Chapter 2

ANSWERING TWO BIOLOGICAL QUESTIONS WITH A LATENT CLASS MODEL VIA MCMC APPLIED TO CAPTURE-RECAPTURE DATA

Francesco Bartolucci,[1] Antonietta Mira[2] and Luisa Scaccia[3]

Istituto di Scienze Economiche, University of Urbino, Italy[1]; Dipartimento di Economia, University of Insubria, Italy[2]; Dipartimento di Scienze Statistiche, University of Perugia, Italy[3]

1. INTRODUCTION

A well-known method for estimating the size, N, of a certain population is the *capture-recapture* method (for a review see Yip *et al.*, 1995a and Schwarz and Seber, 1999). The first motivations to the development of these methods arose in biology where researchers were interested in estimating the number of animals of a certain species (see, for instance, Schnabel, 1938, and Darroch, 1958). Subsequently, this methodology was also applied in medical and social contexts where it is important to estimate the number of subjects with a certain disease or in a particular situation (Yip *et al.*, 1995b).

A typical capture-recapture study consists in capturing and somehow marking subjects at different occasions so that a *capture configuration*, $\mathbf{r} = (r_1 \ \cdots \ r_J)$, may be associated to each subject captured at least once, where J is the number of capture occasions (or *lists*) and r_j is equal to 1 if the subject has been captured at the j-th occasion and 0 otherwise. The population is estimated on the basis of the resulting data that may be arranged in an *incomplete* contingency table with, at most, $2^J - 1$ not empty cells, as the entry corresponding to the subjects never captured is systematically empty. Estimation is usually carried out through a conditional maximum likelihood approach: the multinomial distribution is assumed for the (hypothetical) contingency table including also the missing cell, while the parameters of the model are estimated on the basis of the incomplete

table (Sanathanan, 1972). Consequently, N is estimated as $n/[1 - \hat{p}(\mathbf{0})]$, where $\hat{p}(\mathbf{0})$ is the estimated probability of the missing configuration and n is the sample size. Since a saturated model may not be used for these data, a variety of restricted models has been proposed. Many of these models are of the log-linear type (Fienberg, 1972, Darroch *et al.*, 1993), but a more recent approach is based on the Latent Class (LC) model; Agresti (1994) was one of the first to use the LC approach in this context. This model assumes that the population may be divided into k classes so that the subjects within each class have the same degree of "catchability" with respect to the same list. Moreover, given the latent class, the conditional probability of being captured in a certain list is independent of that of being captured in other lists. This allows to account for heterogeneity between individuals; furthermore, the parameters of the model may be easily interpreted. This model has also many variants as a finite-mixture version of the Rasch model (Lindsay *et al.*, 1991) that has been applied with success in the capture-recapture context by Bartolucci and Forcina (2001).

Bayesian literature on capture-recapture data has had some developments only in the last two decades. One of the first important contributions seems to be that of Castledine (1981) who developed a model in which capture probabilities may vary between sampling occasions and provided an approximation to the posterior distribution of N under the proposed model. His approach has been refined by Smith (1991) who derived the exact posterior distribution of N. Later, George and Robert (1992) dealt with Bayesian inference for several capture-recapture models, while Madigan and York (1997) proposed the use of a class of decomposable Bayesian graphical models which can incorporate covariates. More recent papers are those of Fienberg *et al.* (1999), who dealt with a continuous-trait version of the Rasch model, and Basu and Ebrahimi (2001) and Tardella (2002) who proposed alternative ways for dealing with population heterogeneity. However, we believe that Bayesian inference for the LC model, which represents one of the most sensible ways to account for hetorogeneity between individuals, can be successfully used for capture-recapture data. Thus, our aim is to develop this type of inference using Markov chain Monte Carlo (MCMC) methods (Tierney, 1994) for estimation purposes. In particular, we consider models with a varying number of latent classes and take advantage of the Reversible Jump (RJ) strategy (Green, 1995), which allows joint inference about the model and the parameters. The RJ algorithm has recently been employed in Bayesian log-linear models for estimating population size by King and Brooks (2001). We improve the efficiency of the estimators obtained via RJ by using the Delayed Rejection (DR) strategy (Tierney and Mira, 1999, Green and Mira, 2001).

From the MCMC output, we can easily obtain a point estimate as well as a credibility interval for N. Furthermore, as a result of the simulation, we

also get the estimated probability of the number of classes, given the data at hand, and the estimated posterior distribution for the model parameters, conditional to the number of classes.

The paper is organized as follows. In Section 2 we illustrate the LC model and the prior distributions on the parameters that we shall use for inference. In Section 3 we present the MCMC estimation approach based on the RJ and the DR algorithm. Finally, in Section 4, we analyse two real data sets, the first concerning the number of snowshoe hares in a certain region and the second concerning cases of diabetes in Casale Monferrato, Italy.

2.   BAYESIAN LATENT CLASS MODEL

The basic assumption of the LC model is that the population is made of k homogeneous classes, in the sense that the subjects in any class share the same probability of being caught in the J lists. So, the probability of the capture configuration $\mathbf{r}$ is given by

$$p(\mathbf{r}) = \sum_{c=1}^{k} \pi_c \lambda(\mathbf{r} \mid c), \tag{1}$$

where π_c is the weight of the latent class c and $\lambda(\mathbf{r} \mid c)$ is the conditional probability of $\mathbf{r}$ for the subjects in the same class that, under the assumption of *local independence*, is given by

$$\lambda(\mathbf{r} \mid c) = \prod_{j=1}^{J} \lambda_{j|c}^{r_j} (1 - \lambda_{j|c})^{1-r_j}, \tag{2}$$

where $\lambda_{j|c}$ is the conditional probability of being caught at the j-th occasion; the latter parameters will be collected into the vector $\boldsymbol{\lambda}$ by letting the index j run faster than c. Note that the model above is invariant to permutation of the class labels ($c = 1, 2, ..., k$) therefore we adopt a unique labeling in which the π_c are in increasing order.

We assume the following prior distributions for the parameters of the model above:
- the number of latent classes, k, has Uniform distribution between 1 and k_{max};
- the vector $\boldsymbol{\pi} = (\pi_1 \quad \cdots \quad \pi_k)'$ has Dirichlet distribution with parameter $\mathbf{v} = (v_1 \quad \cdots \quad v_k)'$, restricted to the set $\pi_1 < \pi_2 < \cdots < \pi_k$;

- for any j and c, $\lambda_{j|c}$ has Beta distribution with parameters $\beta = (\beta_1 \quad \beta_2)'$.

Moreover, given the parameters, we assume that the frequencies of the hypothetical table, including also the missing cell, follow the multinomial distribution with parameters N and $\{p(\mathbf{r})\}$. According to Sanathanan (1972), this implies that the likelihood of the observed frequencies, collected in the column vector $\mathbf{y} = \{y(\mathbf{r}), \mathbf{r} \neq \mathbf{0}\}$, is

$$L(\mathbf{y} \mid \mathbf{p}) = \prod_{r \neq 0} \left[\frac{p(\mathbf{r})}{\sum_{r \neq 0} p(\mathbf{r})} \right]^{y(\mathbf{r})},$$

with $\mathbf{p} = \{p(\mathbf{r}), \mathbf{r} \neq \mathbf{0}\}$.

3.   BAYESIAN ESTIMATION

Our main aim is estimating the posterior distribution of N. For this purpose, we construct a Markov chain that has the posterior distribution of interest as its unique stationary and limiting distribution. Since the number of latent classes is not a priori fixed, we use the RJ algorithm (Green, 1995) to construct such a Markov chain. At any step of the algorithm, a candidate move is proposed and accepted with a probability that preserves reversibility with respect to the target posterior. As we show in the remaining part of this section, some of the moves try to jump between models with a different number of latent classes; the others simply update the parameters within the current model and so are of standard Metropolis-Hastings type. In order to improve the efficiency of the algorithm, we adopt the DR strategy (Green and Mira, 2001) for the moves of the first type, that is, upon rejection of a candidate move, instead of retaining the same position as done in a standard Metropolis-Hastings algorithm, we attempt a second stage move whose acceptance probability is appropriately computed. This strategy leads to an algorithm that dominates the standard RJ in the sense of the Peskun ordering (Peskun, 1973); thus, all the resulting MCMC estimates have a smaller asymptotic variance.

3.1   Updating the parameters of the current model

When we do not change the number of latent classes, a Metropolis-Hastings step is performed to update the parameters of the model. As

proposal distribution for the weights of the latent classes we use the distribution

$$\text{Dirichlet}\left(\sigma\frac{\boldsymbol{\pi}}{\min(\boldsymbol{\pi})}\right),$$

where $\boldsymbol{\pi}$ denotes the current value of the weights and σ is a positive constant suitably chosen. If the entries of the proposed parameter vector, $\boldsymbol{\pi}'$, are not in increasing order, such a vector is discarded and a new one is drawn. Then, $\boldsymbol{\pi}'$ is accepted with probability

$$\min\left\{1,\frac{L(\mathbf{y}\,|\,\mathbf{p}')\,p(\boldsymbol{\pi}'\,|\,k)\,q(\boldsymbol{\pi}\,|\,\boldsymbol{\pi}',k)}{L(\mathbf{y}\,|\,\mathbf{p})\,p(\boldsymbol{\pi}\,|\,k)\,q(\boldsymbol{\pi}'\,|\,\boldsymbol{\pi},k)}\right\},$$

where $\boldsymbol{p}'$ is the proposed vector of joint probabilities whose entries are computed according to (1), $p(\boldsymbol{\pi}\,|\,k)$ denotes the density of the prior distribution of $\boldsymbol{\pi}$ and $q(\boldsymbol{\pi}'\,|\,\boldsymbol{\pi},k)$ denotes that of the proposal distribution. Note that, because of the constraint $\pi_1 < \pi_2 < \cdots < \pi_k$, these densities are equal to $k!$ times the corresponding Dirichlet densities.

The conditional probabilities of being caught, $\lambda_{j|c}$'s, are updated independently. For any j and c, we propose $\lambda'_{j|c}$ from the distribution $\text{Beta}(\upsilon,\omega)$, where $\upsilon = \left[\lambda_{j|c}(\omega-2)+1\right]/(1-\lambda_{j|c})$ and $\omega = \alpha + b\pi_c$, with $\alpha > 1$ and $b > 0$; note that υ and ω are chosen so that the proposal distribution has mode equal to the current value of $\lambda_{j|c}$ and variance proportional to $1/\pi_c$. The proposed $\lambda'_{j|c}$ is then accepted with probability

$$\min\left\{1,\frac{L(\mathbf{y}\,|\,\mathbf{p}')\,p(\lambda'_{j|c}\,|\,k)\,q(\lambda_{j|c}\,|\,\lambda'_{j|c},k)}{L(\mathbf{y}\,|\,\mathbf{p})\,p(\lambda_{j|c}\,|\,k)\,q(\lambda'_{j|c}\,|\,\lambda_{j|c},k)}\right\},$$

where $p(\lambda_{j|c}\,|\,k)$ again denotes the density of the prior distribution and $q(\lambda'_{j|c}\,|\,\lambda_{j|c},k)$ that of the proposal distribution.

3.2 Changing the number of latent classes

Updating the value of k implies a change of dimensionality for the parameters of the model. The approach we follow consists in a random choice between splitting an existing latent class into two and combining two existing latent classes into one. Using the standard notation (see, for

instance, Green, 1995), we denote the probabilities of these alternatives by b_k and $d_k = 1 - b_k$, respectively, where k is the current number of latent classes. Of course, $b_1 = 1$ and $b_{k_{max}} = 0$; otherwise, we choose $b_k = d_k = 0.5$, for $k = 2, 3, ..., k_{max} - 1$.

The **combine proposal** consists in randomly choosing a pair of adjacent latent classes (c_1, c_2) that we try to merge into a new one, labelled $c*$, whose parameters are chosen as follows

$$\pi'_{c*} = \pi_{c1} + \pi_{c2}$$

$$\lambda'_{j|c*} = \lambda_{j|c_2}, \qquad j = 1, ..., J.$$

The **split proposal** starts by choosing a latent class c^* at random that we try to split into two new ones, labeled c_1 and c_2, augmenting k by 1. The corresponding parameters are chosen as follows. We generate u and $v_j, j = 1, ..., J$, from $\text{Beta}(2, 4)$ and $\text{Beta}(1,1)$, respectively; then, we set

$$\pi'_{c1} = \pi_{c*} u, \qquad \pi'_{c_2} = \pi_{c*}(1 - u),$$

$$\lambda'_{j|c_1} = v_j, \qquad \lambda'_{j|c_2} = \lambda_{j|c*}, \qquad j = 1, ..., J.$$

Obviously, we need to check that the constraints on the weights are satisfied.

According to the RJ recipe, the acceptance probability for the split move is

$$\alpha_1\left(\{\boldsymbol{\pi}, \boldsymbol{\lambda}\}, \{\boldsymbol{\pi}', \boldsymbol{\lambda}'\}\right) = \min\left(1, A_1\right),$$

where

$$A_1 = \frac{L(\mathbf{y} \mid \mathbf{p}')}{L(\mathbf{y} \mid \mathbf{p})} \times \frac{p(k+1)}{p(k)} \times (k+1) \frac{\left(\pi'_{c_1}\right)^{v_{c_1} - 1}\left(\pi'_{c_2}\right)^{v_{c_2} - 1}}{\pi_{c*}^{v_{c*} - 1}}$$

$$\times \prod_{j=1}^{J} \frac{1}{B(\beta_1, \beta_2)}\left(\frac{\lambda'_{j|c_1}\lambda'_{j|c_2}}{\lambda_{j|c*}}\right)^{\beta_1 - 1}\left[\frac{\left(1 - \lambda'_{j|c_1}\right)\left(1 - \lambda'_{j|c_2}\right)}{1 - \lambda_{j|c*}}\right]^{\beta_2 - 1}$$

$$\times \frac{d_{k+1}}{b_k} \times \left[f(u, 2, 4)\prod_{j=1}^{J} f(v_j, 1, 1)\right]^{-1} \times \pi_{c*}, \tag{3}$$

$B(r,q)$ is the Beta function and $f(x,r,q)$ denotes the Beta(r,q) density. The first two lines of (3) are the product of the likelihood ratio ant the prior ratio for the parameters of the model. The third line is the product of the proposal ratio and the Jacobian of the transformation from $(\boldsymbol{\lambda}, \boldsymbol{\pi})$ to $(\boldsymbol{\lambda}', \boldsymbol{\pi}')$.

The acceptance probability for the combine move is $\min\left(1, A_1^{-1}\right)$, with some obvious adjustments in the expression for A_1.

3.2.1   Delayed Rejection strategy

When a rejection decision is taken in either a split or a combine move, instead of advancing the simulation time and retaining the same position, we make another attempt to move according to the DR strategy of Tierney and Mira (1999). This second attempt is done also in the case in which a split or a combine move is rejected due to mis-ordering of the weights. If at the first stage we rejected a combine move, at the second stage we try again to combine two classes but with a more "timid" move, that is we merge the latent class c_2, chosen at the first stage of the RJ, with the latent class 1, i.e. the one with the smallest weight. We then chose the parameters of the new latent class as follows

$$\pi_{c*}'' = \pi_1 + \pi_{c_2}$$

$$\lambda_{j|c*}'' = \lambda_{j|c_2}, \qquad j = 1,\dots,J.$$

If, instead, at the first stage we rejected a split proposal, at the second stage we try again to split the latent class $c*$, chosen at the first stage, but now into the latent classes 1 and c_2. The corresponding parameters are chosen as at the first stage, namely as

$$\pi_1'' = \pi_{c*}\tilde{u}, \qquad \pi_{c_2}'' = \pi_{c*}\left(1 - \tilde{u}\right),$$

$$\lambda_{j|1}'' = \tilde{v}_j, \qquad \lambda_{j|c_2}'' = \lambda_{j|c*}, \qquad j = 1,\dots,J,$$

but in this case, the random number $\tilde{u}$ is generated from Beta$(1,8)$, while $\tilde{v}_j, j = 1,\dots,J$, is still generated from Beta$(1,1)$. As before we need to check that the constraints on the weights $\boldsymbol{\pi}''$ are satisfied.

Following Green and Mira (2000), the second stage proposal is then accepted with probability equal to

$$\alpha_2\left(\{\boldsymbol{\pi},\boldsymbol{\lambda}\},\{\boldsymbol{\pi}'',\boldsymbol{\lambda}''\}\right)=\min\left(1,A_2\right),$$

where

$$A_2 = \frac{L(\mathbf{y}\mid\mathbf{p}'')}{L(\mathbf{y}\mid\mathbf{p})}\times\frac{p(k+1)}{p(k)}\times(k+1)\frac{\left(\pi_1''\right)^{v_1-1}\left(\pi_{c_2}''\right)^{v_{c_2}-1}}{\pi_{c*}^{v_{c*}-1}}$$

$$\times\prod_{j=1}^{J}\frac{1}{B(\beta_1,\beta_2)}\left(\frac{\lambda_{j|1}''\lambda_{j|c_2}''}{\lambda_{j|c*}}\right)^{\beta_1-1}\left[\frac{\left(1-\lambda_{j|1}''\right)\left(1-\lambda_{j|c_2}''\right)}{1-\lambda_{j|c*}}\right]^{\beta_2-1}$$

$$\times\frac{d_{k+1}}{b_k}\times\left[f(\tilde{u},1,8)\prod_{j=1}^{J}f(\tilde{v}_j,1,1)\right]^{-1}\times\pi_{c*}$$

$$\times\frac{1-\alpha_1\left(\{\boldsymbol{\pi}'',\boldsymbol{\lambda}''\},\{\boldsymbol{\pi}^\dagger,\boldsymbol{\lambda}^\dagger\}\right)}{1-\alpha_1\left(\{\boldsymbol{\pi},\boldsymbol{\lambda}\},\{\boldsymbol{\pi}',\boldsymbol{\lambda}'\}\right)} \tag{4}$$

and $\{\boldsymbol{\pi}^\dagger,\boldsymbol{\lambda}^\dagger\}$ represents the first stage proposal in the reverse path from $\{\boldsymbol{\pi}'',\boldsymbol{\lambda}''\}$ to $\{\boldsymbol{\pi},\boldsymbol{\lambda}\}$; this is given by

$$\pi_{c*}^\dagger = \pi_{c_1}'' + \pi_{c_2}''$$

$$\lambda_{j|c*}^\dagger = \lambda_{j|c_2}'', \qquad j=1,\dots,J.$$

The last line in (4) is now the ratio between the rejection probabilities of the first stage moves (for the forward path at the denominator and for the hypothetical reverse path at the denominator).

The acceptance probability for the second stage combine move is $\min\left(1,A_2^{-1}\right)$, with some obvious adjustments in the expression for A_2 and with

$$\pi_{c_1}^\dagger = \pi_{c*}''u, \qquad \pi_{c_2}^\dagger = \pi_{c*}''(1-u),$$

$$\lambda_{j|c_1}^\dagger = v_j, \qquad \lambda_{j|c_2}^\dagger = \lambda_{j|c*}'', \qquad j=1,\dots,J.$$

Green and Mira (2001) showed that this second stage acceptance probability preserves reversibility (and thus stationarity) of the Markov chain with respect to the posterior distribution of interest.

3.3   Inference from the MCMC output

At the end of the algorithm, we obtain a sample drawn from the joint posterior distribution of $(\boldsymbol{\pi} \quad \boldsymbol{\lambda} \quad k)$; we denoted this sample by $(\boldsymbol{\pi}^{(t)} \quad \boldsymbol{\lambda}^{(t)} \quad k^{(t)})$, $t = 1, \ldots, T$, where T is the number of sweeps. From this output, we can estimate the posterior probability of k as the ratio between the number of times the algorithm visited the model with k latent classes and the total number of sweeps.

For any sweep t we also have that the ratio $n/(\mathbf{1'p}^{(t)})$ is an estimate of the population size. Averaging these estimates across all the MCMC sweeps gives an "overall" point estimate of the population size, i.e.:

$$\hat{N} = \frac{1}{T} \sum_{t=1}^{T} \frac{n}{\mathbf{1'p}^{(t)}} .$$

Credibility intervals can also be easily obtained. Notice that overall estimates are calculated without choosing a specific number of latent classes and, in this sense, they take into account the uncertainty existing about the model. In addition, it is possible to obtain "with-in" model estimates, by averaging across the MCMC sweeps, conditional on fixed values of k.

Even if the focus of our analysis is on estimating the population size, evaluation of the marginal probability of appearing in any list can also be of interest; this is particularly useful when different capturing sources need to be compared on the basis of their effectiveness. Let λ_j denote such probability for the j-th list; a natural MCMC estimator for this quantity is

$$\hat{\lambda}_j = \frac{1}{T} \sum_{t=1}^{T} \sum_{c=1}^{k^{(t)}} \lambda_{j|c}^{(t)} \pi_c^{(t)} .$$

4.   APPLICATIONS

The method for estimating the population size illustrated so far is applied to two different capture-recapture data sets. In the RJ sampler with DR, the following settings were used for previously unspecified constants: $k_{\max} = 10$, $\mathbf{v} = 1$, $\boldsymbol{\beta} = 1$, $\sigma = 10$, $a = 2$ and $b = 10$. Simulation studies not reported here indicate that results are quite robust to reasonable perturbations of these values.

4.1   Rabbits Redux

The first data set refers to a population of snowshoe hares. The study reported by Cormack (1992) is based on $J = 6$ consecutive trapping days. The data set has also been analysed by Agresti (1994). The data on the 68 observed animals are presented in Table 1. The 63 observable cells are given in the standard order 000001, 000010, 000011, … by rows in the table.

Table 2-1. Results of capture-recapture of snowshoe hares and estimated probabilities of capture configurations (in italic).

–	4	4	2	3	4	1	1	5	2	0	1	0	1	0	0
9.7	*4.6*	*3.7*	*1.9*	*3.6*	*2.5*	*1.4*	*1.1*	*2.7*	*1.6*	*1.0*	*0.7*	*1.1*	*1.1*	*0.4*	*0.6*
6	1	3	1	3	3	0	1	0	2	0	1	0	2	0	1
4.2	*2.6*	*1.8*	*1.2*	*1.9*	*2.0*	*0.8*	*1.0*	*1.2*	*1.2*	*0.5*	*0.7*	*0.6*	*1.2*	*0.3*	*0.9*
3	1	2	0	2	0	0	1	1	0	1	0	1	0	0	0
2.8	*1.2*	*1.2*	*0.6*	*1.0*	*0.7*	*0.5*	*0.3*	*0.9*	*0.5*	*0.4*	*0.3*	*0.4*	*0.3*	*0.2*	*0.2*
0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
1.1	*0.6*	*0.5*	*0.4*	*0.5*	*0.5*	*0.3*	*0.4*	*0.3*	*0.3*	*0.2*	*0.3*	*0.2*	*0.4*	*0.1*	*0.4*

The results reported here correspond to 100 000 sweeps of the MCMC algorithm, including a burn-in of 50 000 sweeps. The algorithm seems to mix well over the parameter space, with an acceptance rate of the split/combine move approximately equal to 15% at the first stage and to 20% at the second stage. A plot of the changes in k against the number of sweeps is presented in Figure 1(a) (for sake of clarity, data are plotted every 100 sweeps). It shows that the MCMC algorithm mixes well over k, excursions into very high values being short-lived. A useful check on the stationarity is given by the plot of the cumulative occupancy fractions for different values of k against the number of sweeps. This is presented in Figure 1(b), where it can be seen that the burn-in is more than adequate to achieve stability in the occupancy fractions.

Estimated posterior probabilities for k are given in Figure 1(c). It is immediately apparent that there are a number of competing explanations of the data which are tenable. The posterior for k favors 2 to 4 latent classes. Agresti (1994) also found the ordinary LC model, as well as the quasi-symmetric model, with two classes to fit the data quite well. However, in our context, we do not need to choose a specific number of latent classes, and we can take into account the uncertainty existing about the model, overcoming the model-dependence problem of the classical approach.
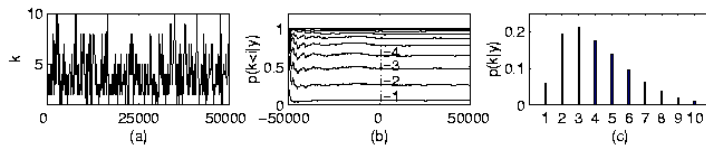
Figure 1. Example of trace of k for 50 000 sweeps after burn-in (a), cumulative occupancy fractions for complete run including burn-in (b) and posterior distribution of k (c).

Overall point estimates for the probability of each capture configuration are given in Table 1. Cellwise inspection reveals a good fit of the model to the observed data. Table 2 provides the posterior model probabilities for the a posteriori most probable models, together with the corresponding with-in model estimates and 95% credibility intervals for N. Overall estimates are also presented, as well as the results obtained by Agresti (1994) with different classical LC models. The overall estimate for N that we obtain is 77.7, suggesting that only 10 snowshoe hares were missed by all 6 trapping occasions. Agresti (1994) obtains a very similar result using a quasi-symmetric LC model with 2 classes. Our overall credibility interval is (72.2, 87.3), which is larger than the confidence interval found by Agresti (1994) for the quasi-symmetric model, but smaller than those he found for ordinary models with 2 or 3 latent classes. It is probably worthwhile to spend a few words on this result. Agresti (1994) stresses that for such sparse data, quite different models can appear to fit adequately yet can provide highly diverse point and interval estimates of N. Particularly with small sample sizes, a model selection procedure may suggest a model that is much simpler than the one that truly represents reality well, leading to confidence intervals for N that tend to be too narrow. We believe, therefore, that the overall credibility interval we estimated, properly reflecting our uncertainty on the "true" model, is not affected by this problem. For this reason it is not as "optimistic" as the one for the simple quasi-symmetric model. On the other side, it has the advantage of being narrower than the confidence intervals found by Agresti for models more complicated than the quasi-symmetric one.

The estimates of the probability of being caught at the j-th occasion are presented in Table 3. In the present case there are no relevant differences between the capture occasions, i.e. between the six consecutive trapping days, revealing that the captures were likely performed under the same conditions.

Table 2-2. Posterior model probabilities for most probable models, together with point and interval estimates of population size.

Model	Posterior probability	Estimate of N	95% CI
Bayesian latent class			
$k = 2$	0.192	79.0	(72.7, 89.0)
$k = 3$	0.211	78.3	(72.5, 87.8)
$k = 4$	0.176	78.0	(72.4, 87.9)
Model-averaged		77.7	(72.2, 87.3)
Classical latent class			
Quasi-symmetric, $k = 2$		77:3	(73.1, 84.9)
Ordinary, $k = 2$		85.2	(76.2, 104.0)
Ordinary, $k = 3$		81.3	(72.0, 103.6)

Table 2-3. Point and interval estimates of the probabilities of being caught at different occasions.

	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
$\hat{\lambda}_j$	0.23	0.37	0.28	0.35	0.31	0.42
95% CI	(0.14,0.33)	(0.26,0.49)	(0.18,0.38)	(0.24,0.46)	(0.21,0.42)	(0.30,0.53)

Table 2-4. Data from prevalent cases of known diabetes for resident of Casale Monferrato, Piemonte, Italy, and estimated probabilities of capture configurations (in italic).

–	10	182	8	74	7	20	14	709	12	650	46	104	18	157	58
	506.2	*8.9*	*179.4*	*8.3*	*72.2*	*7.1*	*27.6*	*12.8*	*708.3*	*16.8*	*649.2*	*45.0*	*102.2*	*18.0*	*155.1 58.2*

4.2　　Diabetes study

The data in Table 4 are taken from Bruno *et al.* (1994) and refer to 2069 cases of diabetes in a small town in northern Italy recorded in October 1988 on the basis of four different lists: (i) clinics (all patients diagnosed as diabetics by the local clinic and/or family physicians); (ii) hospitals (all patients discharged with a primary or secondary diagnosis of diabetes in all private and public hospitals in the region); (iii) prescriptions (a computerized list of insulin and hypoglycemic prescriptions); (iv) reimbursement (all patients who requested reimbursement for insulin and reagent strips). The

same dataset is also analysed by Yip *et al.* (1995a), Biggeri *et al.* (1999), Fienberg *et al.* (1999), and Bartolucci and Forcina (2001).

For this dataset, the MCMC algorithm mixes slightly slower than for the previous one. This is is due to the fact that now the cell counts are larger, the posterior densities under the different models are more picked and, therefore, it is slightly harder to jump from a model to another one. For this reason, we allowed for a longer burn-in: we run 200 000 sweeps of the algorithm and discarded the first 150 000 as burn-in. The slower mixing also determined a high autocorrelation in the MCMC output, with an *integrated autocorrelation time* (see Section 5) equal to 3118.2. In order to reduce this autocorrelation and, consequently, the variance of the estimates, we sub-sampled from the MCMC output and considered only one every 10 sweeps for estimating purpose. This reduced the integrated autocorrelation time to 311.5.

Looking at the changes in k (Figure 2(*a*)) and cumulative occupancy fractions (Figure 2(*b*)), it can be seen that the mixing is still satisfactory (with an acceptance rate of the split/combine move approximately equal to 4% at the first stage and to 8% at the second stage) and the algorithm eventually converges. For this dataset, the posterior density for k is more concentrated than for the previous one. The three most probable models, i.e. those with 3, 4 and 5 latent classes, have a total posterior density equal to 0.97. Moreover, models with 1, 2 and 10 latent classes are never visited. The most probable model seems to be the one with 3 latent classes. Notice that both Biggeri *et al.* (1999) and Bartolucci and Forcina (2001) consider models with only 2 latent classes. However, given the latent class, they allow for interactions between the different lists while we do not. This is likely the reason why we found models with a larger number of latent classes to fit the data better.
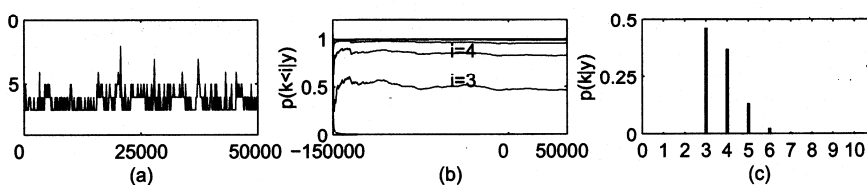

Figure 2. Example of trace of k for 50 000 sweeps after burn-in (a), cumulative occupancy fractions for complete run including burn-in (b) and posterior distribution of k (c).

Table 5 shows with-in model and overall estimates and credibility intervals for N, compared with the results obtained by Biggeri *et al.* (1999) and Bartolucci and Forcina (2001). Note that the with-in model estimates and credibility intervals are based on the whole MCMC run, without sub-

sampling, to avoid the risk of calculating estimates on a too small number of runs.

Both the with-in model and the overall estimates of N we obtained are in between the one obtained by Bartolucci and Forcina (2001) and the one obtained by Biggeri *et al.* (1999). Moreover our credibility intervals include both the confidence intervals in Bartolucci and Forcina (2001) and in Biggeri *et al.* (1999). It is also interesting to notice that, as the number of latent classes considered increases, the estimate of N decreases and the credibility interval gets larger and larger.

Table 2-5. Posterior model probabilities for most probable models, together with point and interval estimates of population size.

Model	Posterior probability	Estimate of N	95% CI
Bayesian latent class			
$k = 3$	0.463	2616.0	(2275.5, 3154.0)
$k = 4$	0.370	2543.2	(2274.1, 3233.3)
$k = 5$	0.132	2534.3	(2269.1, 3355.3)
Model-averaged		2575.5	(2273.6, 3204.2)
Classical latent class			
Biggeri *et al.* (1999)		2696	(2502, 2950)
Bartolucci and Forcina (2001)		2403	(2280, 2585)

In Table 6 the estimated probabilities of being caught by the different lists are reported. For this dataset, a considerable heterogeneity between lists is evident. The list clinics in the one with the highest "catchability" followed by prescriptions, hospitals and reimbursements. This ordering of the lists with respect to their catchability is completely consistent with that found in Bartolucci and Forcina (2001).

Table 2-6. Point and interval estimates of the probabilities of being caught by the different lists.

	$j = 1$	$j = 2$	$j = 3$	$j = 4$
$\hat{\lambda}_j$	0.69	0.18	0.45	0.07
95% CI	(0.55,0.78)	(0.14,0.21)	(0.35,0.51)	(0.05,0.08)

5. COMPARING DELAYED REJECTION AND STANDARD REVERSIBLE JUMP SAMPLER

In general, comparing the performance of MCMC samplers is not an easy task. Following Green and Mira (2001), we define 'performance' as the efficiency of estimating the expectation, under the stationary distribution of the Markov chain, of a given function ϕ on the state space, $E(\phi)$. For our comparisons we focus on the estimating the population size (this specifies indirectly the function ϕ).

There are two quantities to take into account when comparing the performance of the MCMC samplers: the running time R needed to obtain a fixed number of sweeps T plus a fixed number of burn-in steps, and the integrated autocorrelation time, $\tau = \sum_{k=-\infty}^{\infty} \rho_k$ where $\rho_k = \text{Cov}\left[\phi\left(X_0\right), \phi\left(X_k\right)\right]/\sigma^2$, is the lag-$k$ autocorrelation, X_i being the value of the Markov chain at time i, and σ^2 is the variance of ϕ under the stationary distribution (assumed to be finite). The product of τ and R gives a reasonable measure of the efficiency of an MCMC algorithm: it estimates the running time needed to obtain the same precision in estimating $E(\phi)$ as from T independent draws from the stationary distribution.

In our comparison it is reasonable to believe the two codes developed are similarly efficient (and thus an average of the running times of a few simulations is a good estimate of R) since a two stage DR algorithm is compared with a one-stage algorithm (i.e. a standard RJ), which actually forms the first stage of the DR, therefore the DR code is a superset of the other.

To estimate τ we used Sokal's adaptive truncated periodogram estimator (Sokal, 1989), $\sum_{|k| \leq M} \hat{\rho}_k$ where the window width M is chosen adaptively as the minimum integer such that $M \geq 3\hat{\tau}$. We used the fast Fourier transform to estimate the autocorrelations $\hat{\rho}_k$.

The value in the column named 'ratio' is the ratio of R times $\hat{\tau}$ for the standard RJ method and the DR algorithm. Thus values greater than one indicate how much more efficient the DR strategy is. We know that by the Peskun-Tierney theorem (Peskun, 1973; Thierney, 1998) the DR algorithm will always give smaller values of τ than a regular Metropolis-Hastings algorithm, for every function of interest ϕ.

Table 7 and 8 are obtained by averaging 10 simulations and refer to the examples studied in Section 4.1 and 4.2 respectively.

Typically, for a given number of steps, a DR algorithm has a running time which is larger than the corresponding one stage MCMC algorithm, since the DR strategy consists in trying harder to move away from the

current position (and this takes simulation time). Still, in both tables the value in the column named ratio shows that the gain in terms of asymptotic efficiency due to the DR strategy more than compensates for the extra computational time needed to run the DR algorithm.

The high value of $\hat{\tau}$ in Table 8 has motivated the much larger sample size and burn-in used in Section 4.2 to draw our inference and the need for subsampling.

Table 2-7. Rabbits: T = 10 000 burn-in = 5 000

	no delay		delay		ratio
	R in hours	$\hat{\tau}$	R in hours	$\hat{\tau}$	
N	1.82	31.79	1.85	25.87	1.21

Table 2-8. Diabetes: T = 100 000 burn-in = 50 000

	no delay		delay		ratio
	R in hours	$\hat{\tau}$	R in hours	$\hat{\tau}$	
N	4.70	3134.8	4.83	2033.2	1.50

ACKNOWLEDGEMENTS

REFERENCES

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

Bartolucci, F. and Forcina, A. (2001). The Analysis of Capture-Recapture Data with a Rasch-type Model allowing for Conditional Dependence and Multidimensionality. *Biometrics* 57, 714-719.

Basu, S. and Ebrahimi, N. (2001). Bayesian capture-recapture methods for error detection and estimation of population size: hetorogeneity and dependence. *Biometrika* 88, 269-279.

Biggeri, A., Stanghellini, E., Merletti, F. and Marchi, M. (1999). Latent class models for varying catchability and correlation among sources in capture-recapture estimation of the size of a human population. *Statistica Applicata* 11, 563-586.

Bruno, G., Biggeri, A., Merletti, F., Laporte, R., McCarthy, D. and Pagano, G. (1994). Applications of capture-recapture to count diabetes. *Diabetes Care* 17**,** 548-556.

Castledine, B. J. (1981), A Bayesian analysis of multiple-recapture sampling for a closed population, *Biometrika* 68, 197-210.

Cormack, R. M. (1992). Interval estimation for mark-recapture studies of closed populations. *Biometrics* 48, 567-576.

Darroch, J.N. (1958). The multiple-recapture census I. Estimation of a closed population. *Biometrika* 45, 343-359.

Darroch, J. N., Fienberg, S. E., Glonek, G. F. V. and Junker, B. W. (1993). A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *Journal of the American Statistical Association* 88, 1137-1148.

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

Fienberg, S. E., Johnson, M. S. and Junker, B. W. (1999). Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *Journal of the Royal Statistical Society A* 162, 383-405.

George, E. I. and Robert, C. P. (1992). Capture-recapture estimation via Gibbs sampling. *Biometrika* 79, 677-683.

Green, P. J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711-732.

Green, P. J. and Mira, A. (2001), Delayed Rejection in Reversible Jump Metropolis-Hastings. *Biometrika* 88, 1035-1053.

King, R. and Brooks, S. P. (2001). On the Bayesian analysis of population size. *Biometrika* 88, 317-336.

Lindsay, B., Clogg, C. and Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association* 86, 96- 107.

Madigan, D. and York, J. C. (1997). Bayesian methods for estimation of the size of a closed population. *Biometrika* 84, 19- 31.

Peskun, P. H. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika 60, 607-612.*

Sanathanan, L. (1972). Estimating the size of a multinomial population. *The Annals of Mathematical Statistics* 43, 142-152.

Schnabel, Z. E. (1938). The estimation of the total fish population of a lake. *American Mathematical Monthly* 45, 348-52.

Schwarz, C. J. and Seber, A. F. (1999). Estimating animal abundance: review III. *Statistical Science* 14, 427-456.

Smith, P. J. (1991) . Bayesian analyses for a multiple capture-recapture model. *Biometrika 78, 399-407.*

Tardella, L. (2002). A new Bayesian method for nonparametric capture-recapture models in presence of heterogeneity. *Biometrika* 89, 807-817.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics* 22, 1701-1762.

Tierney, L. (1998). A Note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability* 8, 1-9.

Tierney, L. and Mira, A. (1999). Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine* 18, 2507-2515.

Yip, P. S. F., Bruno, C., Tajima, N., Seber, C. A. F., Buckland, S. T., Cormack, R. M., Unwin, N., Chang, Y.-F., Fienberg, S. E., Junker, B. W., LaPorte, R. E., Libman, I. M. and McCarty D. J. (1995a). Capture-recapture and multiple-record systems estimation I: history and theoretical development. *American Journal of Epidemiology* 142, 1047-1058.

Yip, P. S. F., Bruno, C., Tajima, N., Seber, C. A. F., Buckland, S. T., Cormack, R. M., Unwin, N., Chang, Y.-F., Fienberg, S. E., Junker, B. W., LaPorte, R. E., Libman, I. M. and

McCarty D. J. (1995b). Capture-recapture and multiple-record systems estimation II: applications in human diseases. *American Journal of Epidemiology* 142, 1059-1068.