heijden (2009) propose the use of a Poisson regression model. As noted by Zwane & van der Heijden (2004) following McFadden (1973) conditional on being captured at least once, the log linear model yields a multinomial logit. Smoothing spline models have been used in the log-linear formulation of the problem (Zwane & van der Heijden, 2004). That is, the conditional likelihood using the multinomial logit arising from (3) can be derived for each individual as a function of covariates x and parameter vector θ. The product of these terms over the captured individuals gives a conditional likelihood. Following Zwane & van der Heijden (2004) standard implementations of the multinomial likelihood, such as VGAMS (Yee & Wild, 1996; Yee, 2010), may be used to compute estimates $\hat{\theta}$ of θ. Zwane & van der Heijden (2004) adapt the approach of Darroch *et al.* (1993) to obtain an estimate $\widehat{\mu}_0(x_i; \hat{\theta})$ of $\mu_0(x_i; \theta)$ and then take $\widetilde{N} = D + \sum_{i=1}^{D} \widehat{\mu}_0(x_i; \hat{\theta})$. The standard error of this estimator may be estimated using the parametric bootstrap (Zwane & van der Heijden, 2003, 2004).

5 Model M_h, GLM Formulation

5.1 Conditional Likelihood

In model $M_h, p_{ij} \equiv p_i, j = 1, \ldots, \tau$ and $\pi_i = 1 - (1 - p_i)^\tau$ and the number of captures, Y_i, of an individual given it has been captured at least once has a positive binomial distribution (Patil, 1962). The corresponding conditional log-likelihood is $\ell(\eta) = y\eta - b(\eta) + c(y; \tau)$ where $\eta = \log(p/(1 - p))$, $b(\eta) = \tau \log\{1 + \exp(\eta)\} + \log(1 - 1/(1 + \exp(\eta))^\tau)$ and $c(y; \tau) = \log\binom{\tau}{y}$. Note that $p = \exp(\eta)/(1 + \exp(\eta))$, and

$$\mu = b'(\eta) = \tau p(\eta)/\pi(\eta), \text{ and } b''(\eta) = V = \mu \{\mu(\pi - 1) + (1 - p)\}, \qquad (7)$$

which gives the conditional mean and variance of Y. In the GLM formulation we suppose $\eta = g(\mu)$ where $\eta = x^T \beta$, for some covariates x and parameter vector β. Thus, to specify the link we need to find a function g so that $\eta = g(\mu)$ or in our case solve $\mu = \tau p(\eta)/\{1 - \{1 - p(\eta)\}^\tau\}$ for η, which seems analytically intractable. However, it is not necessary to specify the link to fit a GLM and we only need to determine $\partial\mu/\partial\eta$ (McCullagh & Nelder, 1989, §2.5.1), which is given by $\partial\mu/\partial\eta = V\{p(\eta)(1 - p(\eta))\}^{-1}\partial p(\eta)/\partial\eta$. Standard results for GLM's (McCullagh & Nelder, 1989) now give the properties of the estimator. In particular, if X has i-th row x_i^T, $V(\beta) = \text{diag}(V_i)$, and noting that $g'(\mu_i) = 1/V_i$, then $\text{Var}(\hat{\beta} - \beta) = (X^T V(\beta)X^T)^{-1}$ and asymptotically $\hat{\beta} - \beta \sim N_p(0, (X^T V(\beta)X)^{-1})$.

5.2 Population Size

Let C_i take the value 1 if $Y_i > 0$. As $E(C_i) = \pi_i$, $E(\sum_{i=1}^{N} \pi_i^{-1} C_i) = N$ and $\sum_{i=1}^{N} \pi_i^{-1} C_i = \sum_{i=1}^{D} \pi_i^{-1}$ so that if the π_i are known, or can be estimated, we can use the Horvitz-Thompson estimator (Horvitz & Thompson, 1952) to give an unbiased or approximately unbiased estimate of the population size if we have the estimates of capture probabilities. Let $\widehat{\pi}_i = \pi(x_i^T \hat{\beta})$. Then we take $\widehat{N} = \sum_{i=1}^{D} \widehat{\pi}_i^{-1}$. As this estimator involves the parameter estimates from the conditional

likelihood we need to take this into account in computing the standard error. We show in the Appendix that an unbiased estimator of the asymptotic variance is

$$\widehat{\text{Var}}(\widehat{N}) = \sum_{i=1}^{D} \frac{1 - \widehat{\pi}_i}{\widehat{\pi}_i^2} + R(\widehat{\beta})^T X (X^T V(\widehat{\beta}) X)^{-1} X^T R(\widehat{\beta}) \tag{8}$$

where $R(\beta)^T = ((1 - \pi_1)\tau p_1/\pi_1^2, \ldots, (1 - \pi_D)\tau p_D/\pi_D^2)$. In the parametric model M_h the conditional likelihood estimates may be directly computed using the posbinomial family in the VGAM package (Yee, 2010) to fit the positive binomial model using the methods of Yee & Wild (1996).

REMARK.   In the log-linear Poisson model, let λ be the mean capture frequency for an individual and so that $\pi = 1 - \exp(-\lambda)$. Then the conditional log-likelihood is $\ell(\eta) = y\eta - b(\eta) + c(y)$ where $\eta = \log(\lambda)$, $b(\eta) = \exp(\eta) + \log\{1 - \exp(-\exp(\eta))\}$ and $c(y) = -\log(y!)$. Then the mean and variance are $\mu = b'(\eta) = \lambda(\eta)/\pi(\eta)$, and $b''(\eta) = V = \mu\{1 + \lambda(\eta) - \mu\}$. To fit the Poisson model, we may use the pospoisson family in the VGAM package.

5.3 Extensions and Future Work

Section 5 concentrated on model M_h but the model may be extended to incorporate model M_{tbh} as in the original work of Huggins (1989). At present this requires the use of specialized software rather than being included as a family in existing software as is the case with model M_h and the VGAM family posbinomial. We therefore defer a study of models with time varying covariates and time effects, model M_{th} to another article. Ideally we require a generalization of the positive binomial model to a positive multiple Bernoulli model where we observe the outcome of a sequence of Bernoulli trials only if there is at least one success. At the least, it would be a boon to those estimating population size from capture-recapture data if the positive binomial model was included in the available families for any procedure based around GLM's. The use of log-linear Poisson models does allows the fitting of these models through GLMs and the Poisson family function.

Under the GLM framework, it is straightforward to extend the conditional likelihood model of Section 5 to include some modern statistical methods. For example, we may suppose that η is a smooth function of the covariates x rather than a parametric function leading to a non-parametric GLM (Green & Silverman, 1994). Huggins & Hwang (2007) apply local polynomial models (Fan & Gijbels, 1996) to estimate the smooth function using kernel smoothing methods. Similarly, Hwang & Huggins (2007) have also considered a semi-parametric model that allows η consists of a parametric component and a non-parametric component. These models may also be fitted to model M_h using the VGAM package but at present bootstrap methods are required to estimate standard errors.

A second extension is to adjust for the effects of measurement errors. Traditionally, the covariates have been assumed to be measured without error, but this may be unrealistic in practice. It is well known that ignoring measurement errors can yield inconsistent estimation in linear regression models and generalized linear models (Carroll *et al.*, 2006). In addition to obtaining a biased estimate for the regression parameter in p_i, Hwang & Huang (2003) showed that the population size estimator is also sensitive to measurement errors in the covariates. Again, using the GLM formulation, Hwang *et al.* (2007) apply the conditional score method of Stefanski & Carroll (1987) to establish a consistent estimate of the regression parameter and developed a consistent population size estimator based on the conditional score approach.

We saw that the covariate M_h model is within the class of GLM, hence the usual model diagnostic methods (e.g. residuals analysis and goodness of fit), model comparison or model selection may be applied. Nevertheless, as the main concern of the capture-recapture model is to estimate the population size rather than attain the best fit in the GLM part of the model further research is necessary.

In practice, some covariates may be missing for a variety of reasons. A naive strategy is to simply delete individuals with missing values, i.e. complete-case analysis, but in general it is known that this may yield biased parameter estimates when the missing-data mechanism is related to other variables and there is a growing body of literature on missing data problems in general (Little & Rubin, 2002). However, the capture-recapture model is rarely addressed. Yip *et al.* (2005) and Xi *et al.* (2009) have considered the problem and develop a parametric likelihood approach to a case of missing completely at random (MCAR) where missing-data mechanism is totally irrelevant to other variables. As the assumption of MCAR is generally not true, extensions to other models would be important.

It is well known in GLMs that overdispersion may occur in both the binomial and Poisson families, where the variation of residuals is greater than expected under these models (McCullagh & Nelder, 1989). The same situation may occur in capture-recapture data. One possible approach is to incorporate a random effect (latent variable; mixed effect) into the model raising the possibility of extending the generalized linear mixed effect model (McCulloch *et al.*, 2008). The beta-binomial and negative binomial distribution are well known alternatives to the binomial and Poisson models. Cruyff & van der Heijden (2008) have applied the negative binomial regression model to estimate the population size. Through a simulation study and an example, they showed that ignoring an overdispersion may seriously underestimate the population size (and the coverage percentage of interval estimation is very low). At present there is no analogous result for the case of binomial model. Moreover, there is no theoretical examination of the effect of overdispersion on population size estimate.

For likelihood based models, the Akaike information criterion (AIC) or its correction for finite sample sizes AICc are often used, see also Burnham & Anderson (2002). It remains an open problem to perform model selection among the non-likelihood based models (or the combination of likelihood and non-likelihood based), for example, the non-parametric models in Chao *et al.* (2001); Huggins & Chao (2002).

Robustness can also be an issue. The Horvitz-Thompson estimator is the sum of the inverses of estimated probabilities and is sensitive to small estimated probabilities. Following Cantoni & Ronchetti (2001); Heritier *et al.* (2009) it is feasible to robustify the GLM equations by extending known results for GLMs. However, even if we do have robust estimators of the model parameters, the estimated capture probability corresponding to an outlying covariate value may still be extremely small resulting in inflated estimates of the population size. That is, the model that is fitted to the bulk of the data may not be appropriate at the extreme covariate values. The same problem can occur in non-parametric estimation or b-spline models with high degrees of freedom as in this case extreme values of the covariate are not smoothed, as they are not close to the remainder of the data, and this can lead to inflated population size estimators. This is an area of ongoing research.

Finally, the focus on the estimation of the capture probabilities in the conditional likelihood approach suggests a further application of capture-recapture methods in sampling where the sampling probabilities are heterogeneous. The Horvitz-Thompson estimator has many applications, e.g. Overton & Stehman (1995), but requires inverse probability weights. If inclusion probabilities cannot be determined from the stratification or other means, the use of a capture-recapture design allows the estimated capture probabilities to be incorporated in the more general Horvitz-Thompson estimators.

6 Discussion

The genesis of the capture-recapture methodology was in the estimation of the size of human populations. They subsequently became a major tool in ecology where the estimation of both the abundance of a single species and species richness are important. The methods then returned to the estimation of human populations in epidemiological applications. We saw in Section 2 that a capture-recapture experiment is a sampling scheme that allows us to reliably estimate both N and p in a binomial experiment. Thus, the wide range of applications should not be surprising. Despite its size, our bibliography is necessarily incomplete and there is a huge literature on the development and application of capture-recapture methods. Despite this they currently lack much of the sophistication of methods based on the traditional GLM's with independent samples.

We have seen that the log-linear Poisson models estimate the parameters in a conditional multinomial logit model and hence only differ from our approach through being log-linear rather than logistic models and there is no deeper difference. The log-linear models incorporating covariates can be fitted using standard software at the cost of some complexity and perhaps a restriction due to the size of the "data" that must be created: $2^\tau - 1$ capture histories for each individual.

The implementation of the conditional likelihood approach to the analysis of capture-recapture data typically requires the development of new theory and the writing of dedicated software. This has limited the adoption of new statistical methodology into the analysis of mark recapture data. The increasing sophistication of GLM methods now allows some of these models to be applied using standard packages with simple add on functions to estimate the population size. For example, we can use the posbinomial family in the R package VGAM package (Yee, 2010) to allow the use of standard GLM methods in the analysis of capture-recapture experiments. Stoklosa *et al.* (2011) have recently developed a partial likelihood method that allows the direct use of standard GLM methodology such as SIMEX (Cook, 1994) and GAMS (Wood, 2006). The cost of this is that the partial likelihood conditions on the first capture and some information is lost.

References

Agresti, A. (1994). Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics*, **50**, 494–500.

Alho, J.M. (1990). Logistic regression in capture-recapture models. *Biometrics*, **46**, 623–635.

Amstrup, S.C., McDonald, T.L. & Manly, B.F.J. (2005). *Handbook of Capture-Recapture Analysis*. New Jersey: Princeton University Press.

Baker, S.G. (1994). The multinomial-Poisson transformation. *J. R. Stat. Soc. Ser. D*, **43**, 495–504.

Baillargeon, S. & Rivest, L.P. (2007). Rcapture: loglinear models for capture-recapture in R. *J. Stat. Softw.*, **19**, http://www.jstatsoft.org/.

Böhning, D., Dietz, E., Kuhnert, R. & Schön, D. (2005). Mixture models for capture-recapture count data. *Stat. Methods Appl.*, **14**, 29–43.

Böhning, D. & Schön, D. (2005). Nonparametric maximum likelihood estimation of the population size based on the counting distribution. *J. R. Stat. Soc. Ser. C*, **54**, 721–737.

Böhning, D., Suppawattanabodee, B., Kusolvisitkul, W. & Viwatwongkasem, C. (2004). Estimating the number of drug users in Bangkok 2001: a capture-recapture approach using repeated entries in one list. *Eur. J. Epidemiol.*, **19**, 1075–1083.

Böhning, D. & van der Heijden, P.G.M. (2009). A covariate adjustment for zero-truncated approaches to estimating the size of hidden and elusive populations. *Ann. Appl. Stat.*, **3**, 595–610.

Boulinier, T., Nichols, J.D., Sauer, J.R., Hines, J.E. & Pollock, K.H. (1998). Estimating species richness: the importance of heterogeneity in species detection. *Ecology*, **79**, 1018–1028.

Briand, L.C., Emam, K.E., Freimut, B.G. & Laitenberger, O. (2000). A comprehensive evaluation of capture-recapture models for estimating software defect content. *IEEE Trans. Soft. Eng.*, **26**, 518–540.

Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. New York: Springer-Verlag.

Burnham, K.P. & 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. & Overton, W.S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology*, **60**, 927–936.

Cantoni, E. & Ronchetti, E. (2001). Robust inference for generalized linear models. *J. Amer. Statist. Assoc.*, **96**, 1022–1030.

Carroll, R.J., Ruppert, D., Stefanski, L.A. & Crainiceanu, C.M. (2006). *Measurement Errors in Nonlinear Models: A Modern Perspective*, 2nd ed. London: Chapman & Hall.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**, 783–791.

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

Chao, A. (2001). An overview of Closed capture-recapture models. *J. Agric. Biol. Environ. Stat.*, **6**, 158–175.

Chao, A. (2005). Species richness estimation. In *Encyclopedia of Statistical Sciences*, 2nd ed. Eds. N. Balakrishnan, C. B. Read & B. Vidakovic, pp. 7907–7916. New York: Wiley.

Chao, A. & Lee, S.M. (1992). Estimating the number of classes via sample coverage. *J. Amer. Statist. Assoc.*, **87**, 210–217.

Chao, A., Lee, S.M. & Jeng S.L. (1992). Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics*, **48**, 201–216.

Chao, A., Ma, M.C. & Yang, M.C.K. (1993). Stopping rule and estimation for recapture debugging with unequal detection rates. *Biometrika*, **80**, 193–201.

Chao, A. & Tsay, P.K. (1998). A sample coverage approach to multiple-system estimation with application to census undercounts. *J. Amer. Statist. Assoc.*, **93**, 283–293.

Chao, A., Tsay, P.K., Lin, S.H., Shau, W.Y. & Chao, D.Y. (2001). The applications of capture-recapture models to epidemiological data. *Stat. Med.*, **20**, 3123–3157.

Chao, A., Yip, P., Lee, S.M. & Chu, W. (2001). Population size estimation based on estimating functions for closed capture recapture models. *J. Statist. Plann. Inference*, **92**, 213–232.

Chen, K. (2001). Parametric and semiparametric models for recapture and removal studies: a likelihood approach. *J. R. Stat. Soc. Ser. B*, **63**, 607–619.

Cochran, W.G. (1978). Laplace's ratio estimators. *Contributions to Survey Sampling and Applied Statistics*, Ed. H.A. David, pp. 3–10. New York: Academic Press.

Colwell, R.K. & Coddington, J.A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. R. Soc. London, Ser. B*, **345**, 101–118.

Cook, J.R. & Stefanski, L.A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *J. Amer. Statist. Assoc.*, **89**, 1314–1328.

Cormack, R.M. (1989). Log-linear models for capture-recapture. *Biometrics*, **45**, 395–413.

Cormack, R.M. (1999). Problems with using capture-recapture in epidemiology: an example of a measles epidemic. *J. Clin. Epidemiol.*, **52**, 909–914.

Cormack, R.M. & Jupp, P.E. (1991). Inference for Poisson and multinomial models for capture-recapture experiments. *Biometrika*, **78**, 911–916.

Coull, B.A. & Agresti, A. (1999). The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics*, **55**, 294–301.

Cruyff, M.J. & van der Heijden P.G. (2008). Point and interval estimation of the population size using a zero-truncated negative binomial regression model. *Biom. J.*, **50**, 1035–1050.

Darroch, J.N., Fienberg, S.E., Glonek, G.F.V. & Junker, B.W. (1993). A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *J. Amer. Statist. Assoc.*, **88**, 1137–1148.

DasGupta, A. & Rubin, H. (2005). Estimation of binomial parameters when both n, p are unknown. *J. Statist. Plann. Inference*, **130**, 391–404.

Dorazio, R.M. & Royle, J.A. (2003). Mixture models for estimating the size a closed population when capture rates vary among individuals. *Biometrics*, **59**, 351–364.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. London: Chapman & Hall.

Feldman, D. & Fox, M. (1968). Estimation of the parameter n in the binomial distribution. *J. Amer. Statist. Assoc.*, **63**, 150–158.

Fienberg, S.E. (1972). The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika*, **59**, 591–603.

Good, I.J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, **40**, 237–264.

Gotelli, N.J. & Colwell, R.K. (2010). Estimating species richness. *Frontiers in Measuring Biodiversity*, Ed. A. Magurran & B. McGill, pp. 39–54. New York: Oxford University Press.

Green, P.J. & Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. London: Chapman & Hall.

Hald, A. (1990). *A History of Probability and Statistics and Their Applications before 1750*. New York: Wiley.

Heritier, S., Cantoni, E., Copt, S. & Victoria-Feser, M.P. (2009). *Robust Methods in Biostatistics*. Chichester, West Sussex: Wiley.

Holzmann, H., Munk, A. & Zucchini, W. (2006). On identifiability in capture-recapture models. *Biometrics*, **62**, 934–939.

Hook, E.B. & Regal, R.R. (1995). Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol. Rev.*, **17**, 243–264.

Horvitz, D. & Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, **260**, 663–685.

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

Huggins, R.M. (1991). Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics*, **47**, 725–732.

Huggins, R.M. & Chao, A. (2002). Asymptotic properties of an optimal estimating function approach to the analysis of mark recapture data. *Comm. Statist.*, **31**, 575–597.

Huggins, R.M. & Hwang, W.H. (2007). Nonparametric estimation of population size from capture-recapture data when the capture probability depends on a covariate. *J. R. Stat. Soc. Ser. C*, **56**, 429–443.

Hwang, W.H. & Huang, S.Y.H. (2003). Estimation in capture-recapture models when covariates are subject to measurement errors. *Biometrics*, **59**, 1113–1122.

Hwang, W.H. & Huang, Y.H. (2007). Measurement errors in continuous-time capture-recapture models. *J. Statist. Plann. Inference*, **137**, 1888–1899.

Hwang, W.H., Huang, S.Y.H. & Wang, C.Y. (2007). Effects of measurement error and conditional score estimation in capture-recapture models. *Statist. Sinica*, **17**, 301–316.

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

Hwang, W.H. & Huggins, R.M. (2007). Application of semiparametric regression models in the analysis of capture-recapture experiments. *Aust. N. Z. J. Stat.*, **49**, 191–202.

Ismail, A.A., Beeching, N.J., Gill, G.V. & Bellis, M.A. (1999). Capture-recapture-adjusted prevalence rates of type 2 diabetes are related to social deprivation. *Q. J. Med.*, **92**, 707–710.

Kalbfleisch, J.D & Sprott, D.A. (1973). Marginal and conditional likelihoods. *Sankhya Ser. A*, **35**, 311–328.

LaPorte, R.E. (1994). Assessing the human condition: capture-recapture techniques. *BMJ*, **308**, 5–6.

Lee, S.M. & Chao, A. (1994). Estimating population size via sample coverage for closed capture-recapture models. *Biometrics*, **50**, 88–97.

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

Little, R.J.A. & Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. New Jerssy: Wiley.

Mao, C.X. (2007). Estimating population sizes for capture-recapture sampling with binomial mixtures. *Comput. Statist. Data Anal.*, **51**, 5211–5219.

Mao, C.X. & Lindsay, B.G. (2002). Diagnostics for the homogeneity of inclusion probabilities in a Bernoulli census. *Sankhya Ser. A*, **64**, 626–639.

Mao, C.X. & Lindsay, B.G. (2007). Estimating the number of classes. *Ann. Statist.*, **35**, 917–930.

Mao, C.X. & You, N. (2009). On comparison of mixture models for closed population capture-recapture studies. *Biometrics*, **65**, 547–553.

Mastro, T.D., Kitayaporn, D., Weninger, B.G., Vanichseni, S., Laosunthorn, V., Uneklabh, T., Uneklabh, C., Choopanya, K. & Limpakarnjanarat, K. (1994). Estimating the number of HIV-infected injection drug users in Bangkok: a capture-recapture method. *Amer. J. Public Health*, **84**, 1094–1099.

McCarthy, D.J., Tull, E.S., Moy, C.S., Kwoh, C.K. & Laporte, R.E. (1993). Ascertainment of corrected rates: applications of capture-recapture methods. *Int. J. Epidemiol.*, **22**, 559–565.

McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models*, 2nd ed. Boca Raton: Chapman & Hall.

McCulloch, C.E., Searle, S.R. & Neuhaus, J.M. (2008). *Generalized, Linear, and Mixed Models*. New York: Wiley.

McDonald, T. (2010). mra: Analysis of Mark-Recapture data. R package version 2.7. Available online at: http://CRAN.R-project.org/package=mra.

McFadden, D. (1973). Conditional logit analysis of qualitiative choice behavior. In *Frontiers in Econometrics*, Ed. P. Zareembka, pp. 105–142. New York: Academic Press.

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

Norris, J. L. & Pollock, K. H. (1998). Nonparametric MLE for Poisson species abundance models allowing for heterogeneity between species. *Environ. Ecol. Stat.*, **5**, 391–402.

Olkin, I., Petkau, I. & Zidek, J. (1981). A comparison of *n* estimators for the binomial distribution. *J. Amer. Statist. Assoc.*, **76**, 637–642.

Otis, D.L., Burnham, K.P., White, G.C. & Anderson, D.R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, **62**, 1–135.

Overton, W.S. & Stehman, S.V. (1995). The Horvitz-Thompson theorem as a unifying perspective for probability sampling: with examples from natural resource sampling. *Amer. Statist.*, **49**, 261–268.

Palmgren, J. (1981). The Fisher information matrix for log linear models arguing conditionally on observed explanatory variables. *Biometrika*, **68**, 563–566.

Patil, G.P. (1962). Maximum likelihood estimation for generalized power series distributions and its application to a truncated binomial distribution. *Biometrika*, **49**, 227–237.

Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models. *Biometrics*, **56**, 434–442.

Pledger, S. & Phillpot, P. (2008). Using mixtures to model heterogeneity in ecological capture-recapture studies. *Biom. J.*, **50**, 1022–1034.

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

Rivest, L.P. (2008). Why a time effect often has a limited impact on capture-recapture estimates in closed populations. *Canad. J. Statist.*, **36**, 75–84.

Rivest, L.P. & Lèvesque, T. (2001). Improved log-linear model estimators of abundance in capture-recapture experiments. *Canad. J. Statist.*, **29**, 555–572.

Sanathanan, L. (1972a). Estimating the size of a multinomial population. *Ann. Math. Statist.*, **43**, 142–152.

Sanathanan, L. (1972b). Models and estimation methods in visual scanning experiments. *Technometrics*, **14**, 813–829.

Sandland, R.L. & Cormack, R.M. (1984). Statistical inference for Poisson and multinomial models for capture-recapture experiments. *Biometrika*, **71**, 27–33.

Schwarz, C.J. & Seber, G.A.F. (1999). Estimating animal abundance: review III. *Statist. Sci.*, **14**, 427–456.

Seber, G.A.F. (1986). A review of estimating animal abundance. *Biometrics*, **42**, 267–292.

Seber, G.A.F. (1992). A review of estimating animal abundance II. *Int. Statist. Rev.*, **60**, 129–166.

Seber, G.A.F. (2002). *The Estimation of Animal Abundance and Related Parameters*, 2nd ed. Caldwel, New Jersey: Blackburn Press.

Simondon, F. & Khodja, H. (1999). Capture-recapture method for estimating misclassification errors: application to the measurement of vaccine efficacy in randomized controlled trials. *Int. J. Epidemiol.*, **28**, 113–116.

Stefanski, L.A. & Carroll, R.J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, **74**, 703–716.

Stoklosa, J., Hwang, W.H., Huggins, R.M. & Wu, S.H. (2011). Capture-recapture models with covariates: a partial likelihood approach. *Biometrics*, DOI: 10.1111/j.1541-0420.2011.01596.

Tilling, K. & Sterne, J.A.C. (1999). Capture-recapture models including covariate effects. *Amer. J. Epidemiol.*, **149**, 392–400.

van der Heijden, P.G.M., Bustami, R., Cruyff, M., Engbersen, G. & van Houwelingen, H.C. (2003a). Point and interval estimation of the population size using the truncated Poisson regression model. *Statist. Model.*, **3**, 305–322.

van der Heijden, P.G.M., Cruyff, M. & van Houwelingen, H.C. (2003b). Estimating the size of a criminal population from police records using the truncated Poisson regression model. *Statist. Neerlandica*, **57**, 289–304.

Wang, J.P. & Lindsay, B.G. (2005). A penalized nonparametric maximum likelihood approach to species richness estimation. *J. Amer. Statist. Assoc.*, **100**, 942–959.

White, G.C. & Burnham, K.P. (1999). Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, 120–138.

Wood, S.N. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton: Chapman & Hall.

Xi, L., Watson, R., Wang, J.P. & Yip, P. (2009). Estimation in capture-recapture models when covariates are subject to measurement errors and missing data. *Canad. J. Statist.*, **37**, 645–658.

Xu, Y., Liu, L., You, N., Pan, H. & Yip, P. (2007). Estimating population size for a continuous time frailty model with covariates in a capture-recapture study. *Biometrics*, **63**, 917–921.

Yang, H.C. & Chao, A. (2005). Modeling animals' behavioural response by Markov chain models for capture-recapture experiments. *Biometrics*, **61**, 1010–1017.

Yee, T.W. (2010). The VGAM package for categorical data analysis. *J. Stat. Softw.*, **32**, 1–35.

Yee, T.W. & Wild, C.J. (1996). Vector generalized additive models. *J. R. Stat. Soc. Ser. B*, **58**, 481–493.

Yip, P. (1991). A martingale estimating equation for a capture-recapture experiment in discrete time. *Biometrics*, **47**, 1081–1088.

Yip, P., Huggins, R.M. & Lin, D.Y. (1996). Inference for capture-recapture experiments in continuous time with variable capture rates. *Biometrika*, **83**, 477–483.

Yip, P., Lin, H.Z. & Xi, L. (2005). A semiparametric method for estimating population size for capture-recapture experiment with random covariates in continuous time. *Biometrics*, **61**, 1085–1092.

Yip, P., Wang, Y. & Chao, A. (2003). The application of capture-recapture methods in reliability studies. In *Handbook of Reliability Engineering*, Ed. H. Pham, pp. 493–510. London, New York: Springer.

Zwane, E.N. & van der Heijden, P.G.M. (2003). Implementing the parametric bootstrap in capture recapture models with continuous covariate. *Statist. Probab. Lett.*, **65**, 121–125.

Zwane, E.N. & van der Heijden, P.G.M. (2004). Semiparametric models for capture-recapture studies with covariates. *Comput. Statist. Data Anal.*, **47**, 729–743.

Zwane, E.N., van der Pal-de Bruin, K. & van der Heijden, P.G.M. (2004). The multiple-record systems estimator when registrations refer to different but overlapping populations. *Stat. Med.*, **23**, 2267–2281.

Résumé

Nous présentons une perspective moderne de l'approche par vraisemblances conditionnelles de l'analyse des expériences de capture-recapture. Nous montrons que ces vraisemblances conditionnelles relèvent d'un modèle linéaire généralisé, ce qui permet l'application des nombreuses méthodes élaborées dans ce cadre. Pour replacer ces applications dans leur contexte, nous passons en revue quelques-unes des approches existantes dans les modèles de capture-recapture avec probabilités de capture hétérogènes au sein de populations fermées. Nous décrivons, en particulier, l'utilisation de modèles de mélange paramétriques et non paramétriques, et examinons de façon plus détaillée le cas où les probabilités de capture sont fonction de covariables.

Appendix: Population Size Estimator

We outline the derivation of the variance of the population size estimator. Let $U(\beta)$ be the vector $Y_i - \mu_i(\beta)$. Then a second-order expansion yields $\widehat{\beta} - \beta = (X^T V(\beta) X)^{-1} X^T U(\beta) + O_p((\widehat{\beta} - \beta)^2)$ and $X^T V(\beta) X$ is of the form $\sum_{i=1}^{D} X_i^T X_i V_i^{-1}$. We suppose that $N^{-1} X^T V(\beta) X \xrightarrow{p} \Gamma$. Then in particular $(\widehat{\beta} - \beta)^2 = o_p(N^{-1/2})$. A second-order expansion also yields

$$\widehat{N} = \sum_{i=1}^{N} \frac{C_i}{\pi_i} + \sum_{i=1}^{N} \frac{C_i(\widehat{\pi}_i - \pi_i)}{\pi_i^2} + \sum_{i=1}^{N} C_i O_p((\widehat{\pi}_i - \pi)^2). \tag{A1}$$

Similarly, $\widehat{\pi}_i - \pi_i = (1 - p_i)^{\tau-1}\tau(\widehat{p}_i - p_i) + O_p((\widehat{p}_i - p_i)^2)$, and $\widehat{p}_i - p_i = p_i(1 - p_i)x_i^T(\widehat{\beta} - \beta) + O_p((\widehat{\beta} - \beta)^2)$ so that $O_p((\widehat{\pi}_i - \pi)^2) = O_p((\widehat{\beta} - \beta)^2)$ and hence $\widehat{\pi}_i - \pi_i = (1 - \pi_i)\tau p_i x_i^T (X^T V(\beta) X)^{-1} X^T U(\beta) + O_p((\widehat{\beta} - \beta)^2)$.

Thus,

$$\widehat{N} - N = \sum_{i=1}^{N} \frac{C_i - \pi_i}{\pi_i} \tag{A2}$$

$$+ \sum_{i=1}^{N} \frac{(1 - \pi_i)\tau p_i x_i^T C_i}{\pi_i^2} (X^T V(\beta) X)^{-1} X^T U(\beta) \tag{A3}$$

$$+ \sum_{i=1}^{N} C_i O_p((\widehat{\beta} - \beta)^2). \tag{A4}$$

Now, $E\{(Y_i - \mu(x_i^T \beta)|C_i\} = 0$, $E(C_i|x_i) = \pi_i$, $\text{Var}(C_i|x_i) = \pi_i(1 - \pi_i)$ and hence, with $\bar{\pi} = N^{-1} \sum_{i=1}^{N} \pi_i$, $E(\widehat{N}) = N + N\bar{\pi} O_p((\widehat{\beta} - \beta)^2) = N + o(N)$ so the bias is $o(N)$.

Note that as $N^{-1/2} \sum_{i=1}^{N} C_i O_p((\widehat{\beta} - \beta)^2) = N^{1/2} O_p((\widehat{\beta} - \beta)^2) = o_p(1)$ so we may ignore (A4) in deriving the central limit theorem. Consider (A2). Suppose for some constant π, $N^{-1} \sum_{i=1}^{N} (1 - \pi_i)/\pi_i \xrightarrow{P} \pi$, so that $N^{-1/2} \sum_{i=1}^{N} (C_i - \pi_i)/\pi_i \xrightarrow{D} N(0, \pi)$. Multiplying (A3) by $N^{-1/2}$ yields

$$N^{-1} \sum_{i=1}^{N} \frac{(1 - \pi_i)\tau p_i x_i^T C_i}{\pi_i^2} N^{1/2} (X^T V(\beta) X)^{-1/2} (X^T V(\beta) X)^{-1/2} X^T U(\beta) \tag{A5}$$

and we further suppose that $N^{-1} \sum_{i=1}^{N} (1 - \pi_i)\tau p_i x_i^T C_i/\pi_i^2 \xrightarrow{P} \Delta$ so the limiting distribution of (A5) is $N(0, \gamma^{-1} \Delta \Gamma^{-1} \Delta^T)$. Hence, as the first two terms in $\widehat{N}$ are uncorrelated $N^{-1/2}(\widehat{N} - N) \xrightarrow{D} N(0, \pi + \Delta \Gamma^{-1} \Delta^T)$ and $\text{Var}(\widehat{N}) \approx N(\pi + \Delta \Gamma^{-1} \Delta^T)$. An estimator of $N\pi$ is $\sum_{i=1}^{D} (1 - \widehat{\pi}_i)/\widehat{\pi}_i^2$. Now $N\Delta \approx R^T X$ and $N\Delta \Gamma^{-1} \Delta^T \approx R^T X (X^T V(\beta) X)^{-1} X^T R$, which yields (8).