

Capture–recapture estimation via Gibbs sampling

BY EDWARD I. GEORGE

Department of MSIS, The University of Texas at Austin, Austin, Texas 78712, U.S.A.

AND CHRISTIAN P. ROBERT

*Laboratoire de Statistique Théorique et Appliquée, Université Paris VI,
 75252 Paris Cedex 05, France*

SUMMARY

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 easily 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.

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

1. INTRODUCTION

An alternative approach to estimating population size by a census is the capture–recapture methodology which samples the population several times, identifying individuals which appear more than once. 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 (Otis et al., 1978; Seber, 1982; Pollock, 1991). Wolter (1986) describes various implementations of this method in the context of surveys and censuses.

In its most general formulation, the probability model underlying capture–recapture experiments 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 p_{ij} . Assuming that all captures are independent, the likelihood of N and $p = (p_{11}, \dots, p_{IN})$ from data $\mathcal{D}$ is

$$L(N, p | \mathcal{D}) = \prod_{i=1}^I \prod_{j=1}^N p_{ij}^{\delta_{ij}} (1 - p_{ij})^{1 - \delta_{ij}}, \quad (1.1)$$

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 $p_{ij} = p_i$ pertains to experiments where the probability of capture is identical across individuals within each sample; see, for example, Darroch (1958) and Castledine (1981). For the purpose of illustration, we focus on this homogeneous catch model throughout.

As illustrated by the work of Castledine (1981), Jewell (1985) and Smith (1988, 1991), the Bayesian approach readily lends itself to comprehensive probabilistic modelling and offers rich potential in the capture–recapture setting. Unfortunately, the possible difficulty

of calculating posterior quantities of interest is sometimes overcome by introducing unwarranted assumptions. The purpose of this paper is to show how Gibbs sampling, as an alternative to both analytical calculation and numerical approximation, can facilitate Bayesian calculations for capture-recapture models, thereby enhancing their scope.

For a large variety of priors $\pi(N, p)$, the Gibbs sampler yields an approximation to the marginal posterior of N , namely $\pi(N|\mathcal{D})$, by simulated sampling from the conditional distributions

$$\pi(N|p, \mathcal{D}), \quad \pi(p|N, \mathcal{D}), \quad (1.2)$$

(Gelfand & Smith, 1990; Casella & George, 1992). Starting with an initial value $N^{(0)}$ for N , such as the maximum likelihood estimate, iterations of this simulation can be used to produce a 'Gibbs sequence'

$$N^{(0)}, p^{(0)}, N^{(1)}, p^{(1)}, \dots, \quad (1.3)$$

where $N^{(k)} \sim \pi(N|p^{(k-1)}, \mathcal{D})$ and $p^{(k)} \sim \pi(p|N^{(k)}, \mathcal{D})$. It follows from Tierney (1991) that, when the support of $\pi(p|N, \mathcal{D})$ does not depend on N , the subsequence $\{p^{(k)}\}$ in (1.3) is an ergodic Markov chain with stationary distribution $\pi(p|\mathcal{D})$. The duality principle of Diebolt & Robert (1993) then implies that $\{N^{(k)}\}$ is also ergodic with stationary distribution $\pi(N|\mathcal{D})$. Thus, the Gibbs sequence can be used to approximate $\pi(N|\mathcal{D})$. For instance, the ergodic theorem ensures that $\bar{N} = \sum N^{(k)} / K$ is a consistent estimator of $E(N|\mathcal{D})$.

A major reason for the successful implementation of the Gibbs sampler in these capture-recapture set-ups is the simplification provided by the conditioning in (1.2). This stems from the fact that 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 analytical approach of integrating out the p_{ij} , which often leads to intractable expressions, does not exploit this special structure.

The plan of this paper is as follows. In § 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 § 3, we show how these methods may also be applied to natural hierarchical extensions of these models. In § 4, the Gibbs sampler is used to compute hierarchical Bayes estimates for the Gordy Lake sunfish data treated by Castledine (1981) and Smith (1988, 1991). Further extensions are discussed briefly in § 5.

2. BAYES ESTIMATES FOR THE HOMOGENEOUS CATCH MODEL

We now focus on the homogeneous catch model described in the introduction. From a population of unknown size N , I samples of sizes $n_1, \dots, n_I$ are consecutively drawn, marked, and returned to the population. The total number of distinct captured individuals, denoted by r , is recorded. Note that $r = (\sum n_i) - m$, where m is the total number of times a recaptured individual is observed. If $n_1, \dots, n_I$ 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 i th sample is $p_{ij} \equiv p_i$, namely

$$L(N, p|\mathcal{D}) \propto \frac{N!}{(N-r)!} \prod_{i=1}^I p_i^{n_i} (1-p_i)^{N-n_i}, \quad (2.1)$$

where henceforth $p = (p_1, \dots, p_I)$.

Following Castledine (1981), we consider priors of the form $\pi(N, p) = \pi(N)\pi(p)$, so that N and p are a priori independent. Such priors lead to posterior conditionals of the form

$$\begin{aligned}\pi(N|p, \mathcal{D}) &\propto \frac{N!}{(N-r)!} \left\{ \prod_{i=1}^I (1-p_i) \right\}^N \pi(N), \\ \pi(p|N, \mathcal{D}) &\propto \left\{ \prod_{i=1}^I p_i^{n_i} (1-p_i)^{N-n_i} \right\} \pi(p).\end{aligned}\tag{2.2}$$

In the special case where the p_i 's are a priori independent, $\pi(p) = \prod \pi(p_i)$, and $\pi(p_i) = \text{Be}(a, b)$, the beta distribution with mean $\mu = a/(a+b)$ and variance $\tau^2 = \mu(1-\mu)/(a+b+1)$, it is easy to show that

$$\pi(N|\mathcal{D}) = \frac{N!}{(N-r)!} \left\{ \prod_{i=1}^I \frac{\Gamma(N-n_i+b)}{\Gamma(N+a+b)} \right\} \pi(N).\tag{2.3}$$

Since

$$\frac{\pi(N+1|\mathcal{D})}{\pi(N|\mathcal{D})} = \frac{N+1}{N-r} \left(\prod_{i=1}^I \frac{N-n_i+b}{N+a+b} \right) \frac{\pi(N+1)}{\pi(N)},\tag{2.4}$$

(2.3) may sometimes be approximated rather accurately by recursion if $\pi(N+1)/\pi(N)$ is readily available such as when $\pi(N)$ is Poisson.

Obviously, not all prior formulations allow as clean a solution as (2.3) and (2.4). However, if one can simulate from the conditional posteriors $\pi(N|p, \mathcal{D})$ and $\pi(p|N, \mathcal{D})$ as in (1.2), then the Gibbs sampler may be a promising computational approach. It follows from (2.2) that the support of $\pi(p|N, \mathcal{D})$ does not depend on N . Thus, $\{p^{(k)}\}$ and $\{N^{(k)}\}$ will be Markov chains with stationary distributions $\pi(p|\mathcal{D})$ and $\pi(N|\mathcal{D})$ respectively.

To implement the Gibbs sampler, the only practical restrictions on $\pi(N)$ and $\pi(p)$ is that they lead to conditional posteriors (2.2) which are easily simulated. Fortunately, a variety of choices lead to standard distributions for which simulation programs are readily available. For example, for the simulation of $\{N^{(k)}\}$, if $\pi(N)$ is Poisson with mean λ , the conditional posterior of $N-r$ is Poisson with mean $\lambda \prod (1-p_i)$. Alternatively, for the Jeffreys prior $\pi(N) = 1/N$, the conditional posterior of N is negative binomial with parameters $r-1$ and $1 - \prod (1-p_i)$. For the simulation of $\{p^{(k)}\}$, under independent $\text{Be}(a, b)$ priors for the p_i 's, the conditional posteriors of the p_i 's are independent $\text{Be}(n_i+a, N-n_i+b)$.

Even when the conditional posterior of p is nonstandard, $\{p^{(k)}\}$ can be simulated by the adaptive rejection sampling, ARS, algorithm of Gilks & Wild (1992) whenever $\pi(p|N, \mathcal{D})$ is log concave. From (2.2), this will be satisfied whenever $\log \pi(p_i)$ is concave for each i . One might also consider a logit model on the p_i 's, that is

$$\alpha_i = \log \{p_i/(1-p_i)\} \sim N(\mu_i, \sigma^2), \quad \pi(\alpha_1, \dots, \alpha_I | N, \mathcal{D}) = \prod \pi(\alpha_i | N, \mathcal{D}).$$

Because

$$\pi(\alpha_i | N, \mathcal{D}) \propto \exp \{ \alpha_i n_i - \frac{1}{2}(\alpha_i - \mu_i)^2 / \sigma^2 \} / (1 + e^{\alpha_i})^N$$

is log concave in α_i , these α_i can be simulated by adaptive rejection sampling.

3. HIERARCHICAL BAYES EXTENSIONS

The capture-recapture setting can be substantially enhanced by embedding it in a hierarchical framework in which the components of p can be modelled. This may be accomplished with priors of the form

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

where θ is a hyperparameter governing each prior $\pi(p_i | \theta)$ (Castledine, 1981). Here, N and p are a priori independent, and $p_1, \dots, p_I$ are a priori exchangeable.

For the likelihood (2.1) coupled with a prior of the form (3.1), implementing the Gibbs sampler entails producing a Gibbs sequence $\{N^{(k)}, p^{(k)}, \theta^{(k)}\}$ ($k = 0, 1, \dots$), which, except for preselected initial values, is obtained by alternately simulating from the conditional posteriors

$$\pi(N | p, \theta, \mathcal{D}) = \pi(N | p, \mathcal{D}) \propto \frac{N!}{(N-r)!} \left\{ \prod_{i=1}^I (1-p_i) \right\}^N \pi(N),$$

$$\pi(p | N, \theta, \mathcal{D}) \propto \prod_{i=1}^I p_i^{n_i} (1-p_i)^{N-n_i} \pi(p_i | \theta),$$

$$\pi(\theta | N, p, \mathcal{D}) = \pi(\theta | p) \propto \left\{ \prod_{i=1}^I \pi(p_i | \theta) \right\} \pi(\theta).$$

When the support of $\pi(p_i | \theta)$ does not depend on θ , $\{N^{(k)}\}$, $\{p^{(k)}\}$ and $\{\theta^{(k)}\}$ will each be Markov chains with stationary distributions $\pi(N | \mathcal{D})$, $\pi(p | \mathcal{D})$ and $\pi(\theta | \mathcal{D})$ respectively.

Because $\pi(N | p, \theta, \mathcal{D})$ is as before, and $\pi(p | N, \theta, \mathcal{D})$ is a special case of the conditional posteriors in (2.2), the same considerations concerning ease of simulation apply here. For example, if $\pi(N)$ is Poisson we obtain a Poisson posterior, $\pi(N) = 1/N$ yields a negative binomial posterior, and $\pi(p_i | \theta) = \text{Be}(a, b)$ yields Beta posteriors. To simulate $\theta = (a, b)$ for these cases, the conditional posterior is of the form

$$\pi(a, b | N, p, \mathcal{D}) \propto \left\{ \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right\}^I \left(\prod_{i=1}^I p_i \right)^a \left\{ \prod_{i=1}^I (1-p_i) \right\}^b \pi(a, b). \quad (3.2)$$

In many cases, this posterior will be log concave, and so can be simulated by adaptive rejection sampling. For example, the first three terms on the right-hand side of (3.2) comprise a log concave probability distribution in a and b when $I \geq 2$ (George, Makov & Smith, 1993). Thus, whenever $\pi(a, b)$ is log concave or even constant, $\pi(a, b | N, p, \mathcal{D})$ will be log concave. It can also be shown directly that $\pi(a, b | N, p, \mathcal{D})$ will be log concave under priors of the form $\pi(a, b) \propto (a+b)^{-k} I(a, b > 1)$ for $k \leq I$.

An alternative hierarchical model is obtained by extending the logit model introduced previously. The means μ_i of the logits $\alpha_i = \log \{p_i / (1-p_i)\}$ are usually unknown and may be modelled as exchangeable normal variables $N(\eta, \tau^2)$ (Castledine, 1981). This representation is particularly attractive when covariates can be used to model the mean $\eta = X'\beta$ (Huggins, 1989). In any case the simulation of μ_i is straightforward since

$$\pi(\mu_i | N, p, \mathcal{D}) = \pi(\mu_i | p) = N\left(\frac{\tau^2 \alpha_i + \sigma^2 \eta}{\tau^2 + \sigma^2}, \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2}\right)$$

is easily simulated. Further elaborations, such as putting priors on σ , τ and η , are also easily accommodated by Gibbs sampling.

4. THE SUNFISH EXAMPLE

In this section, we briefly illustrate our techniques on the Gordy Lake sunfish data set treated by Castledine (1981) and Smith (1988, 1991). As shown in Table 1, it consists of $I = 14$ capture occasions from a population of sunfish. At the i th capture, n_i fish are captured out of which m_i have been previously captured. Thus, $r = \sum (n_i - m_i) = 138$ is the total number of different fish captured.

Table 1. Capture-recapture counts of Gordy Lake sunfish

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

Castledine (1981) and Smith (1991) applied the homogeneous catch model (2.1) to this data using the prior formulation $\pi(N, p|a, b) = \pi(N) \prod \pi(p_i|a, b)$ with Jeffreys' prior $\pi(N) \propto 1/N$, and $\pi(p_i|a, b) = \text{Be}(a, b)$ for various fixed (a, b) . Table 2 lists characteristics of $\pi(N|\text{data})$ for some of their choices computed using (2.4). The choices $(a, b) = (0, 0)$, $(0, 1)$, $(1, 1)$ can be motivated as noninformative, the choices $(a, b) = (2, 100)$, $(3, 100)$ as informative, and $(a, b) = (1.79, 56.6)$, which maximizes the likelihood, as empirical Bayes. The wide variability in posterior characteristics shows sensitive dependence on fixed choices of (a, b) .

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

a	b	Mean of N	Std. dev. of N	95% credible interval
0	0	441	80.2	316-628
0	1	446	81.4	319-636
1	1	321	43.2	251-419
2	100	507	70.5	389-664
3	100	419	51.2	332-532
1.79	56.6	419	57.0	324-547

Instead of treating (a, b) as fixed, we pursued the hierarchical approach of putting a prior on (a, b) . We considered the five priors: $\pi(a, b) \propto \exp\{-(a+b)/k\}$ for $k = 10, 100, 1000$, $\pi(a, b) = 1$, and $\pi(a, b) \propto (a+b)^{-2}I(a, b > 1)$. These priors all put uniform probability on $\mu = a/(a+b)$, the mean of the beta $\text{Be}(a, b)$, and vary only in the distribution on $(a+b)$. The prior mean of $(a+b)$ is 20, 200, 2000 under the first three priors respectively, and is infinite under the last two priors. For each of these priors, we used the Gibbs sampler to approximate posterior characteristics of N, p, a, b , and $\mu = a/(a+b)$. Smith (1991), who applied the prior $\pi(a, b) \propto (a+b)^{-2}I(a, b > 0)$ to express vague prior information, used a Laplace approximation to obtain posterior characteristics of N .

Setting initial values by maximum likelihood, we simulated the values $p^{(k)}$ and $N^{(k)}$ in the Gibbs sequence using the IMSL routines DRNBET and RNNBN respectively, and we simulated $a^{(k)}$ and $b^{(k)}$ using adaptive rejection sampling. For each prior, we generated a Gibbs sequence of length 10^6 . For each sequence, Table 3 lists the means $\bar{a}$, $\bar{b}$, $\bar{\mu}$, $\bar{N}$, the standard deviation s_N of N , and a 95% credible interval for N obtained from the 2.5% and 97.5% quantiles. These may be regarded as estimates of the corresponding posterior characteristics.

Table 3 shows that the posterior distributions of a and b vary considerably across our choices of $\pi(a, b)$. However, the posterior means $\mu = a/(a+b)$ are relatively constant,

Table 3. *Posterior characteristics of a , b , μ , N for various hierarchical priors*

$\pi(a, b)$	$\bar{a}$	$\bar{b}$	$\bar{\mu}$	$\bar{N}$	s_N	95% credible interval
$\exp\{-(a+b)/10\}$	1.54	42.6	0.036	403	65.8	297-555
$\exp\{-(a+b)/100\}$	3.64	133	0.028	440	77.7	317-617
$\exp\{-(a+b)/1000\}$	5.41	213	0.026	458	85.1	327-661
Constant	6.46	257	0.026	462	86.4	325-661
$(a+b)^{-2}I(a, b > 1)$	3.33	116	0.029	432	76.7	309-615

around 0.03, suggesting that most of the variation is in the distribution of $(a+b)$. As $(\bar{a} + \bar{b})$ increases both $\bar{N}$ and s_N increase, although these values vary less over Table 3 than over Table 2. Note that the priors in Table 2 with similar values of $\mu \approx 0.03$ yield much smaller s_N and credibility intervals.

5. CONCLUSION

We have shown how Gibbs sampling facilitates Bayesian analyses of homogeneous capture-recapture experiments by allowing for a much wider choice of prior distributions. This is especially valuable when only limited prior information is available, as was demonstrated on the sunfish data.

The homogeneous catch model is only one of a large variety of capture-recapture models. The applicability of the Gibbs sampler to these models relies only on the missing value multinomial structure. Fortunately, many of the capture-recapture variants preserve this structure. For instance, the homogeneous model can be replaced by a behavioural model (Wolter, 1986), where the probability of capture is modified after a capture, or a stratified model, where the target population is partitioned into homogeneous subgroups with different catch probabilities. It also appears that a missing data representation would simplify the treatment of open population problems where deaths and immigration can occur, as is often the case in practical problems such as estimating the number of illegal aliens. Such representations would also be appropriate for tag-loss extensions where there is a possible misclassification of recaptured objects (Seber & Felton, 1976).

ACKNOWLEDGEMENTS

The authors would like to thank George Casella and Martin Wells for helpful discussions, and the referees for many suggestions including the recursive relation (2.4).

REFERENCES

- CASELLA, G. & GEORGE, E. I. (1992). Explaining the Gibbs sampler. *Am. Statistician* **46**, 167-74.
 CASTLEDINE, B. (1981). A Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika* **67**, 197-210.
 DARROCH, J. (1958). The multiple-recapture census, I: Estimation of a closed population. *Biometrika* **63**, 435-47.
 DIEBOLT, J. & ROBERT, C. P. (1993). Estimation of finite mixture distributions through Bayesian sampling. *J.R. Statist. Soc. B*. To appear.
 GELFAND, A. & SMITH, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* **85**, 398-409.
 GEORGE, E. I., MAKOV, U. E. & SMITH, A. F. M. (1993). Conjugate likelihood distributions. *Scand. J. Statist.* To appear.
 GILKS, W. R. & WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* **41**, 337-48.

- HUGGINS, R. (1989). On the statistical analysis of capture-recapture experiments. *Biometrika* **76**, 133-40.
- JEWELL, W. (1985). Bayesian estimation of undetected errors. In *Bayesian Statistics 2*, Ed. J. Bernardo, M. DeGroot, D. V. Lindley and A. F. M. Smith, pp. 663-72. Amsterdam: North-Holland.
- LAPLACE, P. S. (1786). Sur les naissances, les mariages et les morts. In *Histoire de L'Académie Royale des Sciences*, 1783, Paris.
- OTIS, D. L., BURNHAM, K. P., WHITE, G. C. & ANDERSON, D. R. (1978). Statistical inference for capture data on closed animal populations. *Wildlife Monographs* **62**, 1-135.
- 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. Am. Statist. Assoc.* **86**, 225-38.
- SEBER, G. (1982). Capture-recapture methods. In *Encyclopedia of Statistical Sciences*, Ed. S. Kotz and N. Johnson, **1**, pp. 367-74. New York: Wiley.
- SEBER, G. & FELTON, R. (1976). Tag loss and the Petersen mark-recapture experiment. *Biometrika* **68**, 211-9.
- SMITH, P. J. (1988). Bayesian methods for multiple capture-recapture surveys. *Biometrics* **44**, 1177-89.
- SMITH, P. J. (1991). Bayesian analyses for a multiple capture-recapture model. *Biometrika* **78**, 399-408.
- TIERNEY, L. (1992). Exploring posterior distributions using Markov chains. In *Computer Science and Statistics: 23rd Symposium on the Interface*, Ed. E. Keramidas, pp. 563-70.
- WOLTER, K. (1986). Some coverage error models for census data. *J. Am. Statist. Assoc.* **81**, 338-46.

[Received August 1990. Revised April 1992]