The International Journal of Biostatistics

Manuscript 1283

A Lower Bound Model for Multiple Record Systems Estimation with Heterogeneous Catchability

Louis-Paul Rivest, *Université Laval*

A Lower Bound Model for Multiple Record Systems Estimation with Heterogeneous Catchability

Louis-Paul Rivest

Abstract

This work considers the estimation of the size N of a closed population using incomplete lists of its members. Capture histories are constructed by establishing the presence or the absence of each individual in all the lists available. Models for data featuring a heterogeneous catchability and list dependencies are considered. A log-linear model leading to a lower bound for the population size is derived for a known set of list dependencies and a latent catchability variable with an arbitrary distribution. This generalizes Chao's lower bound to models with interactions. The proposed model can be used to carry out a search for important list interactions. It also provides diagnostic information about the nature of the underlying heterogeneity. Indeed, it is shown that the Poisson maximum likelihood estimator of N under a dichotomous latent class model does not exist for a particular set of LB models. Several distributions for the heterogeneous catchability are considered; they allow to investigate the sensitivity of the population size estimate to the model for the heterogeneous catchability.

KEYWORDS: capture-recapture model, latent class model, mixture model, Poisson regression

Author Notes: The author is thankful to the referees for their perceptive comments and to the Canada Research Chair in Statistical Sampling and Data Analysis, to the Natural Science and Engineering Research Council of Canada and to the Fonds Québécois Pour la Nature et les Technologies for their financial support.

1 Introduction

The estimation of the size N of a closed population using capture recapture techniques is an important topic in population epidemiology. The capture occasions are incomplete lists of individuals in a target population. The capture history of an individual is obtained by ascertaining his presence, or absence, in all the lists available. The goal of the analysis is to estimate the total number of individuals N in the population; this amounts to predicting the number of individuals that do not appear on any list. An early discussion of this problem appears in IWGDMF (International Working Group for Disease Monitoring and Forecasting) (1995a,b) and several applications can be found in the literature, see for instance chapter 6 of Bishop, Fienberg, and Holland (1975), Bruno, Biggeri, LaPorte, McCarty, Merletti, and Pagano (1994), Murphy (2009). A key assumption highlighted by Hook and Regal (1993) is that of constant catchability by each source. This assumption is often questionable; some units in the population might have larger probabilities of being detected. For instance, when estimating the prevalence of a disease, patients that have a severe form of the illness might use more medical services and therefore appear on more lists, than those that are less affected. Models for such a heterogeneous catchability are investigated in this work.

Motivated by biological applications, the capture-recapture model with heterogeneous capture probabilities and no list effects has been investigated thoroughly. Huggins (2001), Link (2003) highlight the difficulties in estimating N in this context. They show that the distribution of the latent catchability is not estimable; different distributions can lead to models that fit equally well and that give drastically different estimates of N. Partial information on the size of the population is available from an estimator of a lower bound for N proposed in Chao (1998). Still, the problem is important and Link (2006) argues that "rather than relying on untestable assumptions, it seems advisable that we attempt to identify and control for sources of heterogeneity through appropriate covariate analysis." As shown in Section 6, even when important covariates are included in the model, some residual heterogeneity might be present. Techniques are needed to investigate how sensitive the estimates of population sizes are to the model for the heterogenous capture probabilities. Some are proposed in this paper.

In epidemiological data sets, list dependencies are frequent and the basic models developed for biological applications do not fit. Some techniques to cope with heterogeneous catchability are proposed in IWGDMF (International Working Group for Disease Monitoring and Forecasting) (1995a). Many analyses consist of a stepwise search of important log-linear list interactions, see Chao, Tsay, Lin,

Shau, and Chao (2001), Cormack, Chang, and Smith (2000) and Aaron, Chang, Markovic, and LaPorte (2003). More recently, Bartolucci and Forcina (2001, 2006), Stanghelli and van der Heijden (2004) use a few latent classes to account for the heterogeneity; they use discrete distributions for the heterogeneous catchability.

This work considers a mixture model similar to that presented in Bartolucci and Forcina (2001) and Stanghelli and van der Heijden (2004), with list dependencies and heterogeneous capture probabilities. First a lower bound for the population size for a known set of list dependencies that generalizes the bound of Chao (1998) is presented. This lower bound is valid for any distribution, discrete or continuous, of the latent catchability. It can be estimated by fitting the lower bound (LB) model, a log-linear model with a particular set of explanatory variables.

Models obtained with simple catchability distributions such as the logit normal of Coull and Agresti (1999) and the latent class model of Stanghelli and van der Heijden (2004), are considered. Situations where the maximum likelihood estimator of N for the latent class model does not exist are identified. Data dependent mixing distributions that lead to marginal log-linear models for the data are also investigated and their estimates are compared to those obtained with true mixing distributions. The methodology is illustrated by deriving closed form expressions for the new population size estimators when the data set is constructed with three lists and through numerical examples.

2 A general Rasch model for heterogeneous catchability

The data set for analysis contains the $2^t - 1$ frequencies $\{n_{\mathbf{x}}\}$ of the observable capture histories $\mathbf{x} = (x_1, x_2, \ldots, x_t)$, where t is the number of lists and x_i is 1 if the individual appears in the ith list and 0 otherwise for $i = 1, \ldots, t$. The general Rasch model involves the conditional probabilities $p(\mathbf{x}|\alpha)$ for an individual to get capture history $\mathbf{x}$ given that its catchability variable is α. This probability is written in terms of the list effects β_i for $i = 1, \ldots, t$ and of some list interactions λ_I as

$$p(\mathbf{x}|\alpha) = C(\alpha) \exp\left\{ \sum_{i=1}^{t} x_i(\beta_i + \alpha) + \sum_{I \in \mathscr{I}} L(\mathbf{x}, I)\lambda_I \right\}, \tag{1}$$

where $\mathscr{I}$ is a collection of subsets of $(1, \ldots, t)$ determining the interactions of interest, $L(\mathbf{x}, I) = \prod_{i \in I} x_i$, and $C(\alpha)$ is a normalizing constant such that $\sum p(\mathbf{x}|\alpha) = 1$ where the sum is on the 2^t possible capture histories. For the units that were missed, $\mathbf{x} = \mathbf{0}$ is a vector of t zeros and $L(\mathbf{0}, I) = 0$ for all the subsets I, thus the conditional probability of not appearing in any list is $p(\mathbf{0}|\alpha) = C(\alpha)$. Under (1), the variable α measures catchability; units that are likely to appear in several lists have large α values. Note that the log-linear interactions λ_I do not depend on α.

When $\mathscr{I} = \emptyset$ there are no interactions, and

$$p(\mathbf{x}|\alpha) = \frac{\exp\{\sum(\beta_i + \alpha)x_i\}}{\prod_{i=1}^{t}\{1 + \exp(\beta_i + \alpha)\}}.$$

This is the standard Rasch model where, given α, the lists are independent with capture probabilities $p_i(\alpha) = \exp(\beta_i + \alpha)/\{1 + \exp(\beta_i + \alpha)\}$. Suppose now that there is only one interaction, between the first two lists. One has $\mathscr{I} = \{(1,2)\}$ and

$$\begin{aligned} p(\mathbf{x}|\alpha) &= \frac{\exp\{(\beta_1 + \alpha)x_1 + (\beta_2 + \alpha)x_2 + \lambda_{12}x_1x_2\}}{[1 + \exp(\alpha + \beta_1) + \exp(\alpha + \beta_2) + \exp(2\alpha + \beta_1 + \beta_2 + \lambda_{12})]} \\ &\times \frac{\exp\{\sum_{i=3}^{t}(\beta_i + \alpha)x_i\}}{\prod_{i=3}^{t}\{1 + \exp(\beta_i + \alpha)\}}. \end{aligned}$$

Now assume that the catchability variable α is random with a cumulative distribution function $F(\alpha)$. The unconditional capture probabilities are given by

$$p(\mathbf{x}) = p(0)\exp\left\{\sum_{i=1}^{t} x_i\beta_i + \varphi(\sum_{i=1}^{t} x_i) + \sum_{I\in\mathscr{I}} L(\mathbf{x}, I)\lambda_I\right\}, \tag{2}$$

where $p(0) = \int C(\alpha)dF(\alpha)$ is the marginal probability of being missed and $\varphi(k)$ is the cumulant generating function of a density proportional to $C(\alpha)dF(\alpha)$,

$$\exp\varphi(k) = \int_{\Re} \exp(k\alpha)C(\alpha)dF(\alpha)/p(0).$$

Thus the marginal predicted values $\mu_{\mathbf{x}} = Np(\mathbf{x})$ satisfy

$$\log\mu_{\mathbf{x}} = \gamma + \sum_{i=1}^{t} x_i\beta_i + \varphi(\sum_{i=1}^{t} x_i) + \sum_{I\in\mathscr{I}} L(\mathbf{x}, I)\lambda_I, \tag{3}$$

where $\gamma = \log\{Np(0)\}$ is the log-predicted value for the units missed in the experiment. If $F(\alpha)$ belongs to a parametric family of distributions, such as the normal with an unknown variance, then the parameters of (3) are identifiable provided that $\mathscr{I}$ is not too large. On the other hand, if $F(\alpha)$ is unknown then $\varphi(1), \ldots, \varphi(t)$ are parameters to be estimated. The log-linear parameters γ, $\{\beta_i\}$, $\{\lambda_I\}$, and the mixing parameters $\{\varphi(k)\}$ are not identifiable; indeed it is shown in the next section shows that only a subspace of dimension $k-2$ of $\{\varphi(k)\}$ can be identified.

It is interesting to compare (3) to the dichotomous latent class (LC) model of Stanghelli and van der Heijden (2004). They have an unobserved dichotomous heterogeneity variable U and, when there are no covariates, their model is defined by specifying log-linear interactions involving U and the lists E_i, $i = 1, \ldots, t$. If

the heterogeneity variable takes only two values, 0 and $\alpha > 0$, (3) is similar to Stanghellini and van der Heijden's model. Then $\exp\alpha$ can be interpreted as the odds ratio for the conditional association between U and E_1, U and $E_2, \ldots, U$ and E_t. Thus (3) then models the heterogeneity by having the same parameter α for the log-linear interactions $[UE_1]$, $[UE_2], \ldots [UE_t]$ while Stanghelli and van der Heijden (2004) allow this interaction to vary with E_i. This LC model is considered in Section 4.2.

One can reexpress (3) as a standard hierarchical log-linear model with list interactions. This is done by expanding $\varphi(\sum x_i)$ using Δ, the forward difference operator, defined by $\Delta\varphi(k) = \varphi(k+1) - \varphi(k)$. It is shown in the Appendix that

$$\varphi(\sum x_i) = \sum_{m=0}^{t} \sum_{I \in \mathscr{I}(m)} L(\mathbf{x}, I)\Delta^m \varphi(0), \tag{4}$$

where $\mathscr{I}(m)$ is the set of the subsets of $(1, \ldots, t)$ containing exactly m elements. Thus, in general, all the log-linear parameters of (3) are non zero. The log-linear interaction for a set I with m elements is $\lambda_I + \Delta^m\varphi(0)$ if $I \in \mathscr{I}$ and $\Delta^m\varphi(0)$ otherwise. As shown in Mao (2008), see also p. 452 of Marshall and Olkin (1979),

$$\Delta^2\varphi(0) = \varphi(2) - 2\varphi(1) + \varphi(0) = \varphi(2) - 2\varphi(1) \geq 0. \tag{5}$$

Thus, in the log-linear model for (3) the pairwise interactions are either $\lambda_I + \Delta^2\varphi(0)$, or $\Delta^2\varphi(0)$. They are all positive unless the interaction for a set I of size 2 in $\mathscr{I}$ satisfies $\lambda_I < -\Delta^2\varphi(0)$.

3 A log-linear parametrization for the Rasch model and a lower bound for the population size

We reparametrize (3) as follows,

$$\begin{aligned}\log\mu_{\mathbf{x}} = \gamma + 2\varphi(1) - \varphi(2) + \sum_{i=1}^{t} x_i\{\beta_i - \varphi(1) + \varphi(2)\} \\ + \varphi(\sum_{i=1}^{t} x_i) - \varphi(2)(\sum_{i=1}^{t} x_i - 1) + \varphi(1)(\sum_{i=1}^{t} x_i - 2) + \sum_{I \in \mathscr{I}} L(\mathbf{x}, I)\lambda_I.\end{aligned}$$

When $\sum_i x_i$ is equal to 1 or 2, $\varphi(\sum_i x_i) - \varphi(2)(\sum_i x_i - 1) + \varphi(1)(\sum_i x_i - 2) = 0$. Thus the above model can be written in terms of $\gamma_C = \gamma + 2\varphi(1) - \varphi(2)$, $\beta_{Ci} = \beta_i + \varphi(2) - \varphi(1)$, $i = 1, \ldots, t$ and $\tau_m = \varphi(m) - \varphi(2)(m-1) + \varphi(1)(m-2)$, $m = 3, \ldots, t$. This leads to

$$\log\mu_{\mathbf{x}} = \gamma_C + \sum_{i=1}^{t} x_i\beta_{Ci} + \sum_{m=3}^{t} \psi_m(\sum x_i)\tau_m + \sum_{I \in \mathscr{I}} L(\mathbf{x}, I)\lambda_I, \tag{6}$$

where $\psi_m(k)$ is the indicator function of the set $k = m$. Model (6) is a reparametrization of (3). Its parameters are identifiable under some mild conditions on the set $\mathscr{I}$. For instance it cannot contain all the $t(t-1)/2$ pairwise interactions. Observe that $\sum_{I \in \mathscr{I}(2)} L(\mathbf{x}, I) = (\sum x_i)^2/2 - \sum x_i/2$ is a convex function of $\sum x_i$ that is confounded with $\varphi(\sum x_i)$. This shows that the average of the $t(t-1)/2$ pairwise interactions is not estimable in (6). In general all the interaction terms involving a fixed number of lists cannot appear together in $\mathscr{I}$.

Equation (6) highlights that the log-predicted frequency for the missed units γ is not estimable when $F(\alpha)$ is completely unspecified since it cannot be recovered from the parameters of (6). This observation generalizes the findings of Huggins (2001) and Link (2003) to the generalized Rasch model proposed in this work.

In (6), $\{\tau_m : m = 2, \ldots, t\}$ where $\tau_2 = 0$, is a convex, increasing sequence of positive numbers. This is proved by noting that $\tau_3 = \varphi(3) - 2\varphi(2) + \varphi(1) \geq 0$ and $\tau_k - 2\tau_{k-1} + \tau_{k-2} = \varphi(k) - 2\varphi(k-1) + \varphi(k-2) \geq 0$ for $k = 4, \ldots, t$. Now, since $\exp\{\varphi(-k)\}$ is a Laplace transform, $\{\varphi(k) : \ k = 0, \ldots, t\}$ is a convex sequence of numbers such that $\varphi(k) - 2\varphi(k-1) + \varphi(k-2) \geq 0$ for $k = 2, \ldots, t$, see Marshall and Olkin (1979), page 452.

A lower bound for $Np(0)$, the number of units missed, can be derived from (5). Since $\gamma_C = \gamma + 2\varphi(1) - \varphi(2)$

$$Np(0) = \exp \gamma \geq \exp(\gamma_C). \tag{7}$$

This bound is sharp. When there is no heterogeneity $\varphi(k) = 0$, since $\alpha = 0$ with probability 1, and $\gamma = \gamma_C$. For the lower bound to be reached, it suffices that $\varphi(2) = 2\varphi(1)$. This holds with two latent catchability classes when the α value in one class is very large. Most of the individuals in the large α class appear in more than two lists; thus the values of $\varphi(1)$ and $\varphi(2)$ are driven by a single class and satisfy $\varphi(2) = 2\varphi(1)$.

When $\mathscr{I} = \emptyset$ and $\beta_{Ci} = \beta_C$, Chao (1998) derived (7) and showed that the bound has an explicit form, $\exp \gamma_C = (t-1)\mathrm{E}(f_1)^2/\{2t\mathrm{E}(f_2)\}$, where f_i is the number of units captured i times. A tighter bound is derived in Mao (2008) while Rivest and Baillargeon (2007) consider the case where β_{Ci} varies with i and $\mathscr{I} = \emptyset$. When $t = 3$, Table 1 gives explicit expressions for estimators of $\exp(\gamma_C)$ for several sets $\mathscr{I}$.

Given the set of predicted values $\{\mu_{\mathbf{x}} : \mathbf{x} \neq \mathbf{0}\}$, coming from (3), the parameters γ_C, β_{Ci}, τ_m and λ_I of (6) can be evaluated without knowing the distribution $F(\alpha)$ of the catchability variable. The lower bound $\exp(\gamma_C)$ of (7) can easily be calculated. Model (6) is called the lower bound (LB) model for the set of interactions $\mathscr{I}$. Fitting (6) to the sample frequencies $\{n_{\mathbf{x}}\}$, using a Poisson regression, yields a lower bound estimate $n + \exp \hat{\gamma}_C$ for the size of the population, where $n = \sum n_{\mathbf{x}}$

is the number of units appearing in at least one list. When there is no heterogeneity, $n + \exp\hat{\gamma}_C$ is consistent for N when the true model is defined by the interaction set $\mathscr{I}$. This estimator is however not efficient for that model. This is illustrated in Section 5.

4 Data analysis strategy

The data set for analysis contains the sample frequencies $\{n_{\mathbf{x}}\}$, we let $n = \sum n_{\mathbf{x}}$ be the number of units appearing in at least one list. To determine whether a data set has heterogeneous catchability one can fit a log-linear model with all $t(t-1)/2$ pairwise interactions. According to (5), (3) is a likely model when most, if not all, these interactions have positive estimates. To find the set $\mathscr{I}$ of important log-linear interactions, one can use a forward stepwise approach starting with the LB model with $\mathscr{I} = \emptyset$ corresponding to a Poisson regression with explanatory variables $\{x_1, \ldots, x_t, \psi_3(\sum x_i), \ldots, \psi_t(\sum x_i)\}$, see (6). The largest two-way interactions obtained when fitting the pairwise interaction model can be considered first in the stepwise search. Once an LB model is selected, a likelihood ratio test of heterogeneity is carried by testing $H_0 : \tau_3 = \ldots = \tau_t = 0$. If this test is not significant then $\alpha = 0$ is an acceptable value of the latent catchability variable for all the units in the population.

For an LB model to be well defined, the sequence $\{\hat{\tau}_m\}$ has to be positive, convex and increasing. For instance when $m = 5$, it should satisfy $\hat{\tau}_4 \geq 2\hat{\tau}_3 \geq 0$ and $\hat{\tau}_5 \geq 2\tau_4 - \hat{\tau}_3$. When this fails, the model is refitted with some estimates on the boundary of the parameter space. In practice this is best achieved by changing the explanatory variables for the heterogeneity parameters into $\{(\sum x_i - m + 1)_+ : m = 3, \ldots, t\}$, where $x_+ = \max(0, x)$ and by requesting the new log-linear heterogeneity parameters to be non negative. Negative estimates are set to 0 and the model is refitted until all the heterogeneity parameters estimates are non negative. In the next section, the deviance degrees of freedom for an LB model are set to $2^t - 1$ minus the number of parameter estimates that are not on the boundary of the parameter space.

Since the model is fitted using a Poisson regression, we get the so-called conditional estimator of N, as opposed to the unconditional estimator that would have been obtained by maximizing the corresponding multinomial likelihood. The estimator of N and its variance estimate are given by

$$\hat{N} = n + \exp\hat{\gamma}_C \quad v(\hat{N}) = \exp\hat{\gamma}_C + \exp(2\hat{\gamma}_C)v_p(\hat{\gamma}_C), \tag{8}$$

where $v_p(\gamma_C)$ is the variance of intercept in the Poisson regression for the LB model, see Rivest and Lévesque (2001) for a derivation of $v(\hat{N})$. The R package `Rcapture`,

see Baillargeon and Rivest (2007), provides a friendly environment to carry out these calculations.

To supplement the lower bound estimate obtained from (6), it is interesting to estimate N under several specifications for the catchability distribution $F(\alpha)$. Three types of models that (3) associated with parametric families of distributions $F(\alpha)$ are considered. We use $F(\alpha) = \Phi(\sqrt{2}\alpha/\sigma)$, where $\Phi(\cdot)$ is the $N(0,1)$ distribution function and σ is an unknown scale parameter measuring the variability of the latent catchability. A model featuring two latent classes, as presented in Section 2, is also discussed. Some data dependent mixing distributions leading to marginal log-linear models for the observed data are also considered. These models are briefly reviewed.

4.1 Normal mixing distribution

Models with a normal mixing distribution are fitted as in Coull and Agresti (1999). A 20 point Gaussian quadrature formula is used to approximate $\varphi(k)$,

$$\varphi(k) = \log\left\{\sum_{i=1}^{20} w_i \exp(k\sigma z_i) C(\sigma z_i) / \sum_{i=1}^{20} w_i C(\sigma z_i)\right\} \quad k = 1,\ldots,t,$$

where the weights $\{w_i\}$ and the abscissa $\{z_i\}$ are taken from Abramowitz and Stegun (1972), page 924. The predicted values (3) are then evaluated and the Poisson deviance measuring the discrepancy between the observed and the predicted values is minimized using `optim`, a general optimization routine in `R`. An estimator for N and of its variance are calculated as in (8), with $\hat{\gamma}_C$ replaced by the $\hat{\gamma}$, the estimate of the intercept in (3), and $v_p(\hat{\gamma})$ set equal to the (1,1) entry of the inverse of the Hessian matrix calculated at the minimum value of the deviance.

4.2 Latent class models

The dichotomous latent class (LC) model introduced in Section 2 has two additional parameters: α, the heterogeneity parameter for latent class 1 (it is assumed to be 0 in latent class 0), and a parameter γ_U for the relative size of the two classes. Following Stanghelli and van der Heijden (2004) we use the EM algorithm to fit this model. The unobserved complete data vector has entries $y_{\mathbf{x},u}$, the frequency of capture history $\mathbf{x}$ in class u, $u = 0,1$. The model for this dependent variable is log-linear; its parameters are the log-linear parameters in (3) plus γ_U and α. The columns of the design matrix for these parameters are $(u, u \times \sum x_i)$. The M step fits a Poisson regression to the dependent vector $\mathbf{y}$ while the E step estimates the dependent vector using $y_{\mathbf{x},u} = n_{\mathbf{x}} \times \hat{y}_{\mathbf{x},u}/(\hat{y}_{\mathbf{x},0} + \hat{y}_{\mathbf{x},1})$, and $\hat{y}_{\mathbf{x},u}$ stands for the

predicted value for $(\mathbf{x}, u)$ obtained at the previous M step. The conditional estimator for N is $n + \exp(\hat{\gamma}) + \exp(\hat{\gamma} + \hat{\gamma}_U)$ and the variance is calculated as in Stanghelli and van der Heijden (2004).

We experienced numerical difficulties when fitting the LC model. The EM algorithm did not converge to a finite estimate for N when the estimators $\{\hat{\tau}_m : m = 3, \ldots, t\}$ of the heterogeneity parameters of the LB model were such that $\hat{\tau}_3 > 0$ and $\hat{\tau}_m = (m-2)\hat{\tau}_3$ for $m = 4, \ldots, t$. The remainder of this section demonstrates formally the non existence of the Poisson maximum likelihood estimator of N for the LC model in this case.

The heterogeneity parameters obtained in an LB reparametrization of the dichotomous LC model cannot satisfy $\tau_3 > 0$ and $\tau_m = (m-2)\tau_3$ for $m = 4, \ldots, t$. Such heterogeneity parameters can only be obtained as the limit of a sequence of LC models applied to a sequence of populations indexed by a constructed as follows:

- The population size is $N_a = N_0 \exp a$, where N_0 is a positive integer;
- The list effects are $\beta_{ja} = \beta_j - a$ for some fixed $\beta_j,\ j = 1, \ldots, t$;
- The interaction parameters in $\mathscr{I}$ do not depend on a;
- The additional parameters for the LC model are $\alpha_a = a$ and $\gamma_{Ua} = a_0 - a$, where a_0 is a constant.

The limits, as a goes to ∞, of the heterogeneity parameters for this sequence of LC models satisfy $\tau_m = (m-2)\tau_3 > 0$ for $m = 4, \ldots, t$. This result is proved in the appendix. The above construction gives a sequence of populations where the size of latent class 0 diverges to ∞ while that of latent class 1 stays bounded. The probability of being missed in latent class 0, $C(0)$ goes to 1, and all the predicted values (3) have finite limits. When the heterogeneity parameters of the LB model satisfy $\hat{\tau}_3 > 0$ and $\hat{\tau}_m = (m-2)\hat{\tau}_3$ for $m = 4, \ldots, t$, the best fitting LC model is the limit, as a goes to ∞, of such a sequence despite its degenerate nature. Thus any iterative algorithm that fits the LC model gives a diverging sequence of estimates for N and the Poisson maximum likelihood estimator of N does not exist.

4.3 Data dependent mixing distributions

Following Lindsay (1986) and Darroch, Fienberg, Glonek, and Junker (1993), this section assumes that α has a density proportional to $C(\alpha)^{-1} dF_\tau(\alpha)$, where $C(\alpha)$ is defined in (1) and $F_\tau(\alpha)$ is a standard infinitely divisible distribution with shape parameter τ such as the Poisson, the normal or the gamma distribution. The mixture density $C(\alpha)^{-1} dF_\tau(\alpha)$ is typically a mixture of $F_\tau(\alpha)$ that depends on the number

of capture occasions t and on the log-linear parameters of (3). It is very unlikely to be the true mixing density. Still, it provides a simple class of log-linear models that can be fitted easily without encountering numerical problems such as those highlighted in Section 4.2.

With such a data-dependent mixing density, (3) can be expressed in terms of the logarithm of the moment generating function, $\tau\psi(k)$, of $F_\tau(\alpha)$. Suppose, for instance, that $F_\tau(\alpha)$ is the distribution of X, a Poisson random variable with parameter τ, multiplied by $\log a$ where a is a positive number. The moment generating function of $F_\tau(\alpha)$ is

$$E[\exp\{w\log(a)X\}] = \exp\{\tau(a^w - 1)\}.$$

This mixing distribution has $\varphi(k) = \tau(a^k - 1)$ and gives the following marginal model,

$$\log\mu_\omega = \gamma + \sum_j \omega_i\beta_i + \tau\psi(\sum_i \omega_i) + \sum_{I\in\mathscr{I}} L(\mathbf{x},I)\lambda_I, \tag{9}$$

where $\psi(k) = a^k - 1$. In the sequel we use the Poisson model with $a = 2$; this corresponds to a mixing density that takes only positive values. In (4) $\Delta^m\varphi(0) = \tau$ for $m = 1,\ldots,t$; this model accounts for the heterogeneity by adding τ to all the log-linear interactions.

When $F_\tau(\alpha)$ is the normal distribution with mean 0 and variance τ, the logarithm of its moment generating function is $\tau w^2/2$. Assuming that the density of α is proportional to $C(\alpha)^{-1}dF_\tau(\alpha)$ leads to model (9), with $\psi(k) = k^2/2$. This model is considered by Darroch et al. (1993) and Coull and Agresti (1999). This Darroch model has $\Delta^2\varphi(0) = \tau/2$ while $\Delta^m\varphi(0) = 0$ for $m > 2$. It expresses the heterogeneity only through the pairwise log-linear interactions. Treating the heterogeneity with a simple search for significant log-linear interactions will often produce an estimate $\hat{N}$ very similar to Darroch's. This strategy implicitly assumes a latent heterogeneity with a mixed normal distribution. The Darroch and the normal model of Section 4.1 often give similar estimates for N since their mixing distributions are both derived from the normal. These estimates can be much larger than those obtained with the LB or the Poisson2 model whose mixing density is bounded from below. The cumulative distribution function $F_\tau(\alpha)$ of the negative of a gamma variable with shape parameter τ and scale parameter $1/u$ can also be used. It leads to (9), with $\psi(k) = \log a - \log(a+k)$; the value $a = 3.5$ is used in the sequel.

Rivest and Baillargeon (2007) studied the models proposed here to deal with a heterogeneous catchability when $\mathscr{I} = \emptyset$. They give mixing distributions with different characteristics that allow carrying out a sensitivity analysis of the way in which the heterogeneity is modeled. Once a set of interactions $\mathscr{I}$ is selected, a

model for heterogeneity is nested within the LB model. It expresses the parameter $\tau_3, \ldots, \tau_t$ of (6) in terms of the distribution $F(\alpha)$ of the latent catchability. Thus, its fit can be ascertained by comparing its deviance to that of the LB model. This highlights that only $t-2$ degrees of freedom are available for selecting the distribution of the latent variable α.

5 Models for t = 3 lists

When $t = 3$, only one degree of freedom is available to estimate the mixing distribution $F(\alpha)$; the LC model cannot be fitted and all the heterogeneity models of Sections 4.1 and 4.3 have the same deviance. Except when $\mathscr{I} = \emptyset$, the LB estimates for N have a closed form that is given in Table 1. The heterogeneity parameter is estimated by solving the equation

$$\log n_{111} - \hat{\gamma}_C - \sum_{i=1}^{3} x_i \hat{\beta}_{Ci} - \sum_{I \in \mathscr{I}} L(\mathbf{x}, I) \hat{\lambda}_I = \varphi(3) - 2\varphi(2) + \varphi(1).$$

For the estimators of Section 4.3, $\varphi(k) = \tau\psi(k)$, where τ is the heterogeneity parameter. Solving this equation gives a closed form estimate $\hat{\tau}$ for τ. This leads to the expressions for $\exp\hat{\gamma} = \widehat{Np(0)}$ given in Table 1; see the appendix for detailed calculations. The estimators depend on an exponent a that varies with the method for handling the heterogeneity. The largest one is $a = 1.51$ for the Gamma3.5 model.

In Table 1, $\hat{\gamma}_C$ and $\hat{\beta}_{Ci}$ refer to the estimators for the model with $\mathscr{I} = \emptyset$; they have no closed form expressions. The estimators $\exp\hat{\gamma}$ for the four heterogeneity specifications have the same form, namely $\exp\hat{\gamma}_C$, the lower bound estimator, times a statistic for the heterogeneity, S^a. When there is no heterogeneity, the predicted value of S is 1, and all the estimators $n + \exp\hat{\gamma}$ are approximately unbiased for N. When heterogeneity is present, the predicted value of S is larger than 1 and $\exp\hat{\gamma} \geq \exp\hat{\gamma}_C$. The magnitude of the difference increases with the exponent a. A large value of a gives a large correction for heterogeneity. Among all the models considered in Section 4.3, the Gamma3.5 model with $a = 1.51$ gives the largest estimator for N.

When $\mathscr{I} = \{(1,2)\}$ the estimator of the number of units missed in the model without heterogeneity is $(n_{100} + n_{010} + n_{110}) \times n_{001}/(n_{101} + n_{011} + n_{111})$. When the catchability is heterogeneous, (6) implies that n_{111} can be arbitrarily large; then the above estimator has a negative bias. The LB estimator takes n_{110} and n_{111} out of the above formula. This gives a consistent estimator for the $\mathscr{I} = \{(1,2)\}$ model that is less sensitive to heterogeneity than the maximum likelihood estimator. Darroch et al. (1993) considered models with $\varphi(k) = \tau k^2/2$ and $\mathscr{I} = \{(1,2)\}$

Table 1: Estimators of the number of units missed, $\exp\hat{\gamma}$, when $t = 3$ for various models and their deviance degrees of freedom, df. The values of the exponent a are 0, 0.5, 1.0, and 1.51 for the LB, the Poisson2, the Darroch and the Gamma3.5 model respectively.

Model	df	$\exp\hat{\gamma}$
$\beta_{Ci} = \beta_C$, $\mathscr{I} = \emptyset$	4	$\frac{3^{a-1}(n_{100}+n_{010}+n_{001})^{a+2}n_{111}^{a}}{(n_{110}+n_{011}+n_{011})^{2a+1}}$
$\mathscr{I} = \emptyset$	2	$\exp\hat{\gamma}_C \left\{ \frac{n_{111}}{\exp(\hat{\gamma}_C + \sum \hat{\beta}_{Ci})} \right\}^a$
$\mathscr{I} = \{(1,2)\}$	1	$n_{001}^{a+1} \left(\frac{n_{100}+n_{010}}{n_{101}+n_{011}}\right)^{a+1} \left(\frac{n_{111}}{n_{110}}\right)^a$
$\mathscr{I} = \{(1,2),(1,3)\}$	0	$\frac{n_{010} \times n_{001}}{n_{011}} \left(\frac{n_{111} \times n_{100}}{n_{110} \times n_{101}}\right)^a$

Table 2: Data from the 1988 Dress-Rehearsal Census.

				x			
Source	(1,1,1)	(1,1,0)	(1,0,1)	(1,0,0)	(0,1,1)	(0,1,0)	(0,0,1)
$R3$	72	155	7	71	13	53	43

in their Section 4.4. They noted that their predicted values satisfy the constraint $\mu_{011}\mu_{100}/(\mu_{101}\mu_{010}) = 1$.

6 Numerical examples

First consider the $R3$ data, for strata 11, 12, and 13 in Table 2 of Darroch et al. (1993), about a population of black males in a geographic area which is presented in Table 2 here. This data is from the 1988 dress-rehearsal census; list 1 is the census, list 2 is the Post-Enumeration survey to estimate the census undercount and list 3 is an administrative list compiled from federal agencies.

The saturated log-linear model with $\mathscr{I} = \{(1,2),(1,3),(2,3)\}$ gives $\hat{N} = 1240$ $s.e. = 450$; the presence of heterogeneity is likely since the estimates of the three pairwise interaction terms are positive. Removing the $[E_1E_3]$ interaction, which is not significant at the 5% level, leads to $\hat{N} = 850$ $s.e. = 186$ and a deviance of 3.8 for one degree of freedom. This is the best fitting hierarchical log-linear

model; it is built on the premise that given the status, either counted or not at the post-enumeration survey, the census and the administrative list are independent.

A heterogeneous catchability is plausible in this example, since the three lists are associated to the federal government. The best fitting LB model has $\mathscr{I} = \{(1,2)\}$ with a deviance of 3.45 for df=1 and $\hat{N} = 681$ $s.e. = 78$. The model for the heterogeneous catchability has a huge impact on the estimate. For the normal model of section 4.1, $\hat{N} = 1262$ $s.e. = 346$ while the Darroch et al. (1993) model leads to $\hat{N} = 1400$ $s.e. = 559$. This data comes from the aggregation of three sampling strata; to pursue the analysis, models were fitted to one single stratum. Sizeable differences were still present between the LB and the Darroch et al. (1993) estimates for N. Thus a lower bound estimate might be the best one can do to handle a heterogeneous catchability in this example.

6.1 Analysis of the diabetes data of Bruno et al. (1994)

To illustrate the application of the models of Section 4 consider the $t = 4$ diabetes data of Bruno et al. (1994), which has been analyzed repeatedly in the literature. For completeness, this marginal data is reported in Table 3. The four lists are $E_1 =$ Clinic records, $E_2 =$ Hospital discharge, $E_3 =$ Prescrition and $E_4 =$ Syringe usage. It has $n = 2069$ patients appearing on at least one list, with 47% (975/2069) appearing on one list only. The estimates reported in the literature vary between $\hat{N} \approx 2600$, see Bruno et al. (1994) and Chao et al. (2001), and $\hat{N} \approx 4000$ in IWGDMF (International Working Group for Disease Monitoring and Forecasting) (1995a). Table 4 reports estimates calculated with the models proposed in this paper.

A heterogeneity in catchability is likely since the model with pairwise log-linear interactions has interaction estimates that vary between 0.15 to 1.94. The stepwise search for important list interactions suggested in Section 4 leads to the the interaction set $\mathscr{I} = \{(1,3),(2,4),(3,4)\}$ for the LB model. This model did not meet the constraints $\hat{\tau}_4 > 2\hat{\tau}_3$; thus the final LB model has $\hat{\tau}_4 = 2\hat{\tau}_3$ and $\hat{N} = 2588$, $s.e. = 75$; its deviance is 3.13 for 6 degrees of freedom. The hierarchical log-linear model with $\mathscr{I} = \{(1,3),(2,4),(3,4)\}$ has a deviance of 21.97 for 7 degrees of freedom and the χ^2_1 test statistic for a heterogeneous catchability is 18.84. Thus the null hypothesis of the absence of heterogeneity is rejected at the 0.001 level.

The heterogeneity parameters for the LB model satisfy $\hat{\tau}_4 = 2\hat{\tau}_3$. The iterations of the EM algorithm to fit the LC model of Section 4.2 gave diverging sequences for N, β_j, α, and γ_U, with characteristics similar to those of the parameters of the sequence of models indexed by a in Section 4.2. When $t = 4$, the estimate for $\hat{\tau}_4$ is driven by n_{1111}. It is equal to $2\hat{\tau}_3$ when n_{1111} is small. For instance, setting n_{1111} to 88 rather than 58 in Table 3 yields an LB model $\hat{\tau}_4 > 2\hat{\tau}_3$ and a finite LC

Table 3: Diabetes data set, unstratified and stratified by treatment

		Total				Diet			
		E_1=1		E_1=0		E_1=1		E_1=0	
		E_2=1	E_2=0	E_2=1	E_2=0	E_2=1	E_2=0	E_2=1	E_2=0
E_3=1	E_4=1	58	46	14	8	0	1	1	1
E_3=1	E_4=0	157	650	20	182	1	11	4	11
E_3=0	E_4=1	18	12	7	10	1	0	0	0
E_3=0	E_4=0	104	709	74	NA	11	150	13	NA
		Hypoglycemic agents				Insulin			
		E_1=1		E_1=0		E_1=1		E_1=0	
		E_2=1	E_2=0	E_2=1	E_2=0	E_2=1	E_2=0	E_2=1	E_2=0
E_3=1	E_4=1	6	9	3	3	51	36	10	4
E_3=1	E_4=0	98	571	13	163	57	62	3	6
E_3=0	E_4=1	3	5	1	4	14	7	7	6
E_3=0	E_4=0	68	509	58	NA	22	42	2	NA

Poisson estimate for N, $\hat{N} = 3210$. This highlights that the LC model might not be applicable with $t = 4$ lists. As the number of lists increases, it is less likely to get $\hat{\tau}_m = (m-2)\hat{\tau}_3$ with $\hat{\tau}_3 > 0$, and the numerical difficulties associated to the LC model should be less severe. Still, in experiments with $t = 6$ capture occasions, Dorazio and Royle noticed a frequent lack of convergence of the algorithm for the LC maximum likelihood estimator of N in their response to Pledger (2005).

Table 4 presents the results obtained with several models. Because the LB heterogeneity parameters satisfy $\hat{\tau}_4 = 2\hat{\tau}_3$, the best fitting models are those that provide a large correction for heterogeneity. In Table 4, the best fit and the largest $\hat{N}$ are obtained with the Gamma0.5 model. When $t = 3$ this model has an a-value—as defined in Table 1—of 3.37, as compared to 1.5 and 1 for the Gamma3.5 and the Darroch model.

A stepwise search for a hierarchical log-linear model led Bruno et al. (1994) to a model with $\mathscr{I} = \{(1,2),(1,3),(2,3),(2,4),(3,4)\}$. It has $\hat{N} = 2771$ $s.e. = 145$ and a deviance of 7.62 for 5 degrees of freedom. Modeling the heterogeneity directly with a normal distribution or Darroch et al. (1993) model yields similar values of $\hat{N}$, a more parsimonious model, and a smaller standard error for $\hat{N}$. Table 4 suggests however that modeling the heterogeneity with a long tailed distribution as the negative of a gamma random variable would have provided a better fit and larger estimates for N.

Table 4: Models fitted to the diabetes data.

	Total				Diet			
Interaction set $\mathscr{I}$	$\{(1,3),(2,4),(3,4)\}$				$\{(1,2),(1,3),(3,4)\}$			
Model	Dev	df	$\hat{N}$	$s.e.$	Dev	df	$\hat{N}$	$s.e.$
LB	3.13	6	2588	75	5.67	6	236	19
Poisson2	12.88	6	2573	76	6.16	6	241	24
Darroch	8.32	6	2752	133	5.94	6	256	35
Normal	8.22	6	2763	101	5.94	6	256	34
Gamma3.5	6.74	6	2964	218	5.87	6	275	84
Gamma0.5	5.49	6	4238	952				
None	21.97	7	2472	54	6.49	7	234	18
	Hypo				Insulin			
Interaction set $\mathscr{I}$	$\{(1,3),(2,4),(3,4)\}$				$\{(1,4)\}$			
Model	Dev	df	$\hat{N}$	$s.e.$	Dev	df	$\hat{N}$	$s.e.$
LB	5.67	6	1976	81	8.24	7	332	3
Poisson2	7.13	6	2002	99	11.15	8	332	2
Darroch	6.35	6	2103	157	9.10	8	334	4
Normal	6.35	6	2104	114	9.14	8	335	3
Gamma3.5	6.13	6	2208	232	8.51	8	339	6
None	9.80	7	1910	62	26.48	9	330	2

The analysis now focusses on the $n = 2047$ patients that have been stratified according to their diabetes treatment, either Diet, Hypoglycemic agents, or Insulin. The stratum data given in Bruno et al. (1994) is provided in Table 3. Carrying out independent analyses in each stratum should reduce the unobserved heterogeneity in catchability since the proportion of patients seen only once varies substantially between strata. It is $174/205 = 85\%$, $734/1514 = 48\%$, and $56/328 = 17\%$ respectively in the three strata. Separate analyses, stratum by stratum, are reported in Table 4. Tests for heterogeneity were carried within the three strata as proposed in Section 2. The null hypothesis that $\tau_3 = \tau_4 = 0$ in the LB model found after a stepwise selection of important interactions is not rejected, at the 5% level, in the Diet stratum.

The algorithm for fitting the LC model did not converge in the Diet and Hypoglycemic agent strata, as the heterogeneity parameters of the LB model were on the boundary of the parameter space for these two strata. In the insulin stratum the LC model gives $\hat{N} = 335$, which is similar to the other estimates reported in Table 4 for that stratum.

The results reported in Table 4 rule out a combined estimator as large as 4000 that was possible with the unconditional analysis of Table 3. Combining the results for the three stratum of Table 4 yields $\hat{N} = 2544$ $s.e. = 83$ and a deviance of 19.58 with 19 degrees of freedom for the LB model and $\hat{N} = 2695$ $s.e. = 119$ and a deviance of 21.43 with 20 degrees of freedom for the normal mixture model. This normal mixture estimate is our definitive proposal for this data set. A similar estimate, with a slightly larger standard error, could be obtained by fitting Darroch et al. (1993) log-linear model with $\varphi(k) = \tau k^2/2$.

Stanghelli and van der Heijden (2004) do not report a lack of convergence for $\hat{N}$ when fitting the LC model simultaneously to the three strata. Their model differs from the ones considered here as their parameters α depends only on the list. They do not have stratum specific α value that could be obtained with log-linear models involving three way interactions between the latent class, the list, and the stratum. Their selected model, called model 13, has $\hat{N} = 2582$ and a deviance of 30.7 for 23 degrees of freedom. The stratum specific analysis proposed in this work allows reduction of the deviance by 30% with only three additional parameters.

7 Conclusion

This paper introduced the lower bound (LB) log-linear model for multiple record system estimation of N with data exhibiting a heterogeneity in catchability and list interactions. This model can be used to carry out a stepwise search of important list interactions and for testing whether the catchability is heterogeneous. All these

log-linear models can be fitted using the function `closedpCI.t` of the package `Rcapture`.

The heterogeneity parameters of the LB model contain diagnostic information. The configuration $\hat{\tau}_3 > 0$ and $\hat{\tau}_m = (m-2)\hat{\tau}_3$ $m = 4, \ldots, t$ is associated to an extreme form of heterogeneity that cannot be modeled with a dichotomous latent class model.

Once a set of interactions $\mathscr{I}$ has been selected several strategies have been presented to estimate N. One can report a lower bound estimator. The heterogenous catchability can be modeled with either a latent class model or a normal distribution. Simple log-linear models associated with data dependent mixing distributions can also be used. Thus one can investigate how sensitive is $\hat{N}$ to the specification of a model for the heterogeneous catchability.

Appendix 1: Proof of equation (4)

One has $\Delta^2\varphi(k) = \varphi(k+1) - 2\varphi(k+1) + \varphi(k)$ and for any positive integer m

$$\Delta^m\varphi(k) = \sum_{i=0}^{m} \binom{m}{i} (-1)^i \varphi(k+i).$$

Formally the operator Δ can be defined as $E - I$ where $E\ \varphi(k) = \varphi(k+1)$ is the forward operator and I is the identity operator. Thus $E = I + \Delta$ and

$$E^k = (I+\Delta)^k = \sum_{m=0}^{k} \binom{k}{m} \Delta^m$$

Since $\varphi(k) = E^k\varphi(0)$ the above equation yields

$$\varphi(k) = \sum_{m=0}^{k} \binom{k}{m} \Delta^m\varphi(0) \quad k = 1, \ldots, t.$$

If $k = \sum x_i$, then $\binom{k}{m}$ is the number of possible m-term interactions. In other words

$$\binom{k}{m} = \sum_{I \in \mathscr{I}(m)} L(\mathbf{x}, I).$$

When $m > k$, the two sides of this equality are equal to 0 since all the products $L(\mathbf{x}, I)$ are equal to 0 when the subset I contains more than k elements. In general one can write

$$\varphi(\sum x_i) = \sum_{m=0}^{t} \sum_{I \in \mathscr{I}(m)} L(\mathbf{x}, I) \Delta^m\varphi(0).$$

Appendix 2: Derivation of the LC model for which $\tau_3 > 0$ and $\tau_4 = 2\tau_3$

Without losing generality we take $\alpha > 0$. The assumption $\tau_3 > 0$ and $\tau_4 = 2\tau_3$ is equivalent to $\exp\{\varphi(3) - 2\varphi(2) + \varphi(1)\} > 1$ and $\exp\{\varphi(4) - 2\varphi(3) + \varphi(2)\} = 1$. Written in terms of the parameters γ_U and α these two equations become

$$\begin{aligned}
\frac{C(0)^2 + \exp(\gamma_U)C(0)C(\alpha)\{\exp(3\alpha) + \exp(\alpha)\} + C(\alpha)^2\exp(2\gamma_U + 4\alpha)}{C(0)^2 + 2\exp(\gamma_U)C(0)C(\alpha)\exp(2\alpha) + C(\alpha)^2\exp(2\gamma_U + 4\alpha)} &> 1 \\
\frac{C(0)^2 + \exp(\gamma_U)C(0)C(\alpha)\{\exp(4\alpha) + \exp(2\alpha)\} + C(\alpha)^2\exp(2\gamma_U + 6\alpha)}{C(0)^2 + 2\exp(\gamma_U)C(0)C(\alpha)\exp(3\alpha) + C(\alpha)^2\exp(2\gamma_U + 6\alpha)} &= 1,
\end{aligned}$$

where $C(0)$ and $C(\alpha)$ are the probabilities of being missed in the two latent classes. Both need to be non zero to meet these two conditions. The second equality cannot be met for finite values of γ_U and α. It can be true at the limit when α goes to ∞. For the second limit to be 1 one needs $2\alpha + \gamma_U$ to become arbitrarily large while for the first limit to be larger than 1, $\alpha + \gamma_U$ has to converge to a finite value as α goes to ∞. The two conditions are met for the parameters α_a and γ_{Ua} presented in section 4.2. The parameters $\beta_{ja} : \ j = 1, \ldots, t$ of Section 4.2 are needed for the limiting values of τ_m to be non zero.

With the parameters α_a, γ_{Ua}, and β_{ja}, $C(0) = 1 - \sum \exp(\beta_j - a) + o\{\exp(-a)\}$ while $C(\alpha_a)$ does not depend on a. Thus $\varphi(k) = \log[\{C(0) + C(\alpha_a)\exp(\gamma_{Ua} + k\alpha_a)\}/\{C(0) + C(\alpha_a)\exp(\gamma_{Ua})\}]$ converges to $\log\{1 + C(\alpha_a)\exp(a_0)\}$ when $k = 1$ while it is equivalent to $\log\{C(\alpha_a)\} + a_0 + a(k-1) + o(1)$ when $k > 1$. The limiting values of the heterogeneity parameters for the LB model are

$$\tau_m = (m-2)[\log\{1 + C(\alpha_a)\exp(a_0)\} - a_0 - \log\{C(\alpha_a)\}] \quad m = 3, \ldots, t.$$

Finally taking $N_a = N_0 \exp a$ ensures that the log-predicted values (3) have finite limits as a goes to ∞.

Observe also that the limiting values of the parameters for the LB model for the sequence of populations of Section 4.2 are finite. They are given by the above values of τ_m, the log-linear interactions in $\mathscr{I}$, the marginal list effects $\beta_{Ci} = \beta_i - \log\{1 + C(\alpha_a)\exp(a_0)\} + a_0 + \log\{C(\alpha_a)\}$ and the intercept $\gamma_C = \log(N_0) + 2\log\{1 + C(\alpha_a)\exp(a_0)\} - a_0 - \log\{C(\alpha_a)\}$.

Appendix 3: Derivation of the estimates $\exp\hat{\gamma}$ when $t = 3$

When $\mathscr{I} = \emptyset$ and $\beta_{Ci} = \beta_C$, the LB model has three parameters γ_C, β_C and τ_3. The sufficient statistics for this model are (f_1, f_2, f_3) where f_i is the number of units caught i times. The model is saturated and the parameter estimates are obtained by

solving the equations setting f_i equal to their predicted values. This leads to

$$\exp\hat{\gamma}_C = \frac{f_1^2}{3f_2} \quad \hat{\beta}_C = \log(f_2/f_1) \quad \hat{\tau}_3 = \log(3f_1f_3) - 2\log f_2.$$

For any of the heterogeneity model of Section 4.3, $\hat{\tau} = \hat{\tau}_3/\{\psi(3) - 2\psi(2) + \psi(1)\}$ thus

$$\hat{\gamma} = \hat{\gamma}_C + \hat{\tau}_3\ \{\psi(2) - 2\psi(1)\}/\{\psi(3 - 2\psi(2) + \psi(1)\} \tag{10}$$

The multipliers $a = \{\psi(2) - 2\psi(1)\}/\{\psi(3) - 2\psi(2) + \psi(1)\}$ of τ_3 for the LB, the Poisson2, the Darroch and the Gamma3.5 model are equal to 0, 0.5, 1, and 1.51 respectively. The estimates reported for M_h in Table 2 are derived from (10). Under $[E_1]$ $[E_2]$ $[E_3]$, the Poisson regression estimates $\hat{\gamma}_C$, $\hat{\beta}_{Ci}$, $i = 1,2,3$ of the lower bound model (6) do not have simple closed forms. Once they are estimated, $\hat{\tau}_3 = f_3/\exp(\hat{\gamma}_C + \sum\hat{\beta}_{Ci})$. The values of $\hat{\gamma}$ are then derived from (10).

When $\mathscr{I} = \{(1,2)\}$, the Poisson regression estimating equations for the parameters of the LB model are

$$\begin{aligned}
n_{111} &= \exp(\gamma_C + \sum\beta_{Ci} + \lambda_{12} + \tau_3)\\
n_{110} &= \exp(\gamma_C + \beta_{C1} + \beta_{C2} + \lambda_{12})\\
n_{001} &= \exp(\gamma_C + \beta_{C3})\\
n_{100} + n_{101} &= \exp(\gamma_C + \beta_{C1})(1 + \exp\beta_{C3})\\
n_{101} + n_{011} &= \exp\beta_{C3}\{\exp(\gamma_C + \beta_{C1}) + \exp(\gamma_C + \beta_{C2})\}\\
n_{100} + n_{010} &= \exp(\gamma_C + \beta_{C1}) + \exp(\gamma_C + \beta_{C2})
\end{aligned}$$

Solving these equations leads to the following

$$\exp(\hat{\gamma}_C) = n_{001}\frac{n_{100} + n_{010}}{n_{101} + n_{011}} \quad \exp(\hat{\tau}_3) = n_{001}\frac{n_{111}(n_{100} + n_{010})}{n_{110}(n_{101} + n_{011})},$$

and the estimates of Table 2 for this model are derived from (10).

The LB $[E_1E_2]$ $[E_1E_3]$ model is saturated and the estimates of its parameters are

$$\exp(\hat{\gamma}_C) = \frac{n_{010} \times n_{001}}{n_{011}} \quad \exp(\hat{\tau}_3) = \frac{n_{111} \times n_{100}}{n_{110} \times n_{101}}.$$

References

Aaron, D. J., Y.-F. Chang, N. Markovic, and R. E. LaPorte (2003): "Estimating the lesbian population: a capture-recapture approach," *Journal of Epidemiology Community Health*, 57, 207–209.

Abramowitz, M. and I. Stegun (1972): *Handbook of Mathematical Functions. Tenth Edition*, New York: Dover Press.

Baillargeon, S. and L.-P. Rivest (2007): "The Rcapture package: Loglinear models for capture-recapture in R," *Journal of Statistical Software*, 19, URL `http://www.jstatsoft.org/`.

Bartolucci, F. and A. Forcina (2001): "Analysis of capture-recapture data with a Rasch-type model allowing for conditional dependence and multidimensionality," *Biometrics*, 57, 714–719.

Bartolucci, F. and A. Forcina (2006): "A class of latent marginal models for capture recapture data with continuous covariates," *Journal of the American Statistical Association*, 101, 789–794.

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland (1975): *Discrete Multivariate Analyses: Theory and Practice*, MIT Press.

Bruno, G., A. Biggeri, R. E. LaPorte, D. McCarty, F. Merletti, and G. Pagano (1994): "Appplication of capture-recapture to count diabetes?" *Diabetes Care*, 17, 548–556.

Chao, A. (1998): "Estimating animal adundance with capture frequency data," *J. Wildl. Manage.*, 52, 295–300.

Chao, A., P. Tsay, S.-H. Lin, W.-Y. Shau, and D.-Y. Chao (2001): "The applications of capture-recapture models to epidemiological data," *Statistics in Medicine*, 20, 3123–3157.

Cormack, R. M., Y.-F. Chang, and G. S. Smith (2000): "Estimating death from industrial injury by capture-recapture: a cautionary tale," *Journal of Epidemiology*, 29, 1053–1059.

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

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

Hook, E. B. and R. R. Regal (1993): "Effect of variation in probability of ascertainment by sources ("variable catchability") upon "capture-recapture" estimates of prevalence," *American Journal of Epidemiology*, 137, 1148–1166.

Huggins, R. (2001): "A note on the difficulties associated with the analysis of capture-recapture experiments with heterogeneous capture probabilities," *Statist. Probab. Lett.*, 54, 147–152.

IWGDMF (International Working Group for Disease Monitoring and Forecasting) (1995a): "Capture-recapture and multiple record systems estimation i: history and theoretical development," *Am. J. Epidemiol.*, 142, 1047–1058.

IWGDMF (International Working Group for Disease Monitoring and Forecasting) (1995b): "Capture-recapture and multiple record systems estimation ii: Applications in human diseases," *Am. J. Epidemiol.*, 142, 1059–1068.

Lindsay, B. G. (1986): "Exponential family mixture models (With least-squares estimators)," *Ann. Statist.*, 14, 124–137.

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

Link, W. A. (2006): "Reply to a paper by Holzmann, Munk, and Zucchini," *Biometrics*, 62, 936–939.

Mao, C. X. (2008): "Lower bounds to the population size when capture probabilities vary over individuals," *Australian and New Zeland Journal of Statistics*, 50, 125–135.

Marshall, A. and I. Olkin (1979): *Inequalities: Theory of Majorization and Its Applications*, New York: Academic Press.

Murphy, J. (2009): "Estimating the world trade center tower population on september 11, 2001: A capture-recapture approach," *American Journal of Public Health*, 99, 65–67.

Pledger, S. (2005): "The performance of mixture models in heterogeneous closed population capture-recapture," *Biometrics*, 61, 868–873.

Rivest, L.-P. and S. Baillargeon (2007): "Applications and extensions of Chao's moment estimator for the size of a closed population," *Biometrics*, 63, 999–1006.

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

Stanghelli, E. and P. G. M. van der Heijden (2004): "A multiple-record systems estimation method that takes observed and unobserved heterogeneity into account," *Biometrics*, 60, 510–516.