CALCULATING BAYES ESTIMATES FOR CAPTURE–RECAPTURE MODELS

by

Edward I. George and Christian P. Robert
University of Chicago and Université Paris VI

Technical Report #90-36C

Department of Statistics
Purdue University

July 1990
Revised August 1991

*

CALCULATING BAYES ESTIMATES FOR CAPTURE-RECAPTURE MODELS[1]

Edward I. George, University of Chicago

Christian P. Robert, Université Paris VI

July 1990, Revised July 1991

Abstract

Capture-recapture models are widely used in the estimation of population sizes. Based on data augmentation considerations, we show how Gibbs sampling can be applied to calculate Bayes estimates in this setting. As a result, formulations which were previously avoided because of analytical and numerical intractability, can now be considered for practical application. We illustrate this potential by using Gibbs sampling to calculate Bayes estimates for a hierarchical capture-recapture model in a real example.

Keywords: Data augmentation; Estimation of population size; Gibbs sampling; Hierarchical models; Log concavity; Multinomial model; Multiple-recapture sampling.

[1]This report was formerly titled "Capture-Recapture Models and Bayesian Sampling". The authors would like to thank George Casella and Martin Wells for helpful discussions, and a referee for pointing out the recursive relation (2.9). This research was partially supported by the U.S. Army Research Office through the Mathematical Science Institute at Cornell University, and by the Graduate School of Business at the University of Chicago.

1. Introduction

A common experimental setup for estimating 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 (those that appeared in more than one sample). First used by Laplace (1786) to estimate the population size of France, this approach received its main impetus in the context of estimating the size of wildlife populations, where it became known as the capture-recapture methodology (Otis et al. 1978, Seber 1982, and Pollock 1991). 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, Seltzer and Kroti 1974). A recent application which has received much attention is the estimation of coverage error in surveys and censuses (Wolter 1986).

A general formulation of the underlying capture-recapture experiment is as follows. Let N be the unknown size of the population of interest, and let I be the number of samples taken. The probability that individual (j) is captured in sample (i) is given by p_{ij}. Assuming that all captures are independent, the likelihood of N and $p = (p_{11},\ldots,p_{sN})$ is

$$L(N,p \mid \text{data}) = \prod_{i=1}^{I}\prod_{j=1}^{N} p_{ij}^{\delta_{ij}} (1 - p_{ij})^{1-\delta_{ij}} \tag{1.1}$$

where $\delta_{ij} = 1$ or 0 according to whether or not individual j is captured in sample i. Typically, the parameter space of the general model (1.1) is restricted so that information about N can be extracted from the data. For example, the commonly applied restriction $p_{ij} \equiv p_i$ pertains to experiments where the probability of capture is identical across individuals within each sample. This case is the focus of Sections 2, 3 and 4. Models based on other restrictions are discussed in Section 5.

Although a variety of frequentist and likelihood approaches for making inference about N have appeared in the literature (Bishop, Fienberg, and Holland 1975, Burnham et al. 1986, Pickands and Raghavachari 1989, and Huggins 1989), we shall consider only the Bayesian approach (Castledine 1981, Jewell 1985, Smith 1988, and Leite and Pereira 1990). In the capture-recapture setting, the Bayesian approach to extracting information about N proceeds as follows. For a particular setup, the joint posterior of N and p can be obtained from the likelihood $L(N,p \mid \text{data})$ and the (possibly improper) prior $\pi(N,p)$,

$$(1.2) \qquad \pi(N,p \mid \text{data}) \propto L(N,p \mid \text{data})\, \pi(N,p).$$

The marginal posterior of N is then in principle obtained as

$$(1.3) \qquad \pi(N \mid \text{data}) = \int \pi(N,p \mid \text{data})\, dp.$$

A deficiency of this approach is the potential difficulty of calculating these marginals or even just the posterior quantities of interest such as the (formal) posterior mean $E(N \mid \text{data})$. Unfortunately, such difficulties are sometimes overcome by making additional assumptions which may not be justified.

The purpose of this paper is to show how Gibbs sampling, as an alternative to both analytical calculation and numerical approximation, greatly enhances the potential for Bayesian analyses of capture-recapture models. Effectively, the Gibbs sampler allows us to approximate the marginal posterior of N, namely $\pi(N \mid \text{data})$, by simulated sampling from the conditional distributions

$$(1.4) \qquad \pi(N \mid p, \text{data}) \quad \text{and} \quad \pi(p \mid N, \text{data}),$$

see Gelfand and Smith (1990) and Casella and George (1991). Starting with an initial value for N, say $N^{(0)}$, the Gibbs sampler produces an auxiliary "Gibbs sequence"

$$(1.5) \qquad N^{(0)}, p^{(0)}, N^{(1)}, p^{(1)}, \ldots,$$

by alternately sampling from

$$(1.6) \qquad N^{(k)} \sim \pi(N \mid p^{(k-1)}, \text{data}) \quad \text{and} \quad p^{(k)} \sim \pi(p \mid N^{(k)}, \text{data}).$$

Under weak conditions, the distribution of $N^{(k)}$ converges to $\pi(N \mid \text{data})$ as $k \to \infty$ (see

Diebolt and Robert (1990) and Schervish and Carlin (1990)). By simulating a long sequence, this property can then be exploited to use aspects of the Gibbs sequence (1.5) to approximate $\pi(N \mid \text{data})$. For example, $\overline{N} = \frac{1}{K}\sum_{k=1}^{K} N^{(k)}$ would be a consistent estimator of $E(N \mid \text{data})$. An even better estimator of $E(N \mid \text{data})$, as argued in Gelfand and Smith (1990) and Liu, Wong and Kong (1991), is the average of the conditional expectations, namely $\frac{1}{K}\sum_{k=1}^{K} E(N \mid p^{(k)})$. Similarly one could estimate the posterior density $\pi(N \mid \text{data})$ by $\frac{1}{K}\sum_{k=1}^{K} \pi(N \mid p^{(k)})$.

A major reason for the successful implementation of the Gibbs sampler in these capture-recapture setups is the simplification provided by the conditioning in (1.4). This is because the general model (1.1) is multinomial, the analysis of which is well understood when either the population size N or the probabilities p_{ij} are known. Note that the traditional Bayesian calculation of integrating out the p_{ij}, which often leads to intractable expressions, does not exploit this special structure. Our approach of introducing additional structure which treats the p_{ij} as missing values is fundamentally the idea behind data augmentation (Tanner and Wong 1987). As will be seen in the sequel, many priors yield conditional posteriors (1.4) which are well-known and can be efficiently simulated using standard routines in packages such as IMSL, NAG and S. Furthermore, a large class of priors leads to conditional posteriors which turn out to be log concave, and so can be efficiently simulated using the adaptive rejection sampling algorithms of Gilks and Wild (1991).

The plan of this paper is as follows. In Section 2, we show how Gibbs sampling leads to manageable Bayes calculations for the special case of (1.1) where $p_{ij} \equiv p_i$. In Section 3, we show how these methods may also be applied to natural hierarchical extensions of these models. In Section 4, the Gibbs sampler is used to compute hierarchical Bayes estimates for the sunfish data example of Castledine (1981). In Section 5, we show how Gibbs sampling may be applied to various other capture-recapture models including a hierarchical model for adaptive stratification.

2. Bayes Calculations for the Homogeneous Catch Model

In this section we consider the following multiple recapture experiment which has been studied extensively in the literature (Darroch 1958, Castledine 1981, and Seber 1982). From a closed population of unknown size N, I samples of sizes $n_1,\ldots n_I$ are consecutively drawn, marked, and returned to the population. The total number of distinct captured individuals, denoted r, is recorded. Note that $r = \sum_{i=1}^{I} n_i - m$, where m is the total number of times a recaptured individual is observed. (This notation will be used throughout the paper.) If $n_1,\ldots n_I$ are treated as random, the likelihood for this experiment may be obtained as the special case of (1.1) when the probability of any capture in the ith sample is $p_{ij} \equiv p_i$, namely

$$(2.1) \qquad L(N,p \mid \text{data}) \propto \frac{N!}{(N-r)!} \prod_{i=1}^{I} p_i^{n_i} (1-p_i)^{N-n_i}$$

where henceforth $p = (p_1,\ldots,p_I)$. An alternative perspective of this situation which has also been studied treats $n_1,\ldots n_I$ as fixed so that the model is conditional on these sample sizes. In this case the likelihood is obtained from the hypergeometric model as

$$(2.2) \qquad L(N \mid \text{data}) \propto \binom{N}{r} \Big/ \prod_{i=1}^{I} \binom{N}{n_i}.$$

When I = 2, both of these likelihoods correspond to the Lincoln-Petersen capture-recapture model, which dates back to Laplace (1786). Note that for I = 2, the maximum likelihood estimator of N in both (2.1) and (2.2) is

$$(2.3) \qquad \hat{N}_{MLE} = \left\lfloor \frac{n_1 n_2}{m} \right\rfloor$$

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

We now proceed to consider the problem of obtaining Bayesian estimates for the model (2.1) using priors under which N and p are independent, so that

$$\pi(N,p) = \pi(N)\pi(p) \tag{2.4}$$

This mild restriction which seems plausible in many applications, was also considered by Castledine (1981). It is easy to see that (2.4) when combined with (2.1) yields full conditionals of the form

$$\pi(N \mid p, \text{data}) \propto \frac{N!}{(N-r)!}\left(\prod_{i=1}^{I}(1-p_i)\right)^{N}\pi(N) \tag{2.5}$$

$$\pi(p \mid N, \text{data}) \propto \left\{\prod_{i=1}^{I} p_i^{n_i}(1-p_i)^{N-n_i}\right\}\pi(p) \tag{2.6}$$

Both (2.5) and (2.6) turn out to be especially tractable.

For example, suppose $\pi(p) = \prod_{i=1}^{I}\pi(p_i)$, where $\pi(p_i) = Be(a,b)$, the Beta distribution

$$\pi(p_i) \propto p_i^{a-1}(1-p_i)^{b-1} \tag{2.7}$$

with known hyperparameters $a > 0$ and $b > 0$. Then as shown in Castledine (1981),

$$\pi(N \mid \text{data}) = \frac{N!}{(N-r)!}\left(\prod_{i=1}^{I}\frac{\Gamma(N-n_i+b)}{\Gamma(N+a+b)}\right)\pi(N). \tag{2.8}$$

For this case,

$$\frac{\pi(N+1 \mid \text{data})}{\pi(N \mid \text{data})} = \frac{(N+1)}{(N-r)}\left(\prod_{i=1}^{I}\frac{N-n_i+\beta}{N+\alpha+\beta}\right)\frac{\pi(N+1)}{\pi(N)}, \tag{2.9}$$

so that $\pi(N \mid \text{data})$ in (2.8) can be computed exactly by recursion whenever $\pi(N+1)/\pi(N)$

is readily available such as when $\pi(N) = Po(\lambda)$ or $\pi(N) \equiv$ constant. This then allows for at least straightforward numerical approximation of

$$\pi(p \mid \text{data}) = \sum_N \pi(p \mid N, \text{data})\, \pi(N \mid \text{data}), \tag{2.10}$$

if this quantity is of interest. Although these calculations are inexpensive by today's computing standards, it is interesting that Castledine (1981) used a normal approximation rather than (2.9) to compute $\pi(N \mid \text{data})$. Also, note that integrating out p_i with respect to (2.7) for the choice $(a,b) = (0,1)$ yields the hypergeometric model (2.2).

Of course, not all prior formulations allow as clean a solution as (2.9). We now proceed to show that in those cases where $\pi(N \mid \text{data})$ and/or $\pi(p \mid \text{data})$ cannot be conveniently obtained by analytical or numerical methods, Monte Carlo approximation using the Gibbs sampler may be a promising approach. Starting with an initial value $N^{(0)}$ for N, such as $\hat{N}_{MLE}$, the Gibbs sampler simulates a Gibbs sequence

$$N^{(0)}, p^{(0)}, N^{(1)}, p^{(1)}, N^{(2)}, p^{(2)}, \ldots \tag{2.11}$$

by alternately sampling from

$$N^{(k)} \sim \pi(N \mid p^{(k-1)}, \text{data}) \quad \text{and} \quad p^{(k)} \sim \pi(p \mid N^{(k)}, \text{data}). \tag{2.12}$$

It follows from the main results of Schervish and Carlin (1990) that when the support of $\pi(p)$ is $[0,1]$, the distribution of $p^{(k)}$ will converge to $\pi(p \mid \text{data})$. It then follows from the duality principle of Diebolt and Robert (1990) that the distribution of $N^{(k)}$ will converge to $\pi(N \mid \text{data})$. Thus, for such $\pi(p)$, characteristics of these marginal posteriors of N and p can be approximated from the sequence (2.11).

In order to use the Gibbs sampler effectively here, it is only necessary to be able to simulate N and p easily and efficiently from the conditional distributions in (2.5) and (2.6). Fortunately, these are often standard distributions. For example, in the simulation of N, if $\pi(N) = Po(\lambda)$, the Poisson distribution with mean λ, then (2.5) can be replaced by

$$\text{(2.13)} \qquad \pi(N - r \mid p, \text{data}) = \text{Po}\Big(\lambda \, \big(\textstyle\prod_{i=1}^{I}(1 - p_i)\big)\Big)$$

the Poisson distribution with mean $\lambda\left(\prod_{i=1}^{I}(1 - p_i)\right)$. Alternatively, one might consider the improper prior $\pi(N) \equiv 1$, in which case

$$\text{(2.14)} \qquad \pi(N \mid p, \text{data}) = \text{NB}\Big(r,\ 1 - \textstyle\prod_{i=1}^{I}(1 - p_i)\Big),$$

the negative binomial distribution with mean $r \big/ \left[1 - \prod_{i=1}^{I}(1 - p_i)\right]$. Both of these distributions can be readily simulated.

For the simulation of p, one might consider the general adaptive rejection scheme of Gilks and Wild (1991) which requires only that $\pi(p \mid N, \text{data})$ be log concave in each p_i, (i.e. $\partial \log\pi(p \mid N, \text{data})/\partial p_i$ is nonincreasing for each i). From (2.6), this will be satisfied whenever $\pi(p_i)$ is log concave. One might also consider a prior on a natural transformation of p, such as the normal prior on the logits $\alpha_i = \log\{p_i/(1 - p_i)\}$, $i = 1,\ldots,I$, where $\pi(p)$ is such that

$$\text{(2.15)} \qquad \alpha_1,\ldots,\alpha_I \text{ iid} \sim N(\mu,\sigma^2).$$

In this case, the full conditional $\pi(p \mid N, \text{data})$ is obtained from $\pi(\alpha_1,\ldots,\alpha_I \mid N, \text{data}) = \prod_i \pi(\alpha_i \mid N, \text{data})$ where

$$\text{(2.16)} \qquad \pi(\alpha_i \mid N, \text{data}) \propto \frac{e^{\alpha_i n_i - (\alpha_i - \mu)^2/2\sigma^2}}{(1 + e^{\alpha_i})^N}\ .$$

These α_i can also be simulated using the general adaptive rejection methods since $\pi(\alpha_i \mid N, \text{data})$ is log concave in α_i.

3. Hierarchical Bayes Extensions

In this section, we consider the special case of (2.4) where N and p are a priori independent, and $p_1,\ldots,p_I$ are a priori exchangeable. Such priors will then be of the form

$$\pi(N, p \mid \theta)\pi(\theta) = \pi(N)\left\{\prod_{i=1}^{I} \pi(p_i \mid \theta)\right\}\pi(\theta), \tag{3.1}$$

where θ is a hyperparameter governing each prior $\pi(p_i \mid \theta)$. As opposed to the priors in Section 2, this prior formulation, also considered by Castledine (1981), allows for greater flexibility in the specification of the prior $\pi(p)$. Note that (3.1) reduces to the previous formulation when θ is known so that $\pi(\theta)$ is degenerate. Another possible simplification occurs when p can be eliminated from the above by obtaining $\pi(\theta \mid N, \text{data})$ and $\pi(N \mid \theta, \text{data})$. Our focus here is on those cases where such simplifications are not readily available.

Combined with the likelihood (2.1), the full conditional posteriors for N, p and θ are

$$\pi(N \mid p, \theta, \text{data}) = \pi(N \mid p, \text{data}) \propto \frac{N!}{(N-r)!}\left(\prod_{i=1}^{I} (1-p_i)\right)^{N} \pi(N) \tag{3.2}$$

$$\pi(p \mid N, \theta, \text{data}) \propto \prod_{i=1}^{I} p_i^{n_i} (1-p_i)^{N-n_i} \pi(p_i \mid \theta) \tag{3.3}$$

$$\pi(\theta \mid N, p, \text{data}) = \pi(\theta \mid p) \propto \left\{\prod_{i=1}^{I} \pi(p_i \mid \theta)\right\} \pi(\theta) \tag{3.4}$$

Note that $\pi(N \mid p, \theta, \text{data})$ does not depend on θ, and $\pi(\theta \mid N, p, \text{data})$ does not depend on N or the data. This feature of hierarchical models makes them particularly well suited for simulation methods such as Gibbs sampling, see Morris (1987).

As in Section 2, in those cases where the marginal posteriors of N, p or θ cannot be conveniently obtained by analytical or numerical methods, the Gibbs sampler provides an alternative. Here the Gibbs sequence is of the form

$$\theta^{(0)},N^{(0)},p^{(0)},\theta^{(1)},N^{(1)},p^{(1)},\theta^{(2)},N^{(2)},p^{(2)},\ldots \tag{3.5}$$

where except for the preselected initial values $\theta^{(0)}$ and $N^{(0)}$, the sequence is obtained by alternately simulating from

(3.6) $\theta^{(k)} \sim \pi(\theta \mid p^{(k-1)}), \quad N^{(k)} \sim \pi(N \mid p^{(k-1)}, \text{data}), \quad p^{(k)} \sim \pi(p \mid N^{(k)}, \theta^{(k)}, \text{data}),$

By an argument similar to that used in Section 2, when the support of every $\pi(p_i \mid \theta)$ is [0,1], the distributions of $N^{(k)}$, $p^{(k)}$ and $\theta^{(k)}$ will converge to $\pi(N \mid \text{data})$, $\pi(p \mid \text{data})$ and $\pi(\theta \mid \text{data})$ respectively. Thus for such priors, characteristics of these marginal posteriors can be approximated from the sequence (3.5).

The Gibbs sampler is especially attractive here because of the ease with which the three conditionals in (3.6) can be simulated. Noting first that $\pi(N \mid p, \theta, \text{data})$ in (3.2) is of the same form as $\pi(N \mid p, \text{data})$ in (2.5), the same considerations discussed there apply here. For example, the choices $\pi(N) = Po(\lambda)$ and $\pi(N) \equiv$ constant yield the familiar conditionals given in (2.13) and (2.14), respectively.

A natural candidate for this approach to let each $\pi(p_i \mid \theta) = Be(a,b)$, the Beta distribution

(3.7) $$\pi(p_i \mid a,b) \propto p_i^{a-1}(1-p_i)^{b-1}$$

as in (2.7). However, unlike (2.7), $\theta = (a,b)$ is here treated as unknown. In this case (3.3) becomes

(3.8) $$\pi(p \mid N, a,b, \text{data}) \propto \prod_{i=1}^{I} p_i^{n_i+a-1}(1-p_i)^{N-n_i+b-1}.$$

Each p_i is conditionally Beta with parameters $(n_i + a)$ and $(N - n_i + b)$. Fast methods for simulating this well-known distribution are readily available.

For the Beta prior (3.7), the conditional distribution (3.4) is of the form

(3.9) $$\pi(a,b \mid N, p, \text{data}) \propto \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]^I \left[\prod_{i=1}^{I} p_i\right]^a \left[\prod_{i=1}^{I}(1-p_i)\right]^b \pi(a,b).$$

The first three terms on the right hand side of (3.9) are log-concave in a and b. (The log-concavity of the first term follows from the well-known properties of the digamma function $d\log\Gamma(x)/dx$.) Thus, $\pi(a,b \mid N, p, \text{data})$ will be log-concave whenever $\pi(a,b)$ is log-concave. In such cases, (a,b) can be simulated from (3.9) sequentially via $\pi(a \mid b, N, p, \text{data})$ and $\pi(b \mid a, N, p, \text{data})$ using the adaptive rejection routine of Gilks and Wild (1991). For example, a prior such as $\pi(a,b) \propto \exp[-c(a+b)]$ on $a,b > 0$ for $c > 0$ will work. Furthermore, it can be shown that for $\pi(a,b) \equiv$ constant on $a,b > 0$, (3.9) will be a proper posterior when $I \geq 2$. This improper prior might be considered non-informative.

It may be of interest to note that the efficiency of the Gibbs sampler does not appear to benefit from any analytical or numerical simplification of (3.2)-(3.4). For example, using (2.8) and (2.9) one could compute $\pi(N \mid \theta, \text{data})$ and use this to simulate the $N^{(i)}$'s in (3.5). Although this would provide faster convergence of (3.5), it is much more costly in computation requirements. Furthermore, it does not appear that $\pi(\theta \mid N, \text{data})$ can be obtained, making simulation of the $p^{(i)}$'s from (3.3) necessary.

Although Castledine (1981) considered the hierarchical prior (3.1), he avoided the Beta prior (3.7) calling it intractable, and instead obtained approximations for a hierarchical generalization of (2.15), the normal on the logits $\alpha_i = \log\{p_i/(1-p_i)\}$, $i = 1,\ldots,I$. He coupled the model

(3.10) $$\pi(\alpha_1,\ldots,\alpha_I \mid \mu) = \text{iid} \sim N(\mu, \sigma^2)$$

with

(3.11) $$\pi(\mu) = N(\eta,\tau^2).$$

(He treated σ^2, η and τ^2 as known). Here too, the Gibbs sampler is easily implemented. As before, the conditional $\pi(p \mid N, \mu, \text{data})$ is obtained from $\pi(\alpha_1,\ldots,\alpha_I \mid N, \mu, \text{data}) = \prod_i \pi(\alpha_i \mid N, \mu, \text{data})$ where

(3.12) $$\pi(\alpha_i \mid N, \mu, \text{data}) \propto \frac{e^{\alpha_i n_i - (\alpha_i - \mu)^2/2\sigma^2}}{(1+e^{\alpha_i})^N}$$

is log-concave and so can also be simulated using Gilks and Wild (1991). Again the conditionals $\pi(N \mid p, \mu, \text{data})$ are easily simulated since p is a function of $\alpha_1,\ldots,\alpha_I$.

Finally, the conditional

$$\pi(\mu \mid N, p, \text{data}) = \pi(\mu \mid p) = N\left(\frac{\tau^2\bar{\alpha} + (\sigma^2/I)\eta}{\tau^2 + (\sigma^2/I)}, \frac{\tau^2(\sigma^2/I)}{\tau^2 + (\sigma^2/I)}\right) \tag{3.13}$$

where $\bar{\alpha} = \frac{1}{I}\sum_{i=1}^{I}\alpha_i$, is easily simulated. Generalizing this approach to perform the calculations for further elaborations of this model such as putting priors on σ^2, η and/or τ^2 is straightforward.

4. The Sunfish 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, shown in Table 1, is from Seber (1973, p.143) and was also analyzed by Castledine (1981). It consists of 14 capture events from a population of sunfish. At the ith capture, n_i fish are captured, out of which m_i have been previously captured. Thus, $r = \sum_{i=1}^{I}(n_i - m_i) = 138$ is the total number of different fish captured.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n_i	10	27	17	7	1	5	6	15	9	18	16	5	7	19
m_i	0	0	0	0	0	0	2	1	5	5	4	2	2	3

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

For this data, Castledine (1981) considered the homogeneous catch model (2.1) with $I = 14$ capture events. Assuming prior independence, i.e. $\pi(N,p) = \pi(N)\pi(p)$, he considered the prior $\pi(p \mid a,b) = \prod_{i=1}^{I} \pi(p_i \mid a,b)$ where $\pi(p_i \mid a,b)$ is the beta distribution Be(a,b) in (2.7) and (3.7), and $\pi(N) \propto 1/N$ is the improper prior. In Table 2 below, we present characteristics of $\pi(N \mid \text{data})$ for his choices of (a,b). These were computed using (2.9) rather than with Castledine's normal approximations. Note that the estimates depend strongly on the choice of (a,b).

a	b	Mean of N	Std. Dev. of N	95% Credible Interval
0	1	446.1	81.4	319 – 636
2	100	506.9	70.5	389 – 664
3	100	418.8	51.2	332 – 532
10	500	547.4	54.9	450 – 665
15	500	408.9	35.8	345 – 485
20	1000	556.0	49.7	466 – 661
30	1000	406.8	32.3	348 – 475

Table 2. Posterior characteristics of N for fixed a and b.

Following Castledine (1981), we also considered the homogeneous catch model with the priors $\pi(p \mid a,b) = \prod_{i=1}^{I} \pi(p_i \mid a,b)$ and $\pi(N) \propto 1/N$. However, rather than focus on the posterior of N for various values of a and b as Castledine did, we treated a and b as unknown by using the hierarchical improper prior $\pi(a,b) \equiv 1$ on $a,b > 0$. Castledine (1981) mentioned that he did not pursue this hierarchical setup because of its intractability. As indicated in Section 3, this setup is tailor-made for calculation via the Gibbs sampler. This is accomplished by simulating the Gibbs sequence

$$(4.1) \qquad a^{(0)},b^{(0)},N^{(0)},p_1^{(0)},\ldots,p_I^{(0)},a^{(1)},b^{(1)},N^{(1)},p_1^{(1)},\ldots,p_I^{(1)},a^{(2)},b^{(2)},N^{(2)},\ldots$$

where we initialized $a^{(0)} = 1$, $b^{(0)} = 1$ and $N^{(0)} = \hat{N}_{MLE} = 460$, and then generated subsequent values from the following conditional distributions:

$$(4.2) \qquad \pi(p_i^{(k)} \mid N^{(k)}, a^{(k)},b^{(k)}, \text{data}) = Be(n_i + a^{(k)}, N^{(k)} - n_i + b^{(k)}),$$

$$(4.3) \qquad \pi(a^{(k+1)} \mid b^{(k)}, p^{(k)}, \text{data}) \propto \left[\frac{\Gamma(a+b^{(k)})}{\Gamma(a)}\right]^I \left[\prod_{i=1}^{I} p_i^{(k)}\right]^a ,$$

$$\text{(4.4)} \qquad \pi(b^{(k+1)} \mid a^{(k+1)}, p^{(k)}, \text{data}) \propto \left[\frac{\Gamma(a^{(k+1)}+b)}{\Gamma(b)}\right]^I \left[\prod_{i=1}^I (1-p_i)\right]^b,$$

$$\text{(4.5)} \qquad \pi(N^{(k+1)} \mid p^{(k)}, \text{data}) = NB\left(r-1, 1-\prod_{i=1}^I (1-p_i^{(k)})\right).$$

Samples from the distributions (4.2) and (4.5) were simulated using the IMSL routines DRNBET and RNNBN respectively. Samples from the distributions (4.3) and (4.4) were simulated using the adaptive rejection algorithm of Gilks and Wild (1991). Note that in addition to obtaining an approximation to the posterior of N, the Gibbs sequence automatically provides approximations to the posteriors of p and (a,b), features which may be valuable at least in assessing modeling considerations.

Figure 1 below presents histograms of a, b and N from a generated Gibbs sequence of length 10,000 (after 50 initializing iterations to lessen dependence on starting values). These histograms may be considered estimates of the marginal posteriors $\pi(a \mid \text{data})$, $\pi(b \mid \text{data})$ and $\pi(N \mid \text{data})$. In Table 3 below we present estimates of posterior distribution characteristics for a, b and N which were obtained simply from the sample mean, the sample standard deviation, and the 2.5% and 97.5% quantiles of the generated Gibbs sequence. Although it would require substantially more effort, one could obtain even better estimates of all these characteristics by using conditional quantities such as $\frac{1}{K}\sum_{k=1}^K \pi(N \mid p^{(k)})$ or $\frac{1}{K}\sum_{k=1}^K E(N \mid p^{(k)})$. For the sake of brevity we have not reported posterior estimates for $(p_1,\ldots,p_{14})$, however the estimates for a and b provide indirect information about their distribution.

To begin with, note that the posterior means of a and b do not correspond closely to any of the choices in Table 2. Moreover, it turns out that for the fixed choice (a,b) = (5.83, 233.5), the posterior $\pi(N \mid \text{data})$ has mean 463.2, standard deviation 50.25 and 95% credible interval 376-572. Although 463.2 agrees with our estimate 464.0, our posterior is much more widely dispersed, a consequence of treating a and b as random. Furthermore, our approach allows us to obtain information about the appropriate values of a and b, rather than arbitrarily imposing fixed values. This is important in practice because in such problems we are not apt to know much about the probabilities $p_1,\ldots,p_{14}$.

Comparing our posterior estimates of N with those in Table 2, we see that our estimate 464.0 falls in the middle of the range of the means, our estimate 89.4 is larger than all of the standard deviations, and our 95% credible interval (329, 679) contains all the other intervals except for the one corresponding to (a,b) = (0,1). Finally, it is interesting to note that our estimate is essentially the same as $\hat{N}_{MLE} = 460$. This is consistent with the idea that $\pi(a,b) \equiv 1$ is a non-informative prior. This contrasts with Castledine (1981) who suggested that fixing (a,b) = (0,1) corresponds to using a non-informative prior.

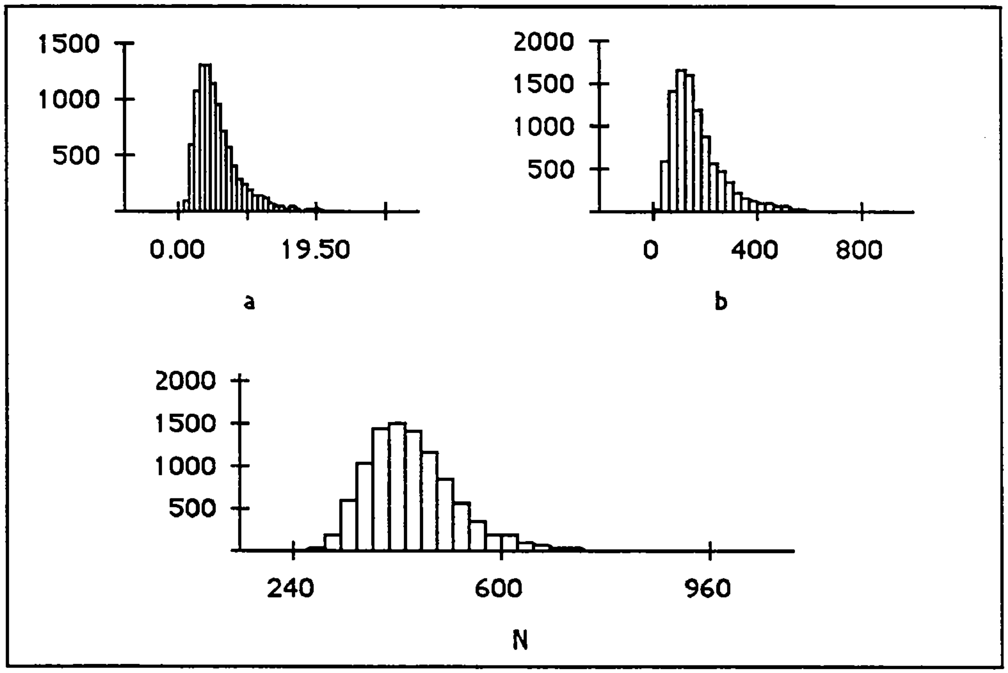

Figure 1. Histograms based on a Gibbs sequence of length 10,000.

Parameter	Mean	St. Dev.	95% Credible Interval
a	5.83	3.7	1.77 – 16.42
b	233.5	157.	59.9 – 673.8
N	464.0	89.4	329 – 679

Table 3. Posterior distribution estimates based on a Gibbs sequence of length 10,000.

5. Various Related Models

The homogeneous catch model (2.1) provides information about the size of N by exploiting the assumption of homogeneous capture probabilities across individuals within capture events. In this section, we consider some examples of alternative capture-recapture models which are based on imposing different homogeneity restrictions on the general model (1.1). For each of these models, we easily obtain the full conditional distributions which provide the basis for Bayesian calculations via the Gibbs sampler.

5.1 The uniform model

Perhaps the simplest model is obtained by assuming complete homogeneity which restricts $p_{ij} \equiv p$ in (1.1), so that all capture probabilities are identical across individuals and samples. The likelihood for this model is

$$L(N, p \mid \text{data}) \propto \frac{N!}{(N-r)!}\, p^{n_+} (1-p)^{I \cdot N - n_+} \tag{5.1}$$

where $n_+ \equiv \sum_{i=1}^{I} n_i$ is the total number of captures. Combining (5.1) with priors of the form $\pi(N,p) = \pi(N)\pi(p)$ as in (2.4) yields full conditionals of the form

$$\pi(N \mid p, \text{data}) \propto \frac{N!}{(N-r)!}\, (1-p)^{I \cdot N}\, \pi(N) \tag{5.2}$$

$$\pi(p \mid N, \text{data}) \propto \{p^{n_+} (1-p)^{I \cdot N - n_+}\}\pi(p). \tag{5.3}$$

Comparing these with (2.5) and (2.6), it is easy to see how the methods of Section 2 may be applied here.

5.2 The behavioral model

An important variation of the capture recapture setup occurs when the probability of recapture is different from the probability of initial capture. This variation is obtained as a special case of the general model (1.1) when the probabilities p_{ij} are allowed to depend on the capture variables $\delta_{i'j}$ where $i' < i$. For example, a special case would have $p_{ij} = p_1$ if individual j has not been captured before time i, and $p_{ij} = p_2$ if individual j has been captured at least once before time i. In the simple case of $I = 2$ capture events, the likelihood for this model becomes

(5.4) $$L(N,p \mid \text{data}) \propto \frac{N!}{(N-r)!} p_1^r p_2^m (1-p_2)^{n_1 - m} (1-p_1)^{2N - n_+}.$$

Combining (5.4) with priors of the form $\pi(N,p) = \pi(N)\pi(p)$ as in (2.4) yields full conditionals of the form

(5.5) $$\pi(N \mid p, \text{data}) \propto \frac{N!}{(N-r)!} (1-p_1)^{2N} \pi(N)$$

(5.6) $$\pi(p \mid N, \text{data}) \propto p_1^r p_2^m (1-p_2)^{n_1 - m} (1-p_1)^{2N - n_+} \pi(p).$$

Comparing these with (2.5) and (2.6), it is easy to see how the methods of Sections 2 and 3 may be applied to this case.

5.3 Stratification models

When the target population can be partitioned into identifiable strata such as *male/female, young/adult/senior, etc....*, it is sometimes more reasonable to assume homogeneity of catch probabilities only within strata. For example, suppose the population was partitioned into S strata such that $p_{ij} = p_s$ if individual i belonged to strata s, where s = 1,...,S. This setup assumes homogeneous catch probabilities within strata across sample events. The likelihood is then obtained as the product of likelihoods (5.1) of the uniform model

(5.7) $$L(N, p \mid \text{data}) \propto \prod_{s=1}^{S} \frac{N_s!}{(N_s - r_s)!} p_s^{n_{+s}} (1-p_s)^{I \cdot N_s - n_{+s}}$$

where N_s is the (unknown) size of the sth stratum, n_{+s} is the total number of captures from the sth stratum, and r_s is the number of distinct individuals captured from the sth stratum. The unknown overall population size is here given by $N = \sum_{s=1}^{S} N_s$. More sophisticated stratification models would be obtained by varying the capture probabilities across samples.

A straightforward approach to analyzing (5.7) would be to treat each of the S problems as separate uniform models of the form (5.1). Using one of the approaches of Section 2, N could then be estimated by the sum of the estimates of $N_1,\ldots,N_S$. A more comprehensive approach would be to use a hierarchical extension similar to those considered in Section 3. For example, suppose it seemed likely that $p_1,\ldots,p_S$ were similar. This could be modeled with an exchangeable prior of the form

$$\pi(N, p \mid \theta)\, \pi(\theta) = \left\{\prod_{s=1}^{S} \pi(N_s)\pi(p_s \mid \theta)\right\}\pi(\theta), \tag{5.8}$$

where

$$\pi(N_s) = Po(\lambda_s), \quad \pi(p_s \mid \theta) = Be(a,b), \quad \pi(\theta) = \pi(a,b) = \text{constant}. \tag{5.9}$$

It can be shown that when $s \geq 2$, this yields a proper posterior. The full conditionals for this prior are

$$\pi(N_s - r_s \mid p, a,b, \text{data}) \propto Po(\lambda_s (1 - p_s)^{I \cdot N_s}), \tag{5.10}$$

$$\pi(p_s \mid N, a,b, \text{data}) \propto Be(n_{+s} + a,\ I \cdot N_s - n_{+s} + b), \tag{5.11}$$

$$\pi(a,b \mid N, p, \text{data}) \propto \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]^S \left[\prod_{s=1}^{S} p_s\right]^a \left[\prod_{s=1}^{S}(1- p_s)\right]^b. \tag{5.12}$$

Both (5.10) and (5.11) can be efficiently simulated by standard methods, whereas (5.12), by virtue of its log concavity, can be efficiently simulated using Gilks and Wild (1991). Thus here too, the Gibbs sampler can here be usefully employed to obtain marginal posterior distributions.

6. Conclusion

We have shown how Gibbs sampling can dramatically facilitate Bayesian analyses of a large variety of capture-recapture models, thus allowing for a much wider choice of the prior distribution. This can be especially valuable when only limited prior information is available. The Gibbs sampler enabled us to use a noninformative hierarchical model to analyze a real example, an analysis that was otherwise intractable. Our analysis led to different results than Castledine (1981), who was forced to choose more limited priors. The reason for the success of the Gibbs sampler in the capture-recapture framework, is the reduction to manageable conditional posterior distributions through data augmentation. This reduction is obtained by alternately treating the population size N and the capture probabilities p_{ij} as missing data.

Obviously, our coverage of capture-recapture models is far from being exhaustive and there are many variants where Gibbs sampling could actually be of use. For instance, a missing data representation would simplify the treatment of open population problems where deaths and immigrations can occur, as is often the case in practice. Such representations would also be appropriate for tag-loss extensions where there is a possible misclassification of recaptured objects (see Seber and Felton (1976)).

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.H. (1986). *Design and analysis methods for fish survival experiments based on release-recapture.* Bethesda, Maryland: America Fisheries Society Monograph.

Casella, G. and George, E.I. (1991). Explaining the Gibbs sampler. *The American Statistician* (to appear).

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* **63** (3), 435-447.

Diebolt, J. and Robert, C. (1990). Estimation of finite mixture distributions through Bayesian sampling. Tech. report 121, LSTA, Université Paris 6.

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

Gilks, W.R. and Wild, P. (1991). Adaptive rejection sampling for Gibbs sampling. *J. Royal Statistical Society, Series C,* (to appear).

Huggins, R. (1989). On the statistical analysis of capture-recapture experiments. *Biometrika* **76** (1), 133-140.

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

Laplace, P.S. (1786). Sur les naissances, les mariages et les morts. In *Histoire de L'Académie Royale des Sciences*, 1783, Paris.

Leite, J.G. and Pereira, C.A. (1990). Bayes estimation of the size of a finite population: Capture/Recapture sequential sample data. *Inter. Statist. Rev.* **58** 3 201-213.

Liu, J., Wong, W.H., and Kong, A. (1991). Correlation structure and convergence rate of the Gibbs sampler (I): Applications to the comparisons of estimators and augmentation schemes. Technical Report 299, Department of Statistics, University of Chicago.

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

Morris, C.N. (1987). "Comment on 'The calculation of posterior distributions by data augmentation' by Tanner and Wong," *J. American Statistical Association* **82** 542-543.

Nayak, T. (1988). Estimating population size by capture-recapture sampling. *Biometrika* **75** (1), 113-120.

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

Pickands, J. and Raghavachari, M. (1989). Exact and asymptotic inference for the size of a population. *Biometrika* **74** (2), 355-363.

Pollock, K.H. (1991). Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: Past, present, and future. *J. American Statistical Association.* **86**, 225-238.

Polya, G. (1976). Probabilities in proof-reading. *Amer. Math. Monthly* **83**, 42.

Schervish, M. and Carlin, B. (1990). On the convergence of succesive substitution sampling. Technical Report 492, Department of Statistics, Carnegie Mellon University.

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

Seber, G. and Felton, R. (1976). Tag loss and the Petersen mark-recapture experiment. *Biometrika* **68** (1), 211-219.

Smith, P. (1988). Bayesian methods for multiple capture-recapture surveys. *Biometrics* **44**, 1177-1189.

Tanner, M. and Wong, W. (1987). The calculation of posterior distributions by data augmentation. *J. American Statistical Association* **81**, 82-86.

Wolter, K. (1986). Some coverage error models for census data. *J. American Statistical Association* **81**, 338-346.