Capture-recapture models and Bayesian sampling

by

Edward I. George and Christian P. Robert †

University of Chicago and Université Paris VI

Summary

Capture-recapture models are widely used to estimate the unknown size of a closed population, N. A successful strategy for exploiting information about N in this setting is obtained through Bayesian modelling, as shown in Castledine (1981). However, direct Bayesian approaches are often cumbersome to implement in this setting. In this paper, we show how Bayesian sampling, using Gibbs sampling and data augmentation, is particularly well suited for use in a wide variety of capture-recapture models, including the multinomial and classical hypergeometric models. This approach can provide accurate approximations of posterior expressions, including the entire posterior distribution.

Keywords: Bayesian computation, data augmentation, Gibbs sampling, conjugate priors, catch propensities, Petersen estimator, behavioral and temporal models.

1. Introduction.

A common experimental setup for estimating N, the unknown size of a closed population, is based on sampling the population more than once, paying special attention to the number of recaptured individuals (i.e. those that appeared in more than one sample). Such setups were first used in the context of estimating the size of wildlife populations, where they were baptised capture-recapture models (see Seber (1982) for an overview and bibliography). This setup has also appeared in proofreading problems (Polya (1976)), in reliability problems in manufacturing quality control and program debugging (Jewell (1985) and Nayak (1988)) and in estimating the number of vital human events (Mark, Setlzer and Kroti (1974)). A recent application which has received much attention is the estimation of coverage error in surveys and censuses (Wolter (1986)).

For the simple case of drawing two samples from a population of unknown size N, Wolter (1986) proposes the following general multinomial setup. He shows that a large variety of capture-recapture models in the literature are equivalent to multinomial experi-

---

ments where the ith population member is sampled according to the following probabilities

$$
\begin{array}{cccc}
 & & \multicolumn{2}{c}{\text{Sample 2}} \\
 & & \text{in} & \text{out} \\
\text{Sample 1} & \text{in} & p_{11}^i & p_{12}^i \\
 & \text{out} & p_{21}^i & p_{22}^i
\end{array}
$$

where $p_{11}^i + p_{12}^i + p_{21}^i + p_{22}^i = 1$. The realisation of such an experiment can be summarised by the counts

$$
\begin{array}{cccc}
 & & \multicolumn{2}{c}{\text{Sample 2}} \\
 & & \text{in} & \text{out} \\
\text{Sample 1} & \text{in} & n_{11} & n_{12} \\
 & \text{out} & n_{21} & n_{22}
\end{array}
$$

where $n_{11} + n_{12} + n_{21} + n_{22} = N$. Note that N remains unknown, since the value n_{22} is not observed. Nonetheless, under certain assumptions on $\mathbf{p}$, the vector of all probabilities p_{jk}^i above, information about N can be extracted from the cell counts n_{11}, n_{12} and n_{21}.

Although a variety of frequentist and likelihood approaches for making inference about N have appeared in the literature (see, e.g., Bishop, Fienberg and Holland (1975, Chapter 6), Burnham *et al.* (1986), Pickands and Raghavachari (1989) and Huggins (1989)), we shall focus on the promising Bayesian approach advocated by Castledine (1981)), Jewell (1985) and Smith (1988). In the capture-recapture setting, the Bayesian approach to exploiting information about N proceeds as follows. For a particular capture-recapture setup, the joint posterior distribution of N and $\mathbf{p}$ can be obtained from the likelihood $L(N, \mathbf{p}|\text{data})$ and the (possibly improper) prior distribution $\pi(N, \mathbf{p})$,

$$
\pi(N, \mathbf{p}|\text{data}) \propto L(N, \mathbf{p}|\text{data})\pi(N, \mathbf{p}). \tag{1.1}
$$

The Bayes estimator for N is then the (formal) posterior mean, $\mathbb{E}[N|\text{data}]$, of the marginal distribution

$$
\pi(N|\text{data}) = \int \pi(N, \mathbf{p}|\text{data})d\mathbf{p} \tag{1.2}
$$

A major deficiency of this approach has been the difficulty of calculating the posterior mean and other posterior quantities. Unfortunately, closed form expressions for the marginal in (1.2) are often unavailable, and thus approximate methods have to be used.

The purpose of this paper is to show how Bayesian sampling is a promising alternative to both analytical calculation and numerical approximation in the Bayesian analysis of capture-recapture models. The essential idea behind Bayesian sampling is to obtain information about marginal posterior distributions by indirect sampling, see Robert (1990). In the capture-recapture setting, this effectively consists of obtaining a random sample from the marginal,

$$
N_1, \ldots, N_n \sim \pi(N|\text{data}), \tag{1.3}
$$

without integrating out $\mathbf{p}$ in (1.2). By taking a large enough sample, the posterior mean or any other posterior quantity can then be estimated to the desired degree of accuracy.

Rather than sample directly from the marginal posterior $\pi(N|\text{data})$, Bayesian sampling exploits the conditional distributions

$$\pi(N|\mathbf{p},\text{data}) \quad \text{and} \quad \pi(\mathbf{p}|N,\text{data}). \tag{1.4}$$

This can be done by using a special case of Gibbs sampling as well as data augmentation, see Gelfand and Smith (1990) or Diebolt and Robert (1990). We note that these two approaches are not the same in general. Starting with an initial value for N, say N_0^*, each sample point in (1.3) is obtained by sampling iteratively from

$$N_k^* \sim \pi(N|\mathbf{p}_{k-1},\text{data}) \quad \text{and} \quad \mathbf{p}_k \sim \pi(\mathbf{p}|N_k^*,\text{data}). \tag{1.5}$$

It follows from Diebolt and Robert (1990) that the distribution of N_k^* converges to $\pi(N|\text{data})$ as k goes to $+\infty$. Thus, for k large enough, $N_1 = N_k^*$ is effectively an observation from $\pi(N|\text{data})$. By repeating this procedure n times, the random sample in (1.3) is obtained.

A powerful advantage of Bayesian sampling over analytical calculation is that obtaining the random sample in (1.3) does not require the marginal posterior (1.2). Instead, all that is needed are the conditional distributions in (1.4). As will be seen in the sequel, these are generally easy to obtain in the capture-recapture setting. Furthermore, useful priors can be chosen so that these conditional distributions will be of standard form, allowing for fast and efficient simulation of the iterative sampling in (1.5). Another advantage of Bayesian sampling is that the multinomial model above can be made to subsume the classical hypergeometric model which also arises in this setting (see, e.g., Darroch (1958)) when the sample sizes are fixed (see Section 4).

This plan of this paper is as follows. In Section 2, we present different models, following the classification of Wolter (1986). In Section 3, we derive the associated Bayesian models and illustrate the advantages of a Bayesian sampling approach. In Section 4, we indicate how our methods also provide a convenient solution for the classical hypergeometric model. Section 5 studies some heterogeneous extensions and gives particular attention to stratified models. Section 6 extends, through an example, the previous results to a multiple recapture setting.

2. Some capture-recapture models.

Consider a closed population of unknown size N. Two random samples are drawn consecutively from this population. Let n_1 and n_2 be the sizes of these two samples, with $n_1 = n_{11} + n_{12}$ and $n_2 = n_{11} + n_{21}$, n_{11} being the size of the intersection of the two samples. The three random variables n_{11}, n_{12} and n_{21} are observed, as described above. We denote by $n.$ the total number of captures, i.e. $n. = n_1 + n_2$, and by n_+ the number of distinct captured objects, i.e. $n_+ = n_{11} + n_{12} + n_{21}$. In the cases considered below, n_{11}, n_{12} and n_{21} are sufficient statistics. Notice that here it is not necessary to know the complete 'history' of each captured individual.

This setup can be generalised for a larger number of recapture events, as in Castledine (1981) or Wolter (1986). However, the extension to these cases is straightforward and,

until Section 6, we focus on the single recapture case. Following Wolter (1986), we present below several capture-recapture models, distinguished by different assumptions on the probabilities of capture.

2.1. The uniform model M_0. For the model M_0, each individual has the same probability p of being captured in the first or the second sample. The likelihood function for this model is

$$L_0(N, p | n_{11}, n_{12}, n_{12}) = \binom{N}{n_{11}\ n_{12}\ n_{21}} p^{n.} (1-p)^{2N-n.}.$$

The maximum likelihood estimator of N is then

$$\hat{N}_0 = \lfloor \frac{n_.^2}{4n_2} \rfloor,$$

where $\lfloor\ .\ \rfloor$ denotes the integer part function.

2.2. The behavioral model M_b. For the model M_b, the probability of recapture c is different from the probability of initial capture, p. If $c > p$, the individuals are said to be 'trap-happy' and if $c < p$, the individuals are 'trap-shy'. For instance, in some wildlife experiments, captured animals are often less likely to be captured a second time, in which case $c < p$. The likelihood function for this model is

$$L_b(N, p, c | n_{11}, n_{12}, n_{21}) = \binom{N}{n_{11}\ n_{12}\ n_{21}} p^{n_+} c^{n_{11}} (1-c)^{n_{21}} (1-p)^{2N-n.}$$

and the maximum likelihood estimator is

$$\hat{N}_b = \left\lfloor n_+ \left(1 - (\frac{n_+ - n_1}{n_1})^2 \right)^{-1} \right\rfloor.$$

2.3. The temporal model M_t. For the model M_t, the probability of capture for the first sample, p_1, is different from the probability of capture in the second sample, p_2. In this case, either the whole population is affected by the first capture or the conditions of the experiment have been modified. The likelihood function for this model is

$$L_t(N, p_1, p_2 | n_{11}, n_{12}, n_{21}) = \binom{N}{n_{11}\ n_{12}\ n_{21}} p_1^{n_1} p_2^{n_2} (1-p_1)^{N-n_1} (1-p_2)^{N-n_2}$$

and the maximum likelihood estimator is

$$\hat{N}_t = \left\lfloor \frac{n_1 n_2}{n_{11}} \right\rfloor.$$

Wolter (1986) also calls M_t the Petersen model, because the maximum likelihood estimator agrees with the maximum likelihood estimator in the classical Petersen model where the sample sizes n_1 and n_2 are fixed (see Section 4).

More complex models can also be introduced. For instance, the combined model M_{bt} takes into account a change between the two captures *and* for the captured objects. Section 5 deals with a general heterogeneous model and Section 6 with a multiple recapture model.

3. Conditional posterior distributions.

A Bayesian analysis of any of the three capture-recapture setups M_0, M_b or M_t described in Section 2, would proceed by multiplying a (possibly improper) prior $\pi(N, \mathbf{p})$ by the corresponding likelihhod L_0, L_b or L_t to obtain the posterior distribution $\pi(N, \mathbf{p}|\text{data})$ as in (1.1). In this section, we illustrate how for certain priors, standard conditional posterior distributions $\pi(\mathbf{p}|N, \text{data})$ and $\pi(N|\mathbf{p}, \text{data})$ are obtained which allow for an efficient sampling simulation. Thus, repeated iterative sampling from these standard conditional posterior distributions as in (1.5) is a readily available method for obtaining a random sample from the marginal posterior $\pi(N|\text{data})$. The Bayes estimates of N, $\mathbb{E}[N|\text{data}]$, or any other posterior quantity of interest, can then be estimated to the desired degree of accuracy. We also show that in these cases, the marginal posterior distribution $\pi(N|\text{data})$ is analytically unwieldy, making infeasible the alternative of 'integrating out' $\mathbf{p}$ in (1.2).

In what follows, we focus on priors of the form

$$\pi(N, \mathbf{p}) = \pi(N)\pi(\mathbf{p}), \tag{3.1}$$

where $\pi(N)$ is Poisson $Po(\lambda)$. Raftery (1988), in the related problem of binomial N estimation, also used a Poisson prior on N. However, it should be apparent that our developments can be adapted to other prior distributions. For instance, Castledine (1981) used an improper prior, $\pi(N) \propto 1/N$ (see also Section 6).

For a prior of the form (3.1) in model M_0, the joint posterior is

$$\pi(N, p|n_{11}, n_{12}, n_{21}) \propto \frac{N!}{(N - n_+)!} p^{n.} q^{2N - n.} \pi(p)\pi(N)$$

$$\propto \frac{(q^2\lambda)^N}{(N - n_+)!}(p/q)^{n.} \pi(p),$$

where $q = 1 - p$. Therefore,

$$\pi(N - n_+|p, n_{11}, n_{12}, n_{21}) \equiv Po(q^2\lambda)$$

$$\pi(p|N, n_{11}, n_{12}, n_{21}) \propto p^{n.} q^{2N - n.} \pi(p).$$

Depending on the prior distribution $\pi(p)$, we may use Bayesian sampling or integrate out the parameter p to obtain the posterior distribution $\pi(N|n_{11}, n_{12}, n_{21})$. For instance, if $\pi(p)$ is $Be(\alpha, \beta)$, the marginal posterior distribution is very cumbersome to compute while Bayesian sampling is straightforward. Indeed, we have

$$\pi(N|n_{11}, n_{12}, n_{21}) \propto \frac{\lambda^N}{(N - n_+)!} \int_0^1 p^{\alpha + n. - 1}(1 - p)^{2N - n. + \beta - 1} dp$$

$$\propto \frac{\lambda^N}{(N - n_+)!} \frac{B(\alpha + n_+, \beta + 2N - n.)}{B(\alpha, \beta)}$$

$$= \frac{\lambda^N}{(N - n_+)!} \frac{(\alpha + n_+ - 1)\dots\alpha(\beta + 2N - n. - 1)\dots\beta}{(\alpha + \beta + 2N - 1)\dots(\alpha + \beta)},$$

while

$$\pi(N - n_+ | p, n_{11}, n_{12}, n_{21}) \equiv Po(q^2\lambda)$$
$$\pi(p | N, n_{11}, n_{12}, n_{21}) \equiv Be(\alpha + n_., \beta + 2N - n_.).$$

Other distributions on N do not modify the complexity of the posterior distribution (except if they have small finite support).

For model M_b, we consider the special case of (3.1) where $\pi(N, \mathbf{p}) = \pi(p)\pi(c)\pi(N)$ and $\pi(N)$ is Poisson $Po(\lambda)$. The joint posterior here is

$$\pi(N, p, c | n_{11}, n_{12}, n_{21}) \propto \frac{N!}{(N - n_+)!} p^{n_+} c^{n_{11}} (1 - c)^{n_{21}} q^{2N - n_.} \pi(p)\pi(c)\pi(N)$$

$$\propto \frac{(q^2\lambda)^N}{(N - n_+)!} p^{n_+} c^{n_{11}} (1 - c)^{n_{21}} q^{-n_.} \pi(p)\pi(c).$$

Therefore,

$$\pi(N - n_+ | p, n_{11}, n_{12}, n_{21}) \equiv Po(q^2\lambda)$$
$$\pi(p | N, n_{11}, n_{12}, n_{21}) \propto p^{n_+} q^{2N - n_.} \pi(p)$$
$$\pi(c | n_{11}, n_{12}, n_{21}) \propto c^{n_{11}} (1 - c)^{n_{21}} \pi(c).$$

Once again, in the case where the prior distributions on p and c are beta distributions, a good approximation to the posterior distribution of N is provided by Bayesian sampling, while a direct approach faces the same computational problems as for the model M_0.

For model M_t, we consider the special case of (3.1) where $\pi(N, \mathbf{p}) = \pi(p_1)\pi(p_2)\pi(N)$ and $\pi(N)$ is Poisson $Po(\lambda)$. The joint posterior here is

$$\pi(N, p_1, p_2 | n_{11}, n_{12}, n_{21}) \propto \frac{N!}{(N - n_+)!} p_1^{n_1} p_2^{n_2} (1 - p_1)^{N - n_1} (1 - p_2)^{N - n_2} \pi(p_1)\pi(p_2)\pi(N)$$

$$\propto \frac{((1 - p_1)(1 - p_2)\lambda)^N}{(N - n_+)!} \left(\frac{p_1}{1 - p_1}\right)^{n_1} \left(\frac{p_2}{1 - p_2}\right)^{n_2} \pi(p_1)\pi(p_2).$$

Therefore,

$$\pi(N - n_+ | p, n_{11}, n_{12}, n_{21}) \equiv Po((1 - p_1)(1 - p_2)\lambda)$$
$$\pi(p_1 | N, n_{11}, n_{12}, n_{21}) \propto p_1^{n_1} (1 - p_1)^{N - n_1} \pi(p_1)$$
$$\pi(p_2 | N, n_{11}, n_{12}, n_{21}) \propto p_2^{n_2} (1 - p_2)^{N - n_2} \pi(p_2).$$

The same remarks as in the previous models apply here.

The prior distributions used above require prior information about the parameters α, β and λ. In the absence of prior information, we suggest $\alpha = \beta = 1$ and either using $\pi(N) = 1/N$ (see Castledine (1981)), which corresponds to the Jeffreys prior $\pi(\lambda) = \lambda^{-1}$, or replacing λ by the maximum likelihood estimator of the appropriate model, thus advocating a pseudo-empirical Bayes approach. Another alternative is developed in the

6

next section, where we show that the classical hypergeometric model can be represented as a limiting case of the temporal model.

4. Hypergeometric model.

We have been dealing until now with the binomial model, where the two sample sizes are supposed to be random. Although adequate for many practical situations, a simpler model has also been studied in the literature, where the two sample sizes are fixed (or the model is conditional on these two sample sizes). It is called the Petersen model or the Darroch model (see Darroch (1958), Casteldine (1981), Seber (1982)). The sample distribution of n_{11} is given by

$$f(n_{11}|N) = \frac{\binom{n_1}{n_{11}}\binom{N-n_1}{n_2-n_{11}}}{\binom{N}{n_2}}. \tag{4.1}$$

It is then obvious that a direct Bayesian approach will lead to rather complicated computations, except in the particular case of bounded uniform priors. We show below that Bayesian sampling allows for a much more efficient treatment.

Starting back with the temporal model M_t, we see that

$$\pi(N|p_1,p_2,n_+) \propto \binom{N}{n_+}\mu^{n_++1}(1-\mu)^{N-n_+}\pi(N),$$

where $\mu = 1 - (1-p_1)(1-p_2)$. In particular, if $\pi(N) = 1$, the posterior distribution of N is $\mathcal{N}eg(n_+,\mu)$. If, in addition, the prior distribution on p_1 and p_2 is $\mathcal{B}e(\alpha,\beta)$, the marginal posterior distribution satisfies

$$\begin{aligned}
\pi(N|n_1,n_2,n_+) &\propto \frac{N!}{(N-n_+)!}\frac{(N-n_1+\beta-1)!}{(N+\alpha+\beta-1)!}\frac{(N-n_2+\beta-1)!}{(N+\alpha+\beta-1)!} \\
&= \frac{N!}{(N-n_+)!}\frac{(N-n_1)!}{N!}\frac{(N-n_2)!}{N!} \\
&\propto \frac{\binom{n_1}{n_{11}}\binom{N-n_1}{n_2-n_{11}}}{\binom{N}{n_2}}
\end{aligned}$$

if $\alpha = 0$ and $\beta = 1$. Therefore, the hypergeometric model can be rewritten as a M_t model with improper prior distributions $\pi(N) = 1$ and $\pi(p) = 1/p$, a fact noted by Castledine (1981). Although these priors may not really correspond to a true prior opinion, combining the hierarchical decomposition of (4.1) with Bayesian sampling provides a very efficient tool for computing the Bayes estimators, since the conditional distributions for p_1 and p_2,

$$\pi(p_i|N,n_i) \propto \mathcal{B}e(n_i,N+1-n_i), \quad (i=1,2)$$

are as easy to simulate as $\mathcal{N}eg(n_+,\mu)$. Note that the temporal aspect of the model is absent from the original formulation, as well as the probability p. In this case, Bayesian sampling makes use of the hierarchical representation to approximate the Bayes estimator, even though it does not necessarily correspond to a true state of Nature. (See Robert (1990) for additional comments on the utilisation of mixture representations by Bayesian sampling.)

5. Heterogeneous extensions.

As considered in Burnham and Overton (1978), Castledine (1981) and Huggins (1989), a more general model can obviously be introduced, namely that the probability is different for each individual (i) and each capture (j) and denoted by p_{ij}. The likelihood is then

$$L(N, \mathbf{p} | \Delta) = \prod_{i=1}^{N} \prod_{j=1}^{2} p_{ij}^{\delta_{ij}} (1 - p_{ij})^{1 - \delta_{ij}},$$

where $\Delta = (\delta_{11}, \ldots, \delta_{N2})$ is the vector of the capture indicators for all members of the population and each capture. By convention, the first members are the captured members. In this more general context, their 'history' obviously counts.

5.1. Bayesian analysis. From a pure Bayesian point of view, we can estimate the parameters of the model if the available prior information is sufficient. For instance, if the priors on the probabilities p_{ij} are all beta $\mathcal{B}e(\alpha, \beta)$ priors, the marginal distribution of Δ given N is

$$
\begin{aligned}
f(\Delta | N) &= \prod_{i=1}^{N} \prod_{j=1}^{2} \int_{0}^{1} \frac{(1 - p_{ij})^{\beta - \delta_{ij}} p_{ij}^{\alpha + \delta_{ij} - 1}}{B(\alpha, \beta)} dp_{ij} \\
&= \prod_{i=1}^{N} \prod_{j=1}^{2} \frac{B(\alpha + \delta_{ij}, \beta + 1 - \delta_{ij})}{B(\alpha, \beta)} \\
&= \prod_{i=1}^{N} \prod_{j=1}^{2} \left(\delta_{ij} \frac{\alpha}{\alpha + \beta} + (1 - \delta_{ij}) \frac{\beta}{\alpha + \beta} \right) \\
&= \left(\frac{\alpha}{\alpha + \beta} \right)^{n.} \left(\frac{\beta}{\alpha + \beta} \right)^{2N - n.}.
\end{aligned}
$$

Therefore, the information about the size of the intersection between the two samples is not used by the likelihood. This result is not very surprising since the probabilities vary between the two captures. If the prior on N is again a Poisson distribution $\mathcal{P}o(\lambda)$, the posterior distribution of $(N - n.)$ is

$$\mathcal{P}o\left(\left(\frac{\beta}{\alpha + \beta} \right)^2 \lambda \right).$$

Here the posterior distribution is directly available and there is no need to call for Bayesian sampling. Note also that the marginal distribution of $n.$ given N is constant when $\alpha = \beta$.

In the case when the probability is the same for each capture and $p_i \sim \mathcal{B}e(\alpha, \beta)$, the marginal distribution of Δ given N is

$$
\begin{aligned}
f(\Delta | N) &= \prod_{i=1}^{N} \int_{0}^{1} 2^{\delta_i (2 - \delta_i)} \frac{(1 - p_i)^{\beta + 1 - \delta_i} p_i^{\alpha + \delta_i - 1} dp_i}{B(\alpha, \beta)} \\
&= \prod_{i=1}^{N} \frac{B(\alpha + \delta_i, \beta + 2 - \delta_i)}{B(\alpha, \beta)},
\end{aligned}
$$

8

where $\delta_i = \delta_{i1} + \delta_{i2}$. Therefore, the distribution of (n_{11}, n_{12}, n_{21}) given N is

$$f(n_{11}, n_{12}, n_{21}|N) = \binom{N}{n_{11}\ n_{12}\ n_{21}} \frac{(2\alpha\beta)^{n_{12}+n_{21}} (\alpha(\alpha+1))^{n_{11}} (\beta(\beta+1))^{2N-n_+}}{[(\alpha+\beta)(\alpha+\beta+1)]^{2N}}$$

In this case, the posterior distribution of $(N - n_+)$ is

$$\mathcal{P}o\left(\left(\frac{\beta(\beta+1)}{(\alpha+\beta)(\alpha+\beta+1)}\right)^2 \lambda\right),$$

thus depends only on the 'effective' size of the double sample, n_+. If $\alpha = \beta = 1$, the density of (n_{11}, n_{12}, n_{21}) is again constant.

These two results show that a totally heterogeneous model actually brings very little information on the population size N since too many parameters have to be controlled at the same time. In addition, the fact that the probabilities are all different calls for many recaptures, as it is the case in Huggins (1988). This is also an example of a situation where Bayesian sampling is of no use, since the marginal posterior distribution of N is much more straightforward than the conditional posterior.

5.2. Known strata. A more amenable case of heterogeneity occurs when the population is divided into known strata such as *male/female*, *young/adult/senior*, etc... This type of division usually occurs in surveys and censuses. We can then use the approach of Section 3 according to two scenarios:

(i) The strata are independent and each stratum size N_s ($s = 1, \ldots, S$) has a prior distribution $\mathcal{P}o(\lambda_s)$. We have then S independent replications of the model considered in Section 3. Bayesian sampling provides an easy approximation of the posterior distributions of $N_1, \ldots, N_S$ and thus of the posterior distribution of $N = N_1 + \ldots + N_S$.

(ii) The population size N has a prior distribution $\mathcal{P}o(\lambda)$ and the strata sizes $(N_1, \ldots, N_S)$ follows a multinomial distribution $\mathcal{M}_S(N; \omega_1, \ldots, \omega_S)$. This assumption is often justified in the case of surveys where the proportions of the subpopulations are rather well-known. However, when N is integrated out, this model appears as a special case of (i), since

$$N_s \sim \mathcal{P}o(\lambda\omega_s).$$

In both cases, we can see that Bayesian sampling allows for an easy computation of the strata sizes estimators.

5.3. Unknown strata. It may also happen that the population is divided into S strata which are impossible to detect, even after individuals have been captured. For instance, this occurs when a part of the population has an undetected disease (e.g., seropositivity) which modifies its behaviour. Therefore, the probability of capture of a given member of the population, p_i, is one of the probabilities $\pi_1, \ldots, \pi_S$ for the different strata, with probability ω_s ($1 \leq s \leq S$) corresponding to the proportion of the stratum in the whole population. We can model the probabilities π_s as beta distributions, $\mathcal{B}e(\alpha_s, \beta_s)$ ($1 \leq s \leq S$). For any prior distribution on N, it is then easy to see that $\pi(N|\text{data})$ cannot be used in practice,

even though it may be expressed in a closed form for some priors. This is often the case with Bayesian analyses of mixture models, since all the partitions of the sample have to be considered (see Diebolt and Robert (1990)).

Consider a Poisson distribution on N associated with the previous model. We then introduce an auxiliary parameter to implement the Bayesian sampling approach. Let z_{is} $(1 \leq i \leq N,\ 1 \leq s \leq S)$ denote the indicator function $\mathbf{I}\{p_i = \pi_s\}$. Therefore,

$$(z_{i1}, \ldots, z_{iS}) \sim \mathcal{M}_S(1; \omega_1, \ldots, \omega_S)$$

and, conditionally on $z_i = (z_{i1}, \ldots, z_{iS})$,

$$p_i = \prod_{s=1}^{S} \pi_s^{z_{is}}$$

(i.e. z_i completely determines p_i). It is easy to see that the simulation steps for Bayesian sampling are then

1. $\pi_s \sim \pi(\pi_s | N, z, \text{data})$ $(s = 1, \ldots, S)$
2. $z_i \sim \pi(z_i | N, \pi, \text{data})$ $(i = 1, \ldots, N)$
3. $N \sim \pi(N | \pi, z, \text{data})$,

where

$$\pi(\pi_s | N, z, \text{data}) \equiv \mathcal{B}e(\alpha_s + m_s, \beta_s + 2n_s - m_s),$$

$$\pi(z_i | N, \pi, \text{data}) \equiv \mathcal{M}_S\left(1; \left(\frac{\pi_s^{\delta_i}(1 - \pi_s)^{(2-\delta_i)}\omega_s}{\sum_{j=1}^{S} \pi_j^{\delta_i}(1 - \pi_j)^{(2-\delta_i)}\omega_j}\right)_s\right)$$

$$\pi(N | \pi, z, \text{data}) \equiv \mathcal{P}o\left(\lambda(1 - \pi_1)^{\tau_1} \ldots (1 - \pi_S)^{\tau_S}\right)$$

and

$$m_s = \sum_{i=1}^{N} z_{is}\delta_i, \quad n_s = \sum_{i=1}^{N} z_{is}, \quad \tau_s = n_s/N.$$

The variables n_s, m_s and τ_s are updated at step 2.

This particular case provides a strong argument in favor of Bayesian sampling since it appears to be the only way to study this intricated setup. A non-Bayesian approach cannot handle all the parameters and a formal Bayesian approach requires an enormous amount of computing time.

6. A multiple recapture example.

In this section, we briefly illustrate the extension of our techniques to the multiple recapture setup by application to a real data set. The data set we consider is from Castledine (1981) and appears in Table 1. It consists of 14 capture events from a population of sunfish. At the ith capture, n_i fishes are captured, out of which m_i have been previously captured. Thus, $n_+ = \sum_i(n_i - m_i) = 138$ is the total number of different fish captured.

i	1	2	3	4	5	6	7
n_i	10	27	17	7	1	5	6
m_i	0	0	0	0	0	0	2

i	8	9	10	11	12	13	14
n_i	15	9	18	16	5	7	19
m_i	1	5	5	4	2	2	3

Table 1. *Multiple recapture data for a population of sunfish.*

Following Castledine (1981), we consider a temporal model M_t with $s = 14$ capture episodes. We put the same prior $Be(\alpha, \beta)$ on each of the capture probabilities $p_1, \ldots, p_s$ and use the prior $\pi(N) \propto 1/N$. Analogous to the developments in Sections 2 and 3, it is straightforward to show that the conditional distributions are

$$\pi(N|p_1, \ldots, p_s, \text{data}) \equiv Neg(n_+ - 1, \mu), \tag{6.1}$$

with $\mu = 1 - (1 - p_1) \ldots (1 - p_s)$, and $(i = 1, \ldots, s)$

$$\pi(p_i|N, \text{data}) \equiv Be(\alpha + n_i, \beta + N - n_i). \tag{6.2}$$

We employed Bayesian sampling to obtain the posterior quantities listed in Table 2 and the posterior histograms in Figure 1. This was done as follows. For each of seven choices of Beta parameters (α, β), we simulated a random sample of size $n = 1000$ from the marginal posterior as in (1.3). We iteratively sampled as in (1.5) from (6.1) and (6.2) a total of fifty times to obtain each sampled observation. (The simulations were performed with the IMSL routines RNNBN and DRNBET.) The mle for N, here 460, was used as the starting value N_0^*. Finally, the posterior quantities were estimated by their sample moments.

Prior parameters α β	Posterior of $N^{(1)}$ Mean St.Dev.	Posterior of $\log(N)^{(1)}$ Mean St.Dev.	Posterior of $\log(N)^{(2)}$ Mean St.Dev.	95% credible interval for N [1]	95% credible interval for N [2]
0 1	448 84	6.09 0.18	6.15 0.18	$312 - 650$	$322 - 656$
2 100	506 70	6.21 0.14	6.21 0.14	$378 - 663$	$393 - 676$
3 100	419 49	6.03 0.12	6.05 0.12	$332 - 520$	$335 - 539$
10 500	548 54	6.30 0.10	6.31 0.10	$454 - 664$	$454 - 671$
15 500	408 37	6.01 0.09	6.02 0.09	$342 - 485$	$347 - 489$
20 1000	548 49	6.32 0.09	6.33 0.09	$471 - 665$	$470 - 666$
30 1000	406 32	6.01 0.08	6.01 0.08	$348 - 475$	$350 - 478$

Table 2. *Comparison between Bayesian sampling approximation* [1] *and the normal approximation in Castledine (1981)* [2].

Figure 1 should appear around here

We have also included in Table 2 the estimates of Castledine (1981) who performed the same analysis, but used a normal approximation to the posterior distribution of $\log(N)$, to obtain the required quantities. The agreement between our numbers and his shows that Castledine's approximations performed remarkably well in most of the cases. Notice that the greatest disagreement between the two approaches occurs for small values of α and β, especially for $\alpha = 0$, $\beta = 1$, when prior information is weakest. One can see in Figure 1 that, in this case, the skewness of the posterior distribution of N is not entirely removed by the log transformation so that Castledine approximation will be off, especially for computing the interval estimates. An important justification of Bayesian sampling techniques is that they can circumvent the problems of inaccurate normal approximations, as pointed out in Tanner and Wong (1987). Finally, note that, in this analysis, the estimates of N seem to be sensitive to the choice of the Beta prior. Thus, it is probably most reasonable to use the estimates for the choice $(\alpha, \beta) = (0, 1)$ which, except for the slightly different prior on N, yields the hypergeometric case discussed in Section 4.

7. Conclusion

This paper has shown that in a large variety of capture-recapture models, Bayesian sampling can provide a fast and efficient approach to obtaining posterior information in Bayes analyses. The particular models we considered were meant to illustrate the general ideas for implementing Bayesian sampling in this context. Of course, our treatment was by no means exhaustive, as there are many other variants and generalisations where Bayesian sampling should also be fruitful. For instance, it would be interesting to apply this approach to the analysis of open population extensions which take into account the deaths and immigrations which actually occur in wildlife populations. Another variation, studied by Seber and Felton (1981), is to consider *tag-loss*, namely the misclassification of recaptured objects as newly captured objects. This last variant may also be used in epidemiology when the proportion of people in each stratum varies between the captures. Another possible extension deals with the estimation of the number of *species*, as considered in Efron and Thisted (1976). The key to the availability of Bayesian sampling methods in all of these extensions is the availability of conditional distributions for which easy computing is practically possible.

Bayesian sampling is not a new inferential procedure but rather an approach for facilitating Bayesian inference, i.e. a powerful tool at the end of the Bayesian processing chain. If a different inferential approach is desired, then Bayesian sampling will be irrelevant. The main criticism of Bayesian inference is its requirement of prior input. If such prior information is available, and it will be in some analyses, we believe a direct Bayesian approach is reasonable. However, when little prior information is available, we would recommend robust Bayesian methods. It turns out that, in capture-recapture settings, such robust approaches are indeed possible. These can be obtained by the hierarchical and empirical Bayesian approaches discussed in Sections 3 and 4. Fortunately, powerful Bayesian sampling methods are also readily available in these cases.

12

References

Bishop, Y., Fienberg, S. and Holland, P. (1975) *Discrete Multivariate Analysis, Theory and Practice*. Cambridge, Mass.: MIT Press.

Burnham, K., Anderson, D., White, G., Brownie, C. and Pollock, K. (1986) *Design and analysis methods for fish survival experiments based on release-recapture*. Bethesda, Maryland: American Fisheries Society Monograph.

Burnham, K. and Overton, W. (1978) Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**(3), 625-33.

Castledine, B. (1981) A Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika* **67**(1), 197-210.

Darroch, J. (1958) The multiple-recapture census. I: Estimation of a closed population. *Biometrika* **45**, 343-59.

Diebolt, J. and Robert, C. (1990) Bayesian estimation of finite mixture distributions (II): Sampling implementation. Tech. Report # 111, Université Paris VI.

Efron, B. and Thisted, R. (1976) Estimating the number of species: How many words did Shakespeare know? *Biometrika* **63**(3), 435-47.

Gelfand, A. and Smith, A. (1990) Sampling based approaches to calculating marginal densities. *J.American Statistical Association*, **85**, 398-409.

Huggins, R. (1989) On the statistical analysis of capture-recapture experiments. *Biometrika* **76**(1), 133-40.

Jewell, W. (1985) Bayesian estimation of undetected errors. In *Bayesian Statistics 2* (J. Bernardo, M. DeGroot, D. Lindley and A. Smith, eds.) North-Holland.

Marks, E., Seltzer, W. and Krotki, K. (1974) *Population Growth Estimation*, New York: Population Council.

Nayak, T. (1988) Estimating population size by capture-recapture sampling. *Biometrika* **75**(1), 113-20.

Pickands, J. and Raghavachari, M. (1987) Exact and asymptotic inference for the size of a population. *Biometrika* **74**(2), 355-63.

Polya, G. (1976) Probabilities in proofreadings. *Am. Math. Monthly*, **83**, 42.

Raftery, A. (1988) Inference for the binomial N parameter : a hierarchical Bayes approach. *Biometrika* **75**(2), 223-28.

Robert, C. (1990) Hidden mixture and Bayesian sampling. Tech. Report # 114, Université Paris VI.

Seber, G. (1982) Capture-recapture methods. In *Encyclopedia of Statistical Sciences* (S. Kotz and N. Johnson, eds.). J. Wiley, New York

Seber, G. and Felton, R. (1981) Tag loss and the Petersen mark-recapture experiment. *Biometrika* **68**(1), 211-9.

Smith, P. (1988) Bayesian methods for multiple capture-recapture surveys. *Biometrics* **44**, 1177-89.

Tanner, M. and Wong, W. (1987) The calculation of posterior distributions by data augmentation. *J. American Statistical Association*, 81, 82-6.

Wolter, K. (1986) Some coverage error models for census data. *J. American Statistical Association* 81, 338-46.

Graduate School of Business
University of Chicago
1101, East 58th Street
Chicago, IL 60637

L.S.T.A. - I.S.U.P.
Université Paris VI
Tour 45-55, 4 place Jussieu
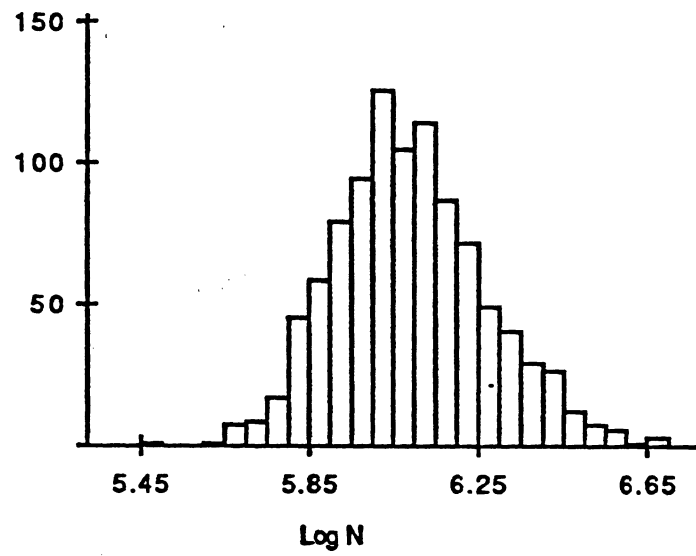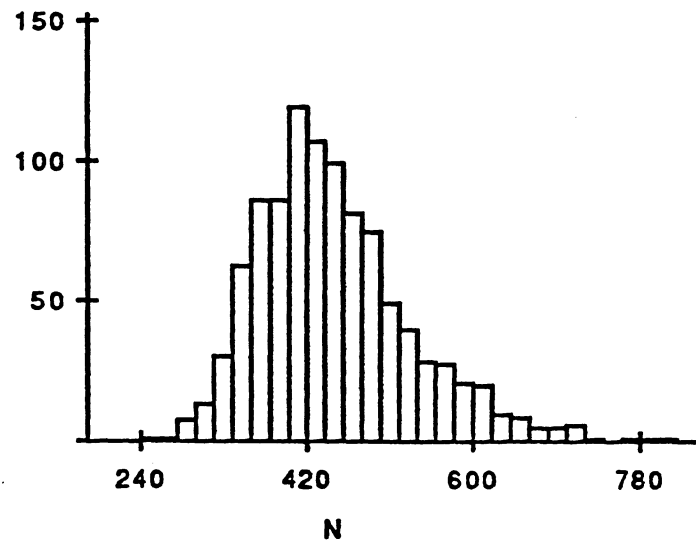*F-75252 Paris Cedex 05*

Figure 1 - Posterior histograms for N and $\log(N)$ and $(\alpha, \beta) = (0, 1)$.