Ann Inst Stat Math (2014) 66:245–278
DOI 10.1007/s10463-013-0413-1

Objective Bayesian analysis for a capture–recapture model

Chang Xu · Dongchu Sun · Chong He

Received: 24 August 2012 / Revised: 29 March 2013 / Published online: 22 June 2013

Abstract In this paper, we study a special capture–recapture model, the M_t model, using objective Bayesian methods. The challenge is to find a justified objective prior for an unknown population size N. We develop an asymptotic objective prior for the discrete parameter N and the Jeffreys' prior for the capture probabilities $\boldsymbol{\theta}$. Simulation studies are conducted and the results show that the reference prior has advantages over ad-hoc non-informative priors. In the end, two real data examples are presented.

Keywords Capture–recapture · Objective prior · Frequentist property

1 Introduction

Capture–recapture is a method used to estimate population sizes. It was originally developed for ecological studies and has been utilized extensively in epidemiological and demographic research. A sequence of classic capture–recapture models were discussed in Otis et al. (1978), commonly referred to as M_0, M_t, M_b, M_h, M_{tb}, M_{th}, M_{bh}, and M_{tbh}. For all of these models, closed populations were assumed. Meanwhile, each model allows for different assumptions regarding capture probabilities. Bayesian approaches have been proved very useful in analyzing both classic capture–recapture models and more complex, hierarchical models that incorporate spatial patterns and covariates (see Royle 2008; Wang et al. 2007; Basu and Ebrahimi 2001; George and

C. Xu (✉) · D. Sun · C. He
Department of Statistics, University of Missouri, Columbia, MO 65211, USA
e-mail: cx3z9@mail.missouri.edu

D. Sun
e-mail: sund@missouri.edu

C. He
e-mail: hezh@missouri.edu

Robert 1992; Smith 1991). Results from these papers show that the Bayesian estimates of population size N are sensitive to hyper-parameters of the prior distribution, especially when the number of sampling occasions k or capture probabilities $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ are small. One common approach for selecting a prior for N is imposing a hierarchical structure and placing priors on the hyper-parameters. For example, Raftery (1988) and Smith (1991) recommended using Poisson(λ) as prior for N and assume Gamma distribution on λ. Another solution is to use non-informative priors. For example, Smith (1991) suggested the Jeffreys prior $1/N$ when very vague prior information is available. George and Robert (1992) also studied M_t model and computed Bayesian estimates via Gibbs sampling. They suggested using the $1/N$ or Poisson prior on N. Bolfarine et al. (1992) used the proper prior $1/(N+1)^2$. Farcdella and collaborators Farcomeni and Tardella (2010); Tardella (2002) proposed both parametric and nonparametric methods to estimate N under M_h model. More recently, Wang et al. (2007) compared a family of non-informative priors $1/N^r$ with $r = 0$ (uniform prior) and $r = 1$ (Jeffreys prior) being two special cases. However, these priors are ad-hoc and do not work well consistently. For instance, Wang et al. (2007) proved that under M_t model with fixed N and $\boldsymbol{\theta}$, Bayesian estimate of N would decrease as the number of sampling occasions increases. Consequently, Bayesian estimates underestimate N for large k.

Objective Bayesian methods have been widely used to develop non-ad-hoc priors for N. Berger et al. (2012) proposed a new approach to develop objective priors for integer parameters. In their approach, the nuisance parameters are integrated out from the model and the marginal likelihood function of the integer parameter is obtained. They treat the integer parameter as a continuous one and obtain the Jeffreys prior. This approach has been proved to work for some challenging problems, such as estimating the number of failures when the survival time is under type II censoring. In our paper, we will apply this approach to develop prior distributions for N in a capture–recapture model. An alternative method was proposed by Barger and Bunge (2010), who developed objective priors for the number of species based on the linear difference score (see Lindsay and Roeder 1987).

We will focus on the M_t model, one of the nine classic capture–recapture models described in Otis et al. (1978). In addition to the common "closed population" assumption, M_t models require (1) homogeneity among all units in the population or in other words, that all different units in the population have the same probability to be captured and (2) capture probabilities vary across multiple sampling occasions. In this model, the observed capture history data follow a multinomial distribution with parameters N and $\boldsymbol{\theta}$. There are several reasons why we choose to study such a simple model. First of all, the M_t model has prodigious applications in wildlife management, ecology, epidemiology, and other areas. Second, this is the first paper that attempts to formally develop the reference prior for population size in a capture–recapture model. We start with a simple model and will continue by working on more complex ones. Third, the formal approach of deriving reference prior is rather complicated even for a model as simple as M_t.

The paper is organized as following: in Sect. 2, the conditional Jeffreys prior for $\boldsymbol{\theta}$ given N will be derived from the conditional likelihood. In Sect. 3, we will develop the asymptotic reference prior for large N based on the marginal likelihood. A discussion of the conditions for proper posterior and the existence of any fixed moment will be given in the end of this section. In Sect. 4.1, we will present and explain the simulation results and show that the reference priors have advantages over more commonly used non-informative priors. An analysis of a real data example is presented in Sect. 4.2. We will discuss future works in the conclusion, Sect. 5.

2 Reference prior of $\boldsymbol{\theta}$

We assume that the capture–recapture experiment consists of k sampling occasions and give the reference prior of $(N, \boldsymbol{\theta})$. Let θ_j be the capture probability and m_j be the number of animals captured during the jth samplings occasion, $j = 1, \ldots, k$. The capture history data may be denoted by k digits binary numbers, within which '0' represents 'not captured' and '1' represents 'captured'. For example, if $k = 3$, Y_{101} denotes the number of animals that are caught in the first and the third trappings, but escaped from the second. Clearly, there are $L = 2^k$ number of Y's. Converting these binary numbers to decimals, we define $n_1, \ldots, n_L$ to be the complete capture record data. Let $n = \sum_{i=1}^{L} n_i$ be the number of animals that were captured at least once and $n_0 = N - n$ be the number of animals that were never captured. The goal is to estimate N, or equivalently n_0, based on $n_1, \ldots, n_L$. For $i = 0, \ldots, L$, n_i follows a binomial distribution with parameters N and $p_i(\boldsymbol{\theta})$. Define

$$p_i(\boldsymbol{\theta}) = \prod_{j=1}^{k} \theta_j^{b_{ij}} (1 - \theta_j)^{1-b_{ij}}, \tag{1}$$

where b_{ij} is the jth binary digit of the integer i. The probability mass function of $n_1, \ldots, n_L$ given N and $\boldsymbol{\theta}$ is

$$f_1(N, \boldsymbol{\theta} \mid n_1, \ldots, n_L) = \frac{N!}{n_0! n_1! \cdots n_L!} \prod_{i=0}^{L} p_i(\boldsymbol{\theta})^{n_i}. \tag{2}$$

The Jeffreys prior Jeffreys (1961) for $\boldsymbol{\theta}$ given N can be derived easily as in the following theorem.

Theorem 1 *The Jeffreys prior for $\boldsymbol{\theta}$ given N is the product of $Beta(\frac{1}{2}, \frac{1}{2})$, and given by*

$$\pi(\boldsymbol{\theta} \mid N) \propto \frac{1}{\prod_{i=1}^{k} \sqrt{\theta_i (1 - \theta_i)}}. \tag{3}$$

The proof is given in the appendix.

3 Parameter-based asymptotic objective prior of N

In general, we assume for $j = 1, \ldots, k$,

$$\theta_j \overset{\text{iid}}{\sim} \text{Beta}(a, b), \tag{4}$$

where a, b are two positive constants. Denote $\pi_{a,b}(\boldsymbol{\theta})$ be the joint density of vector $\boldsymbol{\theta}$. For convenience, we rewrite the likelihood function (2) in the following form:

$$f_2(N, \boldsymbol{\theta} \mid n_1, \ldots, n_L) = \frac{N!}{n_0!n_1! \cdots n_L!} \prod_{i=1}^{k} \theta_i^{m_i} (1-\theta_i)^{N-m_i}. \tag{5}$$

The equivalence of (2) and (5) can be verified by simple algebra. First, we obtain the marginal likelihood of N by averaging $\boldsymbol{\theta}$ over its prior.

$$\begin{aligned} f_3(N \mid n_1, \ldots, n_L) &= \int L_2(N, \boldsymbol{\theta} \mid n_1, \ldots, n_L)\pi_{a,b}(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= \int_0^1 \cdots \int_0^1 \frac{N!}{n_0!n_1! \cdots n_L!} \prod_{i=1}^{k} \theta_i^{m_i}(1-\theta_i)^{N-m_i}\theta_i^{a-1}(1-\theta_i)^{b-1} d\theta_1 \cdots d\theta_k \\ &= \frac{N!}{n_0!n_1! \cdots n_L!} \int_0^1 \cdots \int_0^1 \prod_{i=1}^{k} \theta_i^{m_i+a-1}(1-\theta_i)^{N-m_i+b-1} d\theta_1 \cdots d\theta_k \\ &= \frac{N!}{n_0!n_1! \cdots n_L!} \prod_{i=1}^{k} \frac{\Gamma(m_i+a)\Gamma(N-m_i+b)}{\Gamma(N+a+b)}. \end{aligned} \tag{6}$$

Taking the second derivative of $\log f_3(N \mid n_1, \ldots, n_L)$ with respect to N, we have

$$\begin{aligned} &\frac{\partial^2}{\partial N^2} \log f_3(N \mid n_1, \ldots, n_L) \\ &\quad = \frac{\partial^2}{\partial N^2} \log \Gamma(N+1) - k\frac{\partial^2}{\partial N^2} \log \Gamma(N+a+b) \\ &\qquad + \frac{\partial^2}{\partial N^2} \sum_{l=1}^{k} \log \Gamma(N-m_l+b) - \frac{\partial^2}{\partial N^2} \log \Gamma(N-n+1). \end{aligned} \tag{7}$$

Using the following property of trigamma function (Abramowitz and Stegun 1972),

$$\frac{\partial^2}{\partial N^2} \log \Gamma(N) = \sum_{i=0}^{\infty} \frac{1}{(N+i)^2}, \tag{8}$$

we have

$$\text{The RHS of (7)} = -J_1(N) + J_2(N) + J_3(N) - J_4(N), \tag{9}$$

where

$$J_1(N) = \sum_{i=0}^{\infty} \frac{k}{(N+a+b+i)^2} - \sum_{i=0}^{\infty} \frac{1}{(N+1+i)^2}, \tag{10}$$

$$J_2(N) = \sum_{l=1}^{k} \sum_{i=0}^{\infty} \frac{1}{(N-m_l+b+i)^2} - \sum_{l=1}^{k} \frac{1}{N-m_l+1}, \tag{11}$$

$$J_3(N) = \sum_{l=1}^{k} \frac{1}{N-m_l+1}, \tag{12}$$

$$J_4(N) = \sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2}. \tag{13}$$

The Fisher information of N is then

$$I(N) = J_1(N) - \mathrm{E}(J_2(N)) - \mathrm{E}(J_3(N)) - \mathrm{E}(J_4(N)), \tag{14}$$

where the expectation is taken under the marginal distribution of $(n_1, \ldots, n_L)$, given by (6). We treat N as a continuous random variable and the reference prior of N will be obtained by taking the square root of the Fisher information $I(N)$. The main results are summarized below:

Lemma 1 *For any fixed $k \geq 2$ and $0 < a, b \leq 1$, as $N \to \infty$, we have*

$$E(J_1(N)) \sim \frac{k}{N+a+b} - \frac{1}{N+1} \tag{15}$$

$$E(J_4(N)) \sim \frac{(\log N)^{k+[b]-1}}{(N+1)^b}, \tag{16}$$

where $[b]$ denotes the largest integer that is less than or equal to b. For $0 < b < 1$,

$$E(J_2(N)) \sim \frac{1}{(N+1)^b}, \tag{17}$$

$$E(J_3(N)) \sim \frac{1}{(N+1)^b}. \tag{18}$$

For $b = 1$,

$$E(J_2(N)) \sim \frac{\log(N+a)}{N+1}, \tag{19}$$

$$\frac{a}{N+a} \leq E(J_3(N)) \leq \frac{C_1 \log(N+a)}{N+1}, \tag{20}$$

where C_1 is a positive constant.

The proof the the lemma is given in the appendix.

Theorem 2 *Let $\pi^R(N)$ be the reference prior of N. For any fixed $k \geq 2$ and $0 < a \leq 1$,*

$$\pi^R(N) \propto \frac{(\log N)^{\frac{k+[b]-1}{2}}}{(N+1)^{\frac{b}{2}}}. \tag{21}$$

Proof Since

$$\pi^R(N) = -\mathrm{E}(I(N)),$$

Theorem 2 follows directly from (14) and Lemma 1. □

Corollary 1 *For any fixed $k \geq 2$, $0 < a, b \leq 1$, and non-negative integer g, under prior (21), the gth posterior moment of N exists if there exists $\epsilon > 0$ such that*

$$\sum_{l=1}^{k} m_l - n + ka + b/2 \geq g + 1 + \epsilon. \tag{22}$$

When $g = 0$, the inequality above ensures a proper posterior.

Proof We obtain the marginal likelihood function of N by integrating out $\boldsymbol{\theta}$ over $\pi_{a,b}(\boldsymbol{\theta})$.

$$\begin{aligned} L(N \mid n_1, \ldots, n_L) &= \int L(N, \boldsymbol{\theta} \mid n_1, \ldots, n_L)\pi_{a,b}(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &\propto \frac{\Gamma(N+1)}{\Gamma(N-n+1)} \prod_{l=1}^{k} \frac{\Gamma(m_l+a)\Gamma(N-m_l+b)}{\Gamma(N+a+b)} \\ &\propto \frac{\Gamma(N+1)}{\Gamma(N-n+1)} \prod_{l=1}^{k} \frac{\Gamma(N-m_l+b)}{\Gamma(N+a+b)}. \end{aligned} \tag{23}$$

Gamma functions have the following property (see Erdelyi and Tricomi (1951)): as $z \to \infty$,

$$\frac{\Gamma(z+\alpha)}{\Gamma(z+\beta)} \sim z^{\alpha-\beta}. \tag{24}$$

It follows from (23) to (24) that as $N \to \infty$,

$$L(N \mid n_1, \ldots, n_L) \sim N^{-(ka-n+\sum_{l=1}^{k} m_l)}. \tag{25}$$

The right tail of the marginal posterior density of N can be calculated below. For $0 < b < 1$,

$$\begin{aligned}
\pi(N \mid n_1, \ldots, n_L) &\propto L(N \mid n_1, \ldots, n_L)\pi^R(N) \\
&\sim (\log N)^{\frac{k-1}{2}} (1+N)^{-\frac{b}{2}} N^{-\left[ka-n+\sum_{l=1}^k m_l\right]} \\
&\sim (\log N)^{\frac{k-1}{2}} N^{-\left[ka+b/2+\sum_{l=1}^k m_l - n\right]}. \qquad (26)
\end{aligned}$$

For $b = 1$, similarly we have

$$\pi(N \mid n_1, \ldots, n_L) \sim (\log N)^{\frac{k}{2}} N^{-\left[ka+1/2+\sum_{l=1}^k m_l - n\right]}. \qquad (27)$$

Clearly, if condition (22) holds, for $0 < b < 1$ we have

$$\mathrm{E}(N^g) = \sum_{N=n}^{\infty} N^g \pi(N \mid n_1, \ldots, n_L) \le \sum_{N=n}^{\infty} \frac{(\log N)^{\frac{k-1}{2}}}{N^{1+\epsilon}} < \infty; \qquad (28)$$

for $b = 1$,

$$\mathrm{E}(N^g) = \sum_{N=n}^{\infty} N^g \pi(N \mid n_1, \ldots, n_L) \le \sum_{N=n}^{\infty} \frac{(\log N)^{\frac{k}{2}}}{N^{1+\epsilon}} < \infty. \qquad (29)$$

This completes the proof. □

This condition is very easy to satisfy in practice. For example, when $g = 0$, inequality (22) holds as long as there is at least one recapture.

4 Numerical examples

4.1 Simulation studies

In the simulation study, we compare the estimation of N under four different prior combinations for $(N, \boldsymbol{\theta})$ and the maximum likelihood estimate. With the Jeffreys prior fixed on θ_j, we consider the asymptotic reference prior $\pi^R(N)$, the uniform prior $\pi^U(N) = 1$, and the Jeffreys prior $\pi^J(N) = 1/N$ for N. We also consider the uniform prior for $\boldsymbol{\theta}$ ($a = b = 1$) and the corresponding asymptotic prior for N. These four prior combinations are coded as π_{RU}, π_{RJ}, π_{JJ}, and π_{UJ}. The first subscript indicates the prior for N and the second subscript indicates the prior for $\boldsymbol{\theta}$, where 'R' for the asymptotic reference prior, 'U' for the uniform prior, and 'J' for the Jeffreys prior. The details and results of the simulation study are summarized below.

In the simulation, the number of sampling occasions k varies from 3 to 8. The rest of simulation setups are adopted from Wang et al. (2007). The population size N is set to vary from 50 to 800 with 50 increments. Three sets of capture probabilities are used. They are (0.09, 0.08, 0.07, 0.06, 0.08, 0.09, 0.06, 0.07), (0.09, 0.18, 0.07, 0.16, 0.08, 0.19, 0.06, 0.17), and (0.26, 0.27, 0.28, 0.29, 0.28, 0.29, 0.26, 0.27), representing small, moderate, and large capture probabilities, respectively.

The posterior means are estimated numerically when they exist. For each $(N, \boldsymbol{\theta}, k)$ combination, 2,000 capture history datasets are generated according to likelihood (2). For the jth dataset, the posterior mean, given it exists, is estimated by

$$\hat{N}_j = \frac{\sum_{N=n}^{V} NL(N \mid \text{data}_j)\pi^R(N)}{\sum_{N=n}^{V} L(N \mid \text{data}_j)\pi^R(N)}, \tag{30}$$

where $V = 10^6/2$ (in some cases V is set to be 10^6 to ensure convergence) and $L(N \mid \text{data})$ is given in (2). The mean of the posterior means is denoted as $\hat{N}$. The estimated bias and mean square error can be calculated based on $\hat{N}_j$'s:

$$\widehat{\text{Bias}} = \hat{N} - N, \tag{31}$$

$$\widehat{\text{MSE}} = \frac{1}{2{,}000} \sum_{j=1}^{2{,}000} (\hat{N}_j - N)^2. \tag{32}$$

We also compare 95 % frequentist coverage probabilities and credible intervals of the posterior mean of N. The average 95 % credible intervals of the estimated population size are calculated for each setting under four different priors on θ_j: Beta (0.25, 0.25), Beta (0.5, 0.5), Beta (0.75, 0.75), and Beta (1.0, 1.0). Tables 1 and 2 list the credible intervals under small capture probabilities. The Jeffreys' prior Beta (0.5, 0.5) always produce reasonable credible intervals that cover the true N. However, when $N = 50$ or 100, the credible intervals under the uniform prior (Beta (1.0, 1.0)) fail to or barely cover the true N.

The simulation results on bias, MSE, and coverage probabilities are presented in Figs. 1, 2, 3, 4. The figures summarize the results when the sampling occasion $k = 3, 5, 6, 8$. Each figure contains three panels. The top, middle, and bottom panels correspond to the small, moderate, and large capture probabilities, respectively. In each panel, the left graph compares the bias, the middle one compares the square root of MSE, and the right one compares the 95 % frequentist coverage probability. In every graph, the x-axis denotes the true population sizes from 50 to 800. From the figures, we can see that under all the four priors, the bias decreases as the capture probabilities or the number of sampling occasions increase. This is reasonable because larger $\boldsymbol{\theta}$ or k means more data and thus more accurate estimates. We can also see that the four biases keep the same order under all different $(N, \boldsymbol{\theta}, k)$ combinations. It appears that when capture probabilities are small or moderate and $k \geq 5$, π_{RJ} and MLE perform best since they have the smallest bias. With large capture probabilities, π_{RJ} leads to the minimum bias when $k \geq 6$ and stays stable as k increases. In terms of coverage probability, π_{UJ}, π_{RJ}, and MLE have very similar performances and are better than the other two priors. The MSEs under the four priors and MLE are very close to each other.

4.2 Real data analysis

Example 1 Least Chipmunk. V. Reid (as reported in Otis et al. (1978), and re-analyzed by Wang et al. (2007)) did a capture–recapture experiment on the Least chipmunk *Eutamias minimus* in 1975. In the study, a 9×11 livetrapping grid was set up spaced

Table 1 95 % credible interval under small probability setting: $k = 8, 7, 6$

k	N	$a=b=0.25$		$a=b=0.5$		$a=b=0.75$		$a=b=1$	
		L	U	L	U	L	U	L	U
8	50	32.36	84.71	29.51	60.60	27.91	51.42	26.93	46.35
	100	71.14	133.73	67.30	118.40	64.40	107.84	62.25	100.40
	150	112.34	188.05	107.87	174.34	104.18	163.50	101.24	155.10
	200	154.92	242.64	150.02	229.66	145.78	218.71	142.25	209.79
	250	197.95	295.91	192.77	283.49	188.13	272.63	184.20	263.49
	300	241.89	349.67	236.43	337.55	231.49	326.71	227.18	317.40
	350	286.59	403.75	280.91	391.82	275.69	380.98	271.10	371.51
	400	331.40	457.06	325.55	445.33	320.12	434.53	315.27	424.99
	500	421.85	563.00	415.74	551.58	409.97	540.89	404.74	531.28
	600	512.78	667.68	506.49	656.52	500.49	645.98	495.00	636.35
	700	603.08	770.11	596.66	759.20	590.51	748.83	584.84	739.27
	800	697.06	877.12	690.46	866.28	684.10	855.87	678.20	846.22
7	50			27.51	62.92	25.80	51.54	24.82	45.68
	100	68.19	143.80	63.81	122.59	60.63	109.26	58.35	100.39
	150	108.12	197.14	103.06	179.24	98.99	165.76	95.83	155.72
	200	149.53	251.90	143.95	235.26	139.22	221.69	135.37	211.00
	250	191.68	305.56	185.77	289.89	180.56	276.51	176.21	265.51
	300	234.62	359.50	228.41	344.35	222.84	331.08	218.08	319.89
	350	278.11	413.10	271.67	398.40	265.80	385.25	260.71	373.96
	400	321.99	466.68	315.37	452.25	309.25	439.17	303.85	427.79
	500	411.13	573.88	404.14	559.88	397.61	546.90	391.77	535.41
	600	500.75	679.05	493.55	665.48	486.75	652.73	480.57	641.26
	700	590.59	783.41	583.22	770.12	576.18	757.52	569.77	746.07
	800	682.45	889.31	674.89	876.20	667.65	863.70	661.00	852.26
6	50			25.75	67.35	23.99	52.51	23.01	45.75
	100	65.51	158.64	60.61	128.64	57.20	111.88	54.88	101.42
	150	104.16	210.39	98.50	186.59	94.10	169.75	90.77	157.77
	200	144.56	265.07	138.34	243.65	133.18	226.91	129.08	214.15
	250	185.83	319.34	179.22	299.39	173.53	282.90	168.84	269.78
	300	227.61	372.77	220.70	353.95	214.60	337.79	209.47	324.53
	350	270.26	427.13	263.06	408.90	256.60	392.88	251.08	379.44
	400	313.27	480.79	305.86	463.15	299.10	447.36	293.27	433.86
	500	399.06	585.11	391.41	568.48	384.26	553.21	377.97	539.84
	600	488.62	693.97	480.59	677.67	473.04	662.47	466.32	648.99
	700	577.09	799.40	568.84	783.44	560.98	768.43	553.93	754.95
	800	666.57	904.41	658.15	888.79	650.08	873.96	642.77	860.54

50 feet apart and Least chipmunks were trapped once a day, for 6 consecutive days ($k = 6$). Otis et al. (1978) suggested using the M_t model based on a discrimination procedure. Using Bayes factor, we can show that M_t model fits the data at least better

Table 2 95 % credible interval under small probability setting: k = 5,4,3

k	N	$a = b = 0.25$		$a = b = 0.5$		$a = b = 0.75$		$a = b = 1$	
		L	U	L	U	L	U	L	U
5	50			22.61	76.71	20.69	53.14	19.74	44.53
	100			54.56	144.22	50.56	115.98	48.07	101.34
	150			90.05	205.79	84.65	177.44	80.86	159.91
	200	135.74	304.08	127.51	263.78	121.10	236.72	116.33	218.00
	250	175.00	357.04	166.23	321.00	159.08	294.77	153.53	275.45
	300	214.98	408.89	205.87	377.07	198.13	351.67	191.94	332.09
	350	255.50	461.77	246.07	432.18	237.88	407.55	231.17	387.93
	400	297.18	516.93	287.42	488.54	278.78	464.23	271.58	444.45
	500	379.67	622.37	369.50	595.98	360.26	572.58	352.38	552.91
	600	465.60	732.17	454.93	706.70	445.08	683.59	436.56	663.78
	700	552.72	840.86	541.70	816.18	531.42	793.47	522.40	773.69
	800	641.10	950.17	629.74	926.01	619.06	903.53	609.58	883.72
4	50			19.29	103.99	17.17	55.88	16.24	43.85
	100			48.03	192.61	43.10	126.09	40.48	102.93
	150			80.54	257.14	73.64	193.86	69.42	165.09
	200			114.86	311.64	106.68	256.67	101.21	226.05
	250			151.13	371.66	141.81	319.57	135.21	287.48
	300	199.69	488.49	187.19	421.32	177.32	375.76	169.99	344.52
	350	239.16	543.46	225.98	482.26	215.27	437.71	207.07	405.70
	450	317.37	645.58	303.61	594.18	291.82	552.84	282.43	521.10
	500	358.64	703.24	344.44	653.50	332.12	612.65	322.12	580.66
	600	437.32	801.59	422.98	757.62	410.17	719.69	399.52	688.88
	700	520.10	911.84	505.21	869.77	491.67	832.63	480.25	801.81
	800	606.10	1,025.16	590.68	984.49	576.52	947.92	564.40	917.09
3	50					14.08	62.65	13.35	45.13
	100					36.07	144.76	33.66	107.91
	150					62.65	223.91	58.49	175.16
	200					91.95	293.67	86.32	241.28
	250					123.47	362.42	116.47	307.50
	300					156.03	425.22	147.98	370.35
	350					190.02	485.68	181.04	432.61
	400			240.81	634.81	227.12	551.99	217.17	499.29
	500			314.38	749.80	299.16	674.13	287.72	621.86
	600	411.43	955.55	392.33	865.83	376.03	796.48	363.31	745.40
	700	490.32	1, 061.85	470.47	980.81	453.23	914.99	439.49	864.65
	800	564.79	1,151.19	545.02	1,079.07	527.49	1,018.23	513.23	970.31

than the simplest model M_0, which assumes constant capture probability throughout the experiment. Integrating out the parameters, we obtain the density functions of data under M_t as

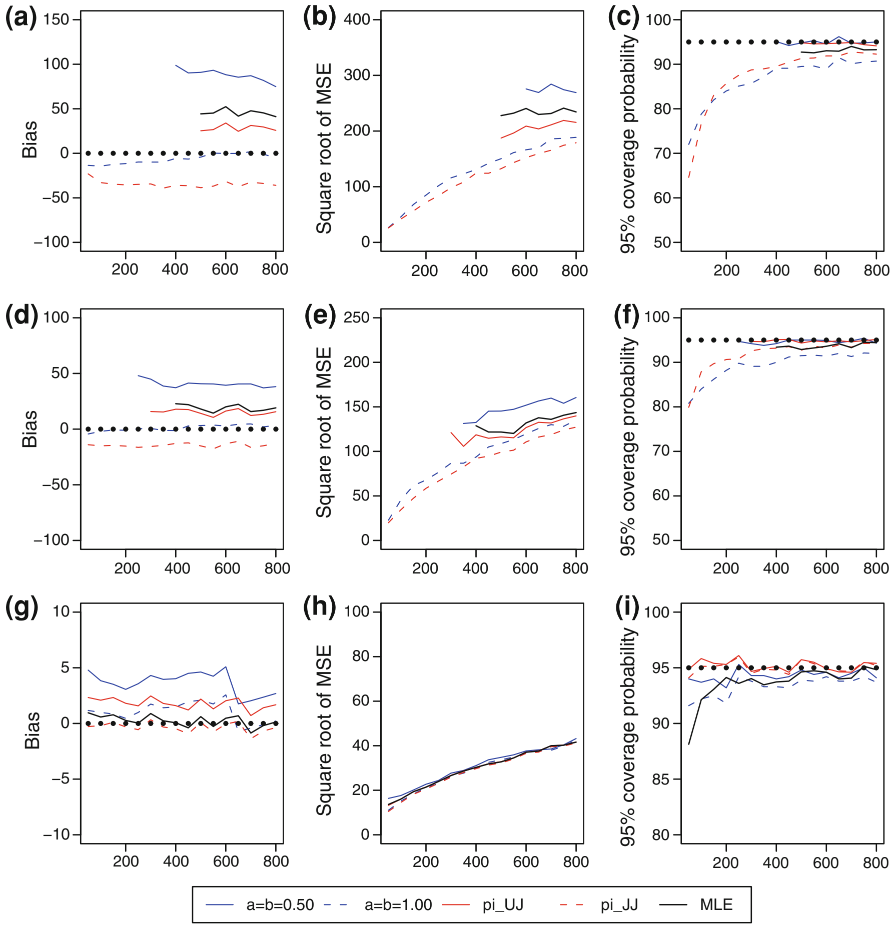

Fig. 1 Simulation results for $k = 3$. Four priors $\pi_{RU}, \pi_{RJ}, \pi_{JJ}, \pi_{UJ}$, and MLE are compared. The *top*, *middle*, and *bottom panel* corresponds to the small, moderate, and large capture probabilities. In each panel, the *left graph* compares the bias, the *middle graph* compares the square root of MSE, and the *right graph* compares the 95 % coverage probability. In each graph, the x-axis denotes the true values of N from 50 to 800

$$f_t(Data \mid M_t) = \sum_{N=n}^{\infty} \pi^R(N) \int f_2(N, \boldsymbol{\theta} \mid n_1, \ldots, n_L)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$
$$= \sum_{N=n}^{\infty} \frac{\pi^R(N)N!}{D\pi^k(N-n)!} \prod_{i=1}^{k} \frac{\Gamma(m_i + 1/2)\Gamma(N - m_i + 1/2)}{\Gamma(N+1)}, \qquad (33)$$

where $D = n_1!n_2! \cdots n_{L-1}!$. On the other hand, M_0 model assumes that the capture probability remains constant during the experiment and has the likelihood function

$$L_0(N, \theta \mid n_1, \ldots, n_L) = \frac{N!}{D(N-n)!} \theta^{\sum m_i} (1-\theta)^{kN - \sum m_i}. \qquad (34)$$

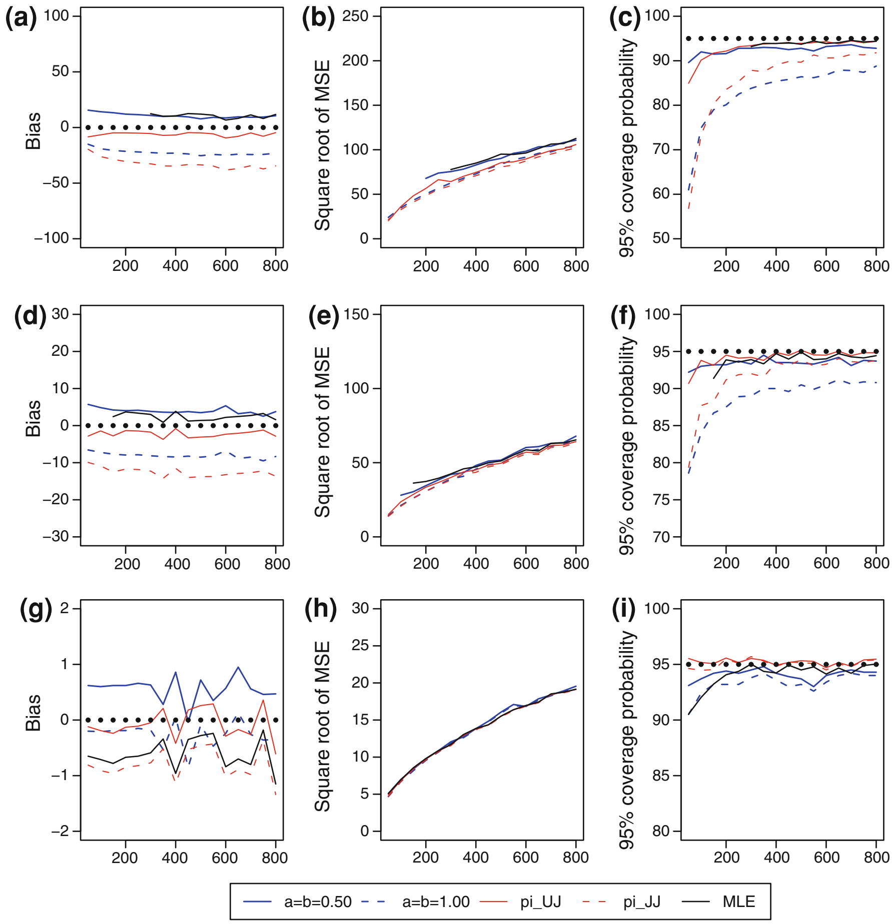

Fig. 2 Simulation results for $k = 5$. Four priors $\pi_{RU}, \pi_{RJ}, \pi_{JJ}, \pi_{UJ}$, and MLE are compared. The *top*, *middle*, and *bottom panel* corresponds to the small, moderate, and large capture probabilities. In each panel, the *left graph* compares the bias, the *middle graph* compares the square root of MSE, and the *right graph* compares the 95 % coverage probability. In each graph, the x-axis denotes the true values of N from 50 to 800

The density function of data under M_0 is

$$f_0(Data \mid M_0) = \sum_{N=n}^{\infty} \frac{\pi^R(N) N!}{\pi D(N-n)!} \frac{\Gamma(\sum m_i + 1/2)\Gamma(kN - \sum m_i + 1/2)}{\Gamma(kN+1)}. \tag{35}$$

The Bayes factor of M_t over M_0 is defined by $\frac{f(Data|M_t)}{f(Data|M_0)}$.

The capture history data, after being processed, are given in Table 3. Since k is relatively large, we choose the prior $\pi^R(N)$ and Jeffrey prior for $\boldsymbol{\theta}$ ($a = b = 0.5$). This combination of priors is denoted by π_{RJ}. The Bayes factor under π_{RJ} is 2.81,

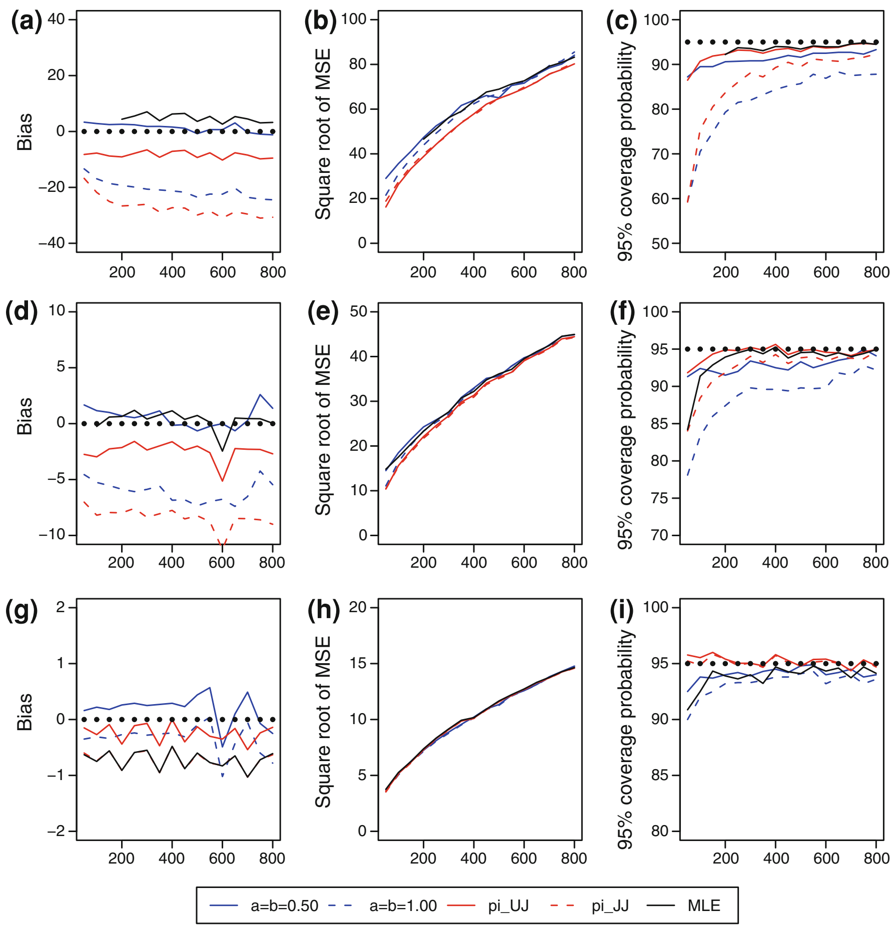

Fig. 3 Simulation results for $k = 6$. Four priors $\pi_{RU}, \pi_{RJ}, \pi_{JJ}, \pi_{UJ}$, and MLE are compared. The *top*, *middle*, and *bottom panel* corresponds to the small, moderate, and large capture probabilities. In each panel, the *left graph* compares the bias, the *middle graph* compares the square root of MSE, and the *right graph* compares the 95 % coverage probability. In each graph, the x-axis denotes the true values of N from 50 to 800

indicating M_t fits the data better. We compare π_{RJ} with the ad-hoc priors π_{UU}, π_{UJ}, π_{JU}, and π_{JJ} proposed in Wang et al. (2007). The first subscript indicates the prior for N, where 'U' and 'J' stand for the uniform prior $\pi(N) \propto 1$ and the Jeffreys prior $\pi(N) = N^{-1}$, respectively. The second subscript indicates the prior for $\boldsymbol{\theta}$, where 'U' and 'J' stand for the uniform prior $\pi(\boldsymbol{\theta}) \propto 1$ and the Jeffreys prior $\pi(\boldsymbol{\theta}) = \prod_{j=1}^{k} \theta_j^{-1/2}(1-\theta_j)^{-1/2}$. We present our results, as well as Wang et al. (2007)'s, in Table 4. Clearly, although all five estimates are very close to each other, π_{RJ} produces the shortest 95 % credible interval. The estimated capture probabilities under π_{RJ} are (0.137, 0.294, 0.314, 0.471, 0.373, 0.137).

Example 2 Turtle. In summer 2011, the research team in the department of fisheries and wildlife sciences at the University of Missouri conducted a mark–recapture study

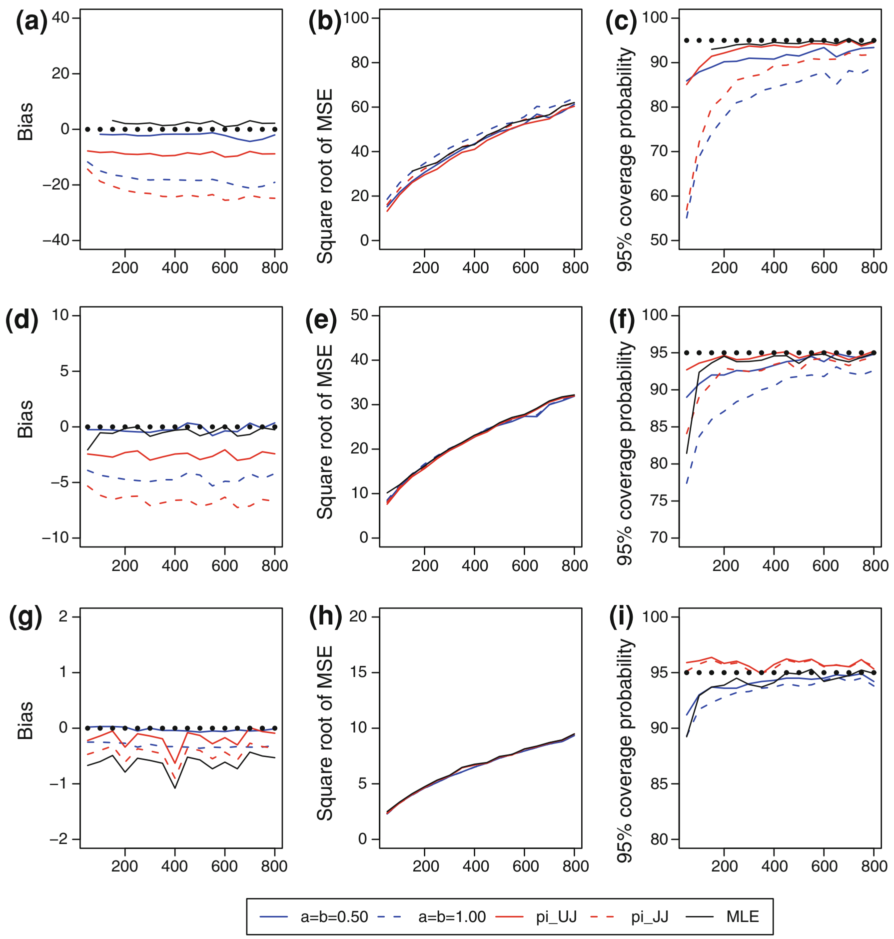

Fig. 4 Simulation results for $k = 8$. Four priors $\pi_{RU}, \pi_{RJ}, \pi_{JJ}, \pi_{UJ}$, and MLE are compared. The *top*, *middle*, and *bottom panel* corresponds to the small, moderate, and large capture probabilities. In each panel, the *left graph* compares the bias, the *middle graph* compares the square root of MSE, and the *right graph* compares the 95 % coverage probability. In each graph, the x-axis denotes the true values of N from 50 to 800

Table 3 Least Chipmunk data by V. Reid (1975)

Sampling occasion (i)	1	2	3	4	5	6
Animals caught (m_i)	7	15	16	24	19	7
Total caught	45					

of smooth softshell turtles on the Missouri River for eight trap nights ($k = 8$). The purpose of this study was to compare the specie abundance at sites where turtles are commercially harvested and where they are not. The Bayes factor of M_t and M_0 is 5.17, favoring M_t model. The sufficient statistics are summarized in Table 5. The estimates of N, 95 % credible intervals, and standard deviations under 5 priors and the MLE are summarized in Table 6. The estimates appear to be dependent on the

Table 4 Least Chipmunk example results. π_{UJ}, π_{JJ}, π_{UU}, and π_{JU} denote the four combinations of non-informative priors studied in Wang et al. (2007)

	Estimate of N	S.d.	95 % C.I.
π_{UJ}	53	3.54	(47, 61)
π_{JJ}	51	3.27	(46, 59)
π_{UU}	52	3.78	(47, 60)
π_{JU}	51	3.15	(46, 58)
MLE	50	3.14	(44, 56)
π_{RJ}	51	3.38	(46, 56)

The first subscript represents the prior for N, and the second subscript represents the prior for $\boldsymbol{\theta}$ with U uniform prior and J Jeffreys prior

Table 5 Smooth softshell Turtle data (2011)

Sampling occasion (i)	1	2	3	4	5	6	7	8
Animals caught (m_i)	12	14	18	8	23	26	22	24
Total caught	115							

Table 6 Smooth softshell Turtle example results

	Estimate of N	S.d.	95 % C.I.
π_{UJ}	245	34	(190, 322)
π_{JJ}	240	32	(188, 315)
π_{UU}	230	29	(182, 297)
π_{JU}	227	28	(180, 291)
MLE	251	35	(182, 320)
π_{RJ}	247	34	(191, 325)

π_{UJ}, π_{JJ}, π_{UU}, and π_{JU} denote the four combinations of non-informative priors studied in Wang et al. (2007). The first subscript represents the prior for N, and the second subscript represents the prior for $\boldsymbol{\theta}$ with U uniform prior and J Jeffreys prior

prior specification and in the same order as in the simulation. The estimated capture probabilities under π_{RJ} are (0.047, 0.056, 0.073,0.032, 0.093, 0.104, 0.090, 0.097), corresponding to the small capture probabilities case in the simulation. According to the upper panel in Fig. 4, both the MLE and π_{RJ} will give accurate estimates.

5 Comments

In summary, we developed a set of asymptotic reference priors for discrete parameter N in the capture–recapture model M_t and compared them with several commonly used non-informative priors and MLE. Based on the simulation results, we recommend using π_{RJ}, i.e., asymptotic reference prior for N and Jeffreys prior for $\boldsymbol{\theta}$ when k is relative large ($k \geq 6$) because of the small bias under this prior. For smaller k, π_{JU} performs well, resulting in small biases and accurate coverage probabilities.

The methodology we developed in this paper can be easily extended to other simple capture–recapture models such as M_0 and M_b, for they have similar likelihood functions as M_t. It will be more challenging to extend our method to M_h model, as constraints or hierarchical structures must be introduced on $\boldsymbol{\theta}$ to ensure the identifiability of N, making the likelihood function more complicated. Furthermore, our method can be applied to epidemiology research. For example, Seber et al. (2000) and Wang et al. (2005) analyzed the patient list mismatch data in medical records with capture–recapture models. The objective Bayesian method in our paper can be applied in this problem and potentially lead to more accurate estimates of the number of patients.

Appendix A: Proof of Theorem 1

The likelihood function (2) can be written as the product of the probability mass functions of iid random vectors $\boldsymbol{e}_j$'s, $j = 1, \ldots, N$, where $\boldsymbol{e}_j = (e_{j0}, e_{j1}, \cdots, e_{jL})$ and

$$\boldsymbol{e}_j \mid \boldsymbol{\theta} \sim \text{Multi}(1; p_0(\boldsymbol{\theta}), \ldots, p_L(\boldsymbol{\theta})). \tag{36}$$

The logarithm of the density of $\boldsymbol{e}_j$ is

$$\log p(\boldsymbol{e}_j \mid \boldsymbol{\theta}) = \sum_{i=0}^{L} e_{ji} \log p_i(\boldsymbol{\theta}). \tag{37}$$

For convenience, define

$$C_j \doteq \left\{ s \in \{0, 1, \ldots, L = 2^k\} : \text{the } j\text{th binary digit (from left to right) of } s \text{ is } 1 \right\}$$

For any $r = 1, \ldots, k$, we have

$$\begin{aligned} &\frac{\partial^2}{\partial \theta_r^2} \sum_{i=0}^{L} e_{ji} \log p_i(\boldsymbol{\theta}) \\ &= \sum_{i=0}^{L} e_{ij} \sum_{j=1}^{k} \frac{\partial^2}{\partial \theta_r^2} \left(b_{ij} \log \theta_j + (1 - b_{ij}) \log(1 - \theta_j) \right) \\ &= -\sum_{i=0}^{L} e_{ij} \left(\frac{b_{ir}}{\theta_r^2} + \frac{1 - b_{ir}}{(1 - \theta_r)^2} \right). \end{aligned} \tag{38}$$

Since $\mathrm{E}(e_{ji}) = p_i(\boldsymbol{\theta})$, we have

$$\mathrm{E}\left(\frac{\partial^2}{\partial \theta_r^2} \sum_{i=0}^{L} e_{ji} \log p_i(\boldsymbol{\theta}) \right) = -\sum_{i=0}^{L} \left(\frac{p_i(\boldsymbol{\theta}) b_{ir}}{\theta_r^2} + \frac{p_i(\boldsymbol{\theta})(1 - b_{ir})}{(1 - \theta_r)^2} \right). \tag{39}$$

Let $Q_{ir} = \frac{p_i(\boldsymbol{\theta})b_{ir}}{\theta_r^2} + \frac{p_i(\boldsymbol{\theta})(1-b_{ir})}{(1-\theta_r)^2}$. If $b_{ir} = 0$,

$$Q_{ir} = \frac{p_i(\boldsymbol{\theta})}{(1-\theta_r)^2} = (1-\theta_r)^{-1} \prod_{j \neq r} \theta_j^{b_{ij}} (1-\theta_j)^{1-b_{ij}}. \tag{40}$$

If $b_{ir} = 1$,

$$Q_{ir} = \frac{p_i(\boldsymbol{\theta})}{\theta_r^2} = \theta_r^{-1} \prod_{j \neq r} \theta_j^{b_{ij}} (1-\theta_j)^{1-b_{ij}}. \tag{41}$$

Notice that $\prod_{j \neq r} \theta_j^{b_{ij}} (1-\theta_j)^{1-b_{ij}}$ does not depend on b_{ir} and equals 1. Therefore, we have

$$\mathrm{E}\left(\frac{\partial^2}{\partial \theta_r^2} \sum_{i=0}^{L} e_{ji} \log p_i(\boldsymbol{\theta}) \right) = \frac{1}{\theta_r(1-\theta_r)}. \tag{42}$$

Also, we have

$$\frac{\partial^2}{\partial \theta_s \theta_t} \sum_{i=0}^{L} e_{ji} \log p_i(\boldsymbol{\theta}) = 0. \tag{43}$$

Therefore, the Fisher information matrix for $\boldsymbol{\theta}$ is Diag $\left\{\theta_1^{-1}(1-\theta_1)^{-1}, \ldots, \theta_k^{-1} (1-\theta_k)^{-1}\right\}$, which yields the result. □

Appendix B: Proof of Lemma 1

B.1 Proof of (15)

We start from studying $J_1(N)$ to prove (15). Since $\frac{1}{(N+a+b+i)^2}$ is decreasing in i, for $i = 0, 1, \cdots$,

$$\frac{k}{(N+a+b+i+1)^2} < \int_i^{i+1} \frac{k}{(N+a+b+x)^2} dx < \frac{k}{(N+a+b+i)^2}.$$

Adding up the inequality from $i = 0$ to infinity, we get

$$0 \leq \sum_{i=0}^{\infty} \frac{k}{(N+a+b+i)^2} - \int_0^{\infty} \frac{k}{(N+a+b+x)^2} dx \leq \frac{k}{(N+a+b)^2}. \tag{44}$$

The integral in (44) is $\frac{k}{N+a+b}$. Therefore,

$$\frac{k}{N+a+b} \le \sum_{i=0}^{\infty} \frac{k}{(N+a+b+i)^2} \le \frac{k}{N+a+b} + \frac{k}{(N+a+b)^2}. \tag{45}$$

Apply (45) for $k = 1$ and $a + b = 1$, we have

$$\frac{1}{N+1} \le \sum_{i=0}^{\infty} \frac{1}{(N+1+i)^2} \le \frac{1}{N+1} + \frac{k}{(N+1)^2}. \tag{46}$$

Combing (45) and (46), we can bound $J_1(N)$ by

$$\begin{aligned} \frac{k}{N+a+b} - \frac{1}{N+1} - \frac{1}{(N+1)^2} &\le J_1(N) \\ \le \frac{k}{N+a+b} - \frac{1}{N+1} + \frac{k}{(N+a+b)^2}&. \end{aligned} \tag{47}$$

$J_1(N)$ does not depend on the data, so $\mathrm{E}(J_1(N)) = J_1(N)$. So (15) follows from (47) immediately.

B.2 Proofs of (17) and (19)

The asymptotic order of $\mathrm{E}(J_2(N))$ depends on the value of b and is summarized in (17) and (19). To prove these equations, we need the following result from Chao and Strawderman (1972). For $X \sim Bin(N, p)$,

$$\mathrm{E}\left(\frac{1}{X+1}\right) = \frac{1-(1-p)^{N+1}}{(N+1)p}.$$

For all $1 \le l \le k$, since $m_l \mid N, \theta_l \sim \mathrm{Bin}(N, 1-\theta_l)$, we can apply the above result and get

$$\begin{aligned} \frac{1}{k}\mathrm{E}(J_2(N)) &= \mathrm{E}\left[\mathrm{E}\left(\frac{1}{N-m_1+1} \mid \theta_1\right)\right] \\ &= \int_0^1 \frac{1-\theta_1^{N+1}}{(N+1)(1-\theta_1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta_1^{a-1}(1-\theta_1)^{b-1} d\theta_1 \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)(N+1)} \sum_{j=0}^{N} \int_0^1 \theta_1^{j+a-1}(1-\theta_1)^{b-1} d\theta_1 \\ &= \frac{\Gamma(a+b)}{\Gamma(a)(N+1)} \sum_{j=0}^{N} \frac{\Gamma(j+a)}{\Gamma(j+a+b)}. \end{aligned} \tag{48}$$

Therefore, we have the following decomposition which will be used for several times throughout the proof of Lemma 1.

$$\frac{1}{k}\mathrm{E}\left(J_2(N)\right) = J_{21}(N) + J_{22}(N) + J_{23}(N), \tag{49}$$

where

$$J_{21}(N) = \frac{\Gamma(a+b)}{\Gamma(a)(N+1)} \int_0^N \frac{1}{(x+a)^b} dx, \tag{50}$$

$$J_{22}(N) = \frac{\Gamma(a+b)}{\Gamma(a)(N+1)} \left(\sum_{j=0}^{N} \frac{1}{(j+a)^b} - \int_0^N \frac{1}{(x+a)^b} dx \right), \tag{51}$$

$$J_{23}(N) = \frac{\Gamma(a+b)}{\Gamma(a)(N+1)} \sum_{j=0}^{N} \left(\frac{\Gamma(j+a)}{\Gamma(j+a+b)} - \frac{1}{(j+a)^b} \right). \tag{52}$$

Next, we will study the limiting behaviors of these three terms. The order of $J_{21}(N)$ can be found out straightforwardly. When $0 < b < 1$, $J_{21}(N)$ can be written as

$$J_{21}(N) = \frac{\Gamma(a+b)}{(1-b)\Gamma(a)(N+1)} \left((N+a)^{1-b} - a^{1-b} \right) \sim \frac{1}{N^b}. \tag{53}$$

When $b = 1$,

$$J_{21}(N) = \frac{\Gamma(a+b)}{(1-b)\Gamma(a)(N+1)} \log\left(\frac{N+a}{a}\right) \sim \frac{\log(N+a)}{N+1}. \tag{54}$$

For the second term $J_{22}(N)$, we show that it is asymptotically smaller than N^{-1}. For any $0 < a, b \leq 1$, due to the monotonicity of $(x+a)^{-b}$, we can bound the integral as

$$\sum_{j=1}^{N} \frac{1}{(j+a)^b} < \int_0^N \frac{1}{(x+a)^b} dx < \sum_{j=0}^{N-1} \frac{1}{(j+a)^b}, \tag{55}$$

which is equivalent to

$$\frac{1}{(N+a)^b} < \sum_{j=0}^{N} \frac{1}{(j+a)^b} - \int_0^N \frac{1}{(x+a)^b} dx < \frac{1}{a^b}. \tag{56}$$

Therefore, $J_{22}(N)$ can be bounded as

$$\frac{\Gamma(a+b)}{\Gamma(a)(N+1)(N+a)^b} < J_{22}(N) < \frac{\Gamma(a+b)}{\Gamma(a)a^b(N+1)}, \tag{57}$$

where the upper bound is in the order of $1/N$.

Lastly, we prove that third term $J_{23}(N)$ approaches $1/N$ as $N \to \infty$ for any $0 < a, b \leq 1$. It follows from Erdelyi and Tricomi (1951) that for fixed $\alpha, \beta \geq 0$, as $z \to \infty$,

$$\frac{\Gamma(z+\alpha)}{\Gamma(z+\beta)} = z^{\alpha-\beta}\left(1 + \frac{(\alpha-\beta)(\alpha+\beta-1)}{2z} + O\left(\frac{1}{z^2}\right)\right). \tag{58}$$

Applying this expansion with $z = j + a$, $\alpha = 0$, and $\beta = b$,

$$\frac{\Gamma(j+a)}{\Gamma(j+a+b)} = \frac{1}{(j+a)^b}\left(1 + \frac{b(1-b)}{2(j+a)} + J_{23}^*(j)\right), \tag{59}$$

where $J_{23}^*(j) \sim O\left((j+a)^{-2}\right)$, as $j \to \infty$. Thus, we can rewrite $J_{23}(N)$ as

$$J_{23}(N) = \frac{\Gamma(a+b)}{\Gamma(a)(N+1)}\left(\frac{b(1-b)}{2}\sum_{j=0}^{N}\frac{1}{(j+a)^{1+b}} + \sum_{j=0}^{N}\frac{J_{23}^*(j)}{(j+a)^b}\right).$$

The facts that

$$\sum_{j=0}^{\infty}\frac{1}{(j+a)^{1+b}} < \infty \quad \text{and} \quad \sum_{j=0}^{\infty}\frac{J_{23}^*(j)}{(j+a)^b} < \infty$$

ensure that as $N \to \infty$,

$$J_{23}(N) \sim \frac{1}{N}. \tag{60}$$

Applying (53), (54), (57), and (60) to (49) will prove (17) and (19). □

B.3 Proofs of (18) and (20)

Note that $J_3(N)$ can be bounded from both sides. On one hand,

$$\begin{aligned}
J_3(N) &\leq \sum_{l=1}^{k}\left[\frac{1}{(N-m_l+b)^2} + \int_0^{\infty}\frac{1}{(N-m_l+b+x)^2}dx\right] - \sum_{l=1}^{k}\frac{1}{N-m_l+1} \\
&= \sum_{l=1}^{k}\frac{1}{(N-m_l+b)^2} + \sum_{l=1}^{k}\left(\frac{1}{N-m_l+b} - \frac{1}{N-m_l+1}\right) \\
&= \sum_{l=1}^{k}\frac{1}{(N-m_l+b)^2} + \sum_{l=1}^{k}\frac{1-b}{(N-m_l+b)(N-m_l+1)}
\end{aligned}$$

$$\leq \sum_{l=1}^{k} \frac{1}{(N-m_l+b)^2} + \sum_{l=1}^{k} \frac{1-b}{(N-m_l+b)^2}$$
$$= \sum_{l=1}^{k} \frac{2-b}{(N-m_l+b)^2}. \tag{61}$$

On the other hand,

$$J_3(N) \geq \sum_{l=1}^{k} \left(\int_0^\infty \frac{1}{(N-m_l+b+x)^2} dx - \frac{1}{N-m_l+1} \right)$$
$$= \sum_{l=1}^{k} \frac{1-b}{(N-m_l+b)(N-m_l+1)} \geq \sum_{l=1}^{k} \frac{b(1-b)}{(N-m_l+b)^2}. \tag{62}$$

Let

$$J_3^*(N) = \mathrm{E}\left(\frac{1}{(N-m_1+b)^2}\right).$$

Inequalities (61) and (62) imply that $\mathrm{E}(J_3(N))/J_3^*(N)$ can be controlled by positive constants.

$$b(1-b)J_3^*(N) \leq \mathrm{E}(J_3(N)) \leq (2-b)J_3^*(N). \tag{63}$$

It suffices to find out the asymptotic order of $J_3^*(N)$. In fact,

$$J_3^*(N) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \frac{1}{(N-m_l+b)^2} \sum_{m_l=0}^{N} \binom{N}{m_l} \theta_l^{m_l+a-1}(1-\theta_l)^{N-m_l+b-1} d\theta_l$$
$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \sum_{i=0}^{N} \frac{\Gamma(N+1)\Gamma(i+a)\Gamma(N-i+b)}{(N-i+b)^2\Gamma(i+1)\Gamma(N-i+1)\Gamma(N+a+b)}, \tag{64}$$

where m_l is replaced by i for notation simplicity. Since $J_3^*(N)$ is a finite summation of positive components, it can be bounded from below by its last term,

$$J_3^*(N) \geq \frac{\Gamma(a+b)\Gamma(N+a)}{b^2\Gamma(a)\Gamma(N+a+b)} \approx \frac{\Gamma(a+b)}{b^2\Gamma(a)(N+a)^b}. \tag{65}$$

Now, we bound $J_3^*(N)$ from above. Applying (17) and (19) , we have for any $0 < b \leq 1$, as $N \to \infty$,

$$J_3^*(N) \leq \frac{1}{b^2}\mathrm{E}\left(\frac{1}{(N-m_1+1)}\right) \sim \frac{\log^{[b]} N}{(N+1)^b}. \tag{66}$$

It follows from (65) and (66) that, as $N \to \infty$, when $0 < b < 1$,

$$J_3^*(N) \sim N^{-b}. \tag{67}$$

When $b = 1$,

$$\frac{a}{N+a} \le J_3^*(N) \le \frac{C_1 \log(N+a)}{N+1}. \tag{68}$$

Combining (63), (67), and (68), we completed the proof of (18) and (20). □

B.4 Proof of (16)–Basis: $k = 2,\ 0 < b < 1$

$\mathrm{E}(J_4(N))$ is the leading term. Finding out its asymptotic order is the most challenging part of this proof. We will use induction method to study $\mathrm{E}(J_4(N))$ and prove (16). In subsections B.4 and B.5, we will focus on the base case (basis) when $k = 2$. Specifically, in B.4, we will show that (16) holds when $k = 2,\ 0 < b < 1$. In B.5, we will prove (16) when $k = 2,\ b = 1$ with a different approach. Finally, we will prove (16) for any arbitrary $k \ge 3$ (the inductive step) in B.6.

When $k = 2$, n follows a Binomial distribution:

$$n \mid N, \theta_1, \theta_2 \sim \mathrm{Bin}\left(N, 1 - (1-\theta_1)(1-\theta_2)\right).$$

Let $\delta_i = 1 - \theta_i$, $i = 1, 2$. For $0 < b < 1$, the expectation of $J_4(N)$ when $k = 2$, denoted by $\mathrm{E}_2(J_4(N))$, can be expressed as

$$\begin{aligned}
\mathrm{E}_2(J_4(N)) &= \sum_{n=0}^{N}\left(\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\right)\frac{\Gamma(N+1)}{\Gamma(n+1)\Gamma(N-n+1)}\\
&\quad \int_0^1\int_0^1 (1-\delta_1\delta_2)^n(\delta_1\delta_2)^{N-n}B^{-2}(a,b)(\delta_1\delta_2)^{b-1}(1-\delta_1)^{a-1}(1-\delta_2)^{a-1}d\delta_1 d\delta_2\\
&= \sum_{n=0}^{N}\left(\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\right)\frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(n+1)\Gamma(N-n+1)}\\
&\quad \int_0^1\int_0^1 (1-\delta_2+\delta_2(1-\delta_1))^n(\delta_2\delta_2)^{N-n+b-1}(1-\delta_1)^{a-1}(1-\delta_2)^{a-1}d\delta_1 d\delta_2\\
&= \sum_{n=0}^{N}\left(\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\right)\frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(n+1)\Gamma(N-n+1)}\sum_{j=0}^{n}\frac{\Gamma(n+1)}{\Gamma(j+1)\Gamma(n-j+1)}\\
&\quad \int_0^1\int_0^1 \delta_1^{N-n+b-1}(1-\delta_1)^{n-j+a-1}\delta_2^{N-j+b-1}(1-\delta_2)^{j+a-1}d\delta_1 d\delta_2\\
&= J_{41} + J_{42} + J_{43},
\end{aligned} \tag{69}$$

where for $l = 1, 2, 3$,

$$J_{4l}(N) = \frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=0}^{N} \sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \frac{\Gamma(N-n+b)}{\Gamma(N-n+1)} J^*_{4l}(n), \tag{70}$$

and

$$J^*_{41}(n) = \int_0^n \frac{1}{(x+1)^{1-a}(n-x+1)^{1-a}(N-x+1)^a} dx, \tag{71}$$

$$J^*_{42}(n) = \sum_{j=0}^{n} \frac{1}{(j+1)^{1-a}(n-j+1)^{1-a}(N-j+1)^a} - J^*_{41}(n), \tag{72}$$

$$J^*_{43}(n) = \sum_{j=0}^{n} \frac{\Gamma(j+a)\Gamma(n-j+a)\Gamma(N-j+b)}{\Gamma(j+1)\Gamma(n-j+1)\Gamma(N-j+a+b)} - J^*_{41}(n) - J^*_{42}(n). \tag{73}$$

Notice that the forms of $J^*_{42}(n)$, $J^*_{43}(n)$ are similar with those of $J_{22}(N)$, $J_{23}(N)$. Therefore, the asymptotic bounds of $J^*_{42}(n)$, $J^*_{43}(n)$ can be obtained by adopting the similar approaches in the proof of (51) and (52). We omit the details and present the results below:

$$J_{4l}(N) \leq \frac{1}{(N+1)^b}, \quad l = 2, 3. \tag{74}$$

The term $J_{41}(N)$ can be decomposed into two partial sums,

$$J_{41}(N) = J_{411}(N) + J_{412}(N), \tag{75}$$

where

$$J_{411}(N) = \frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=0}^{N-\log N-1} \sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \frac{\Gamma(N-n+b)}{\Gamma(N-n+1)} J^*_{41}(n), \tag{76}$$

$$J_{412}(N) = \frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=N-\log N}^{N} \sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \frac{\Gamma(N-n+b)}{\Gamma(N-n+1)} J^*_{41}(n). \tag{77}$$

Interestingly, although $J_{412}(N)$ has much less terms than $J_{411}(N)$, it is the leading term. In fact, for $k = 2$ and $0 < b < 1$, there exists positive constants C_2, C_3 such that as $N \to \infty$,

$$J_{411}(N) \leq \frac{C_2 \log^b N}{(N+1)^b}, \tag{78}$$

$$J_{412}(N) \sim \frac{C_3 B^{-2}(a,b) \log N}{(N+1)^b}. \tag{79}$$

For $J_{411}(N)$, let $y = x/n$ and we can bound $J^*_{41}(n)$ by

$$\begin{aligned} J^*_{41}(n) &= \frac{1}{n^{1-a}} \int_0^1 \frac{1}{(y+\frac{1}{n})^{1-a}(1-y+\frac{1}{n})^{1-a}(1-y+\frac{1}{n}+\frac{N-n}{n})^a} dy \\ &\le \frac{1}{n^{1-a}} \int_0^1 \frac{1}{(y+\frac{1}{n})^{1-a}(1-y+\frac{1}{n})} dy. \end{aligned} \tag{80}$$

The integral can be bounded as following. Notice that

$$\begin{aligned} 0 \le &\int_0^{1/2} \frac{1}{(y+1/n)^{1-a}(1-y+1/n)} dy \\ &\le \frac{1}{1/2+1/n} \int_0^{1/2} \frac{1}{(y+1/n)^{1-a}} dy \le \frac{2}{a}\left(\frac{1}{2}+\frac{1}{n}\right)^a \to \frac{2^{1-a}}{a}. \end{aligned} \tag{81}$$

Similarly, we have

$$\log n/2 \le \int_{1/2}^1 \frac{1}{(y+1/n)^{1-a}(1-y+1/n)} dy \le 2^{1-a} \log(1+n/2). \tag{82}$$

It follows from (80), (81), (82) that,

$$J^*_{41}(n) \le \frac{2^{1-a}\log(n+1)}{(n+1)^{1-a}}. \tag{83}$$

and consequently, applying the same decomposition and arguments as in (50), (51), (52), we have

$$\begin{aligned} J_{411}(N) &\le \frac{2B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=0}^{N-\log N} \frac{\Gamma(N-n+b)}{\Gamma(N-n+2)} \frac{2^{1-a}\log(n+1)}{(n+1)^{1-a}} \\ &\le \frac{2^{4-2a}B^{-2}(a,b)\log^b N}{(1-b)N^b}. \end{aligned} \tag{84}$$

Let $C_2 = \frac{2^{4-2a}B^{-2}(a,b)}{1-b}$ and we proved (78).

For $J_{412}(N)$, note that $J^*_{41}(n) \le \frac{\log(n+1)}{(n+1)^{1-a}}$, we get

$$\begin{aligned} J_{412}(N) \le &\frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=N-\log N}^{N} \left(\sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2}\right) \\ &\times \frac{\log(n+1)\Gamma(N-n+b)}{(n+a)^{1-a}\Gamma(N-n+1)} \\ &\le \frac{B^{-2}(a,b)\log(N+1)\Gamma(N+1)}{(N-\log N+1)^{1-a}\Gamma(N+a+b)} H(N), \end{aligned} \tag{85}$$

where

$$H(N) = \sum_{n=N-\log N}^{N} \left(\sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \right) \frac{\Gamma(N-n+b)}{\Gamma(N-n+1)}. \tag{86}$$

On the other hand, notice that

$$\begin{aligned} J_{41}^*(N) &\geq \frac{1}{n^{1-a}} \int_0^1 \frac{1}{\left(y+\frac{1}{n}+\frac{N-n}{n}\right)^{1-a} \left(1-y+\frac{1}{n}+\frac{N-n}{n}\right)} dy \\ &\geq \frac{1}{n^{1-a}\left(1+\frac{1}{n}+\frac{N-n}{n}\right)^{1-a}} \int_0^1 \frac{1}{\left(1-y+\frac{1}{n}+\frac{N-n}{n}\right)} dy \\ &= \frac{1}{(N+1)^{1-a}} \left(\log n - \log(N-n+1)\right). \end{aligned}$$

Since $n < N - \log N$ and $N - n > \log N$, we have

$$J_{41}^*(N) \geq \frac{1}{(N+1)^{1-a}} \left(\log(N-\log N) - \log(1+\log N)\right). \tag{87}$$

Consequently, we obtain a lower bound of $J_{412}(N)$ by

$$J_{412}(N) \geq B^{-2}(a,b) \frac{[\log(N-\log N) - \log(1+\log N)]\Gamma(N+1)}{(N+1)^{1-a}\Gamma(N+a+b)} H(N). \tag{88}$$

Let $m = N - n$ and we have

$$\begin{aligned} H(N) &= \sum_{m=0}^{\log N} \left(\sum_{i=0}^{\infty} \frac{1}{(m+1+i)^2} \right) \frac{\Gamma(m+b)}{\Gamma(m+1)} \\ &\leq \sum_{m=0}^{\log N} \frac{2\Gamma(m+b)}{(m+1)\Gamma(m+1)} \leq \sum_{m=0}^{\infty} \frac{2\Gamma(m+b)}{\Gamma(m+2)} \sim \sum_{m=0}^{\infty} \frac{1}{m^{2-b}} < \infty. \end{aligned} \tag{89}$$

Since $H(N)$ is an increasing function of N, the limit of $H(N)$ as $N \to \infty$ exists and we assume $\lim_{N\to\infty} H(N) = C_4$, where C_4 is a positive constant. Combining (85) and (88), we have

$$\begin{aligned} &\frac{C_4 B^{-2}(a,b)[\log(N-\log N) - \log(1+\log N)]\Gamma(N+1)}{(N+1)^{1-a}\Gamma(N+a+b)} \\ &\quad \leq J_{412}(N) \leq \frac{C_4 B^{-2}(a,b)\log(N+1)\Gamma(N+1)}{(N-\log N+1)^{1-a}\Gamma(N+a+b)}, \end{aligned} \tag{90}$$

which yields (79) immediately.

It follows from (75), (78), (79) that as $N \to \infty$,

$$J_{41}(N) \sim \frac{C_3 B^{-2}(a,b)\log N}{(N+1)^b}. \tag{91}$$

Combining with (69) and (74), it is clear that $J_{41}(N)$ is the leading term of $\mathrm{E}_2(J_4(N))$, hence (16) holds for $k=2,\ 0<b<1$. □

B.5 Proof of (16)–Basis: $k=2, b=1$

To show (16) for $k=2,\ b=1$, a different approach is required. Because $J_4(N)$ can be bounded by

$$\frac{1}{N-n+1} \le J_4(N) \le \frac{1}{N-n+1} + \frac{1}{(N-n+1)^2}, \tag{92}$$

it suffices to obtain the asymptotic orders of $\mathrm{E}\frac{1}{N-n+1}$ and $\mathrm{E}\frac{1}{(N-n+1)^2}$. Applying Chao and Strawderman (1972)'s result, the former expectation can be decomposed as

$$\begin{aligned}
\mathrm{E}_2\Big(\frac{1}{N-n+1}\Big) &= \mathrm{E}\Big(\mathrm{E}\Big(\frac{1}{N-n+1} \mid \boldsymbol{\theta}\Big)\Big)\\
&= B^{-2}(a,1)\int_0^1\int_0^1 \frac{1-\Big(1-\prod_{i=1}^2(1-\theta_i)\Big)^{N+1}}{(N+1)\prod_{i=1}^2(1-\theta_i)}\prod_{i=1}^2\theta_i^{a-1}d\theta_1 d\theta_2\\
&= B^{-2}(a,1)\int_0^1\int_0^1 \frac{1-\Big(1-\prod_{i=1}^2\delta_i\Big)^{N+1}}{(N+1)\prod_{i=1}^2\delta_i}\prod_{i=1}^2(1-\delta_i)^{a-1}d\delta_1 d\delta_2\\
&= \frac{B^{-2}(a,1)}{N+1}\sum_{j=0}^N\int_0^1\int_0^1 (1-\delta_2+\delta_2(1-\delta_1))^j\prod_{i=1}^2(1-\delta_i)^{a-1}d\delta_1 d\delta_2\\
&= \frac{B^{-2}(a,1)}{N+1}\sum_{j=0}^N\frac{\Gamma(j+1)}{\Gamma(j+a+1)}\sum_{i=0}^j\frac{\Gamma(i+a)\Gamma(j-i+a)}{\Gamma(i+1)\Gamma(j-i+a+1)}\\
&= J_{44}(N)+J_{45}(N)+J_{46}(N),
\end{aligned} \tag{93}$$

where for $l=4,5,6$,

$$J_{4l}(N) = \frac{B^{-2}(a,1)}{N+1}\sum_{j=0}^N\frac{\Gamma(j+1)}{\Gamma(j+a+1)}J^*_{4l}(j), \tag{94}$$

and

$$J_{44}^{*}(j) = \int_0^j \frac{1}{(x+a)^{1-a}(j-x+a)}dx, \tag{95}$$

$$J_{45}^{*}(j) = \sum_{i=0}^{j} \frac{1}{(i+a)^{1-a}(j-i+a)} - \int_0^j \frac{1}{(x+a)^{1-a}(j-x+a)}dx, \tag{96}$$

$$J_{46}^{*}(j) = \sum_{i=0}^{j} \frac{\Gamma(i+a)\Gamma(j-i+a)}{\Gamma(i+1)\Gamma(j-i+a+1)} - \sum_{i=0}^{j} \frac{1}{(i+a)^{1-a}(j-i+a)}. \tag{97}$$

The asymptotic orders or bounds of $J_{4l}(N)$, $l = 4, 5, 6$, and $\mathrm{E}\left(\frac{1}{(N-n+1)^2}\right)$ are summarized below. For fixed $0 < a \le 1$ and $b = 1$, sending N to infinity, we have

$$\frac{B^{-2}(a,1)\log^2 N}{2(N+1)} \le J_{44}(N) \le \frac{B^{-2}(a,1)\log^2 N}{2(N+1)} + \frac{B^{-2}(a,1)(\log N)^{3/2}}{a(N+1)}, \tag{98}$$

$$J_{4l}(N) \le \frac{B^{-2}(a,1)\log N}{N+1}, \quad l = 5, 6, \tag{99}$$

$$\mathrm{E}\left(\frac{1}{(N-n+1)^2}\right) \le \frac{2B^{-2}(a,1)\log(N+1)}{N}. \tag{100}$$

Due to the similarity of the proofs in nature, we will only prove (98), which gives the asymptotic order of the leading term. In fact, we let $y = x/j$ and write $J_{44}^{*}(j)$ as

$$J_{44}^{*}(j) = \frac{1}{j^{1-a}} \int_0^1 \frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy. \tag{101}$$

This integral can be shown to have an order of $\log j$ for j large enough. Let $\epsilon = (\log j)^{-1/2}$. First, we integrate from 0 to $1-\epsilon$.

$$\begin{aligned} \int_0^{1-\epsilon} \frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy &\le \frac{1}{\epsilon + a/j} \int_0^{1-\epsilon} \frac{1}{(y+a/j)^{1-a}}dy \\ &= \frac{1}{a(\epsilon+a/j)}\left[(1-\epsilon+a/j)^a - (a/j)^a\right] \le \frac{(1-\epsilon+a/j)^a}{a(\epsilon+a/j)}. \end{aligned} \tag{102}$$

Second, we integrate from $1-\epsilon$ to 1.

$$\begin{aligned} &\int_{1-\epsilon}^{1} \frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy \\ &\quad \le \frac{1}{(1-\epsilon+a/j)^{1-a}} \int_{1-\epsilon}^{1} \frac{1}{1-y+a/j}dy = \frac{1}{(1-\epsilon+a/j)^{1-a}} \log \frac{\epsilon j + a}{a}. \end{aligned} \tag{103}$$

On the other hand,

$$\int_{1-\epsilon}^{1} \frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy$$
$$\geq \frac{1}{(1+a/j)^{1-a}} \int_{1-\epsilon}^{1} \frac{1}{1-a+a/j}dy = \frac{1}{(1+a/j)^{1-a}} \log \frac{\epsilon j + a}{a}. \tag{104}$$

It follows from (81), (103), and (104) that

$$\frac{1}{(1+a/j)^{1-a}} \log \frac{\epsilon j + a}{a} \leq \int_{0}^{1} \frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy$$
$$\leq \frac{1}{(1-\epsilon+a/j)^{1-a}} \log \frac{\epsilon j + a}{a} + \frac{(1-\epsilon+a/j)^{a}}{\epsilon(\epsilon+a/j)}. \tag{105}$$

Clearly,

$$\lim_{j\to\infty} \frac{1}{(1+a/j)^{1-a}} \frac{\log(\epsilon j + a)/a}{\log j} = 1, \tag{106}$$

$$\lim_{j\to\infty} \frac{1}{(1-\epsilon+a/j)^{1-a}} \frac{\log(\epsilon j + a)/a}{\log j} = 1. \tag{107}$$

Also, by the definition of ϵ, we have

$$\frac{(1-\epsilon+a/j)^{a}}{a(\epsilon+a/j)\log j} \sim \frac{1}{a(\log j)^{1/2}}. \tag{108}$$

It follows from (105), (106), (107), and (108) that

$$\log j \leq \int_{0}^{1} \frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy \leq \log j + \frac{1}{a}(\log j)^{1/2}. \tag{109}$$

Combining (109) and (101), we have

$$\frac{\log j}{j^{1-a}} \leq J_{44}^{*}(j) \leq \frac{\log j}{j^{1-a}} + \frac{(\log j)^{1/2}}{a j^{1-a}}. \tag{110}$$

From (110), we can apply the same decomposition as in (50), (51), (52) and prove (16) for $k = 2,\ b = 1$. The details of the decomposition are omitted to reduce the redundancy of the proof. □

B.6 Proof of (16)–Inductive step

Here, we prove (16) for general $k \geq 3,\ 0 < b \leq 1$. Based on the discussions in B.4 and B.5, the asymptotic order of $\mathrm{E}_2(J_4(N))$ can be summarized as the following. For $k = 2$ and $0 < a \leq 1$,

(1) when $0 < b < 1$,

$$\frac{C_2 B^{-2}(a,b)\log(N+1)}{N^b} \le \mathrm{E}_2(J_4(N)) \le \frac{C_2 B^{-2}(a,b)\log(N+1)}{N^b} + \frac{\log^b N}{N^b}; \tag{111}$$

(2) when $b = 1$,

$$\frac{B^{-2}(a,1)\log^2 N}{2(N+1)} \le \mathrm{E}_2(J_4(N)) \le \frac{B^{-2}(a,1)\log^2 N}{2(N+1)} + \frac{B^{-2}(a,1)\log^{3/2} N}{a(N+1)}. \tag{112}$$

In general, we assume the capture–recapture experiment consists of $k+1$ sampling occasions. So

$$n \mid N, \boldsymbol{\theta} \sim \mathrm{Bin}\Big(N, 1 - \prod_{i=1}^{k+1}(1-\theta_i)\Big).$$

For any fixed $0 < a, b \le 1$ and $u = 1, \cdots, k+1$, we let $\delta_u = 1 - \theta_u$, $r = n - j$, and $m = N - j$. The expected value of $J_4(N)$ can be expressed as

$$\begin{aligned}
&\mathrm{E}_{k+1}(J_4(N)) \\
&\quad = \sum_{n=0}^{N}\sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2}\frac{N!}{n!(N-n)!}\int_0^1 \cdots \int_0^1 \left(1 - \prod_{u=1}^{k+1}\delta_u\right)^n \left(\prod_{u=1}^{k+1} x_u\right)^{N-n} \\
&B^{-(k+1)}(a,b)\prod_{u=1}^{k+1}\delta_u^{b-1}\prod_{u=1}^{k+1}(1-\delta_u)^{a-1}d\delta_1\cdots d\delta_{k+1} \\
&\quad = B^{-(k+1)}(a,b)\sum_{n=0}^{N}\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\frac{\Gamma(N+1)}{\Gamma(n+1)\Gamma(N-n+1)} \\
&\qquad \times \sum_{j=0}^{n}\frac{\Gamma(n+1)}{\Gamma(j+1)\Gamma(n-j+1)} \\
&\quad \int_0^1 \cdots \int_0^1 (1-\delta_{k+1})^j \delta_{k+1}^{n-j}\left(1 - \prod_{u=1}^{k}\delta_u\right)^{n-j} \delta_u^{N-n+b-1} \\
&\qquad \times \prod_{u=1}^{k+1}(1-\delta_u)^{a-1}d\delta_1\cdots d\delta_{k+1}.
\end{aligned}$$

In this way, we can integrate out δ_{k+1} and get the recursive formula.

$$\mathrm{RHS} = \frac{B^{-(k+1)}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)}\sum_{n=0}^{N}\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\frac{1}{\Gamma(N-n+1)}$$

$$
\begin{aligned}
&\times \sum_{j=0}^{n} \frac{\Gamma(j+a)\Gamma(N-j+b)}{\Gamma(j+1)\Gamma(n-j+1)} \\
&\int_0^1 \cdots \int_0^1 \left(1-\prod_{u=1}^{k}\delta_u\right)^{n-j} \prod_{u=1}^{k}\delta_u^{N-n+b-1}\prod_{u=1}^{k}(1-\delta_u)^{a-1}d\delta_1\cdots d\delta_k \\
&= \frac{B(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{m=0}^{N} \frac{\Gamma(N-m+a)\Gamma(m+b)}{\Gamma(N-m+1)\Gamma(m+1)} \sum_{r=0}^{m}\left(\sum_{i=0}^{\infty}\frac{1}{(m-r+1+i)^2}\right) \\
&\int_0^1 \cdots \int_0^1 \frac{B^{-k}(a,b)\Gamma(m+1)}{\Gamma(r+1)\Gamma(m-r+1)}\left(1-\prod_{u=1}^{k}\delta_u\right)^{r} \\
&\times \prod_{u=1}^{k}\delta_u^{m-r+b-1}\prod_{u=1}^{k}(1-\delta_u)^{a-1}d\delta_1\cdots d\delta_k \\
&= \frac{B(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{m=0}^{N} \frac{\Gamma(m+b)\Gamma(N-m+a)}{\Gamma(m+1)\Gamma(N-m+1)} \mathrm{E}_k(J_4(m)),
\end{aligned}
$$

where $\mathrm{E}_k(J_4(N))$ denotes the expectation of $J_4(N)$ when there are k sampling occasions. For fixed $k \geq 3$, $0 < a \leq 1$, and $0 < b < 1$, we assume that for some constants $C_5(k)$, $C_6(k)$ depending on k, such that as $m \to \infty$,

$$
\left|\mathrm{E}_k(J_4(m)) - \frac{C_2 B^{-k}(a,b)\log^{k-1}(m)}{m^b}\right| \leq \frac{C_5(k)\log^{k-2+b}(m)}{m^b}. \tag{113}
$$

For $b = 1$, we assume that

$$
\left|\mathrm{E}_k(J_4(m)) - \frac{B^{-k}(a,1)\log^{k}(m)}{m}\right| \leq \frac{C_6(k)\log^{k-1/2}(m)}{m}. \tag{114}
$$

Clearly, assumptions (113) and (114) hold for $k = 2$.

For $k \geq 3$, we will show the following results that directly lead to (16). Assuming $k+1$ sampling occasions and assumptions (113) and (114), we have for $0 < b < 1$,

$$
\left|\mathrm{E}_{k+1}(J_4(N)) - \frac{C_2 B^{-(k+1)}(a,b)\log^{k}(N)}{N^b}\right| \leq \frac{C_5(k+1)\log^{k-1+b}(N)}{N^b}. \tag{115}
$$

For $b = 1$,

$$
\left|\mathrm{E}_{k+1}(J_4(N)) - \frac{B^{-(k+1)}(a,1)\log^{k+1}(N)}{N}\right| \leq \frac{C_6(k+1)\log^{k+1/2}(N)}{N}. \tag{116}
$$

We focus on the case when $0 < b < 1$ and prove (115). $\mathrm{E}_{k+1}(J_4(N))$ can be decomposed after removing the first $M+1$ terms (M is a fixed integer).

$$
\begin{aligned}
\mathrm{E}_{k+1}(J_4(N)) = {} & \frac{C_2 B^{-(k+1)}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{m=M}^{N} \frac{\Gamma(m+b)\Gamma(N-m+a)\log^{k-1} m}{m^b \Gamma(m+1)\Gamma(N-m+1)} \\
& + \frac{B(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{m=M}^{N} \frac{\Gamma(m+b)\Gamma(N-m+a)}{\Gamma(m+1)\Gamma(N-m+1)} \\
& \times \left(\mathrm{E}_k(J_4(m)) - \frac{C_2 B^{-k}(a,b)\log^{k-1} m}{m^b} \right) \\
\equiv {} & J_{4k1}(N) + J_{4k2}(N). \qquad (117)
\end{aligned}
$$

By the similar approach in the proof of (50), (51), (52), we can approximate $J_{4k1}(N)$ by

$$
J_{4k1}(N) \approx \frac{C_2 B^{-(k+1)}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} I_k(N), \qquad (118)
$$

where

$$
I_k(N) = \int_M^N \frac{\log^{k-1} x}{(x+1)(N-x+1)^{1-a}} dx. \qquad (119)
$$

Below, we show that for any $0 < b < 1$ and $k \geq 2$, there exist positive constants C_7 such that

$$
\frac{\log^k N}{N^{1-a}} - \frac{C_7 \log^{k-1} N}{N^{1-a}} \leq I_k(N) \leq \frac{\log^k N}{N^{1-a}}. \qquad (120)
$$

To prove (120), we first find the order of $I_k(N)$ for $k = 2$. Let $y = N - x$,

$$
\begin{aligned}
I_2(N) &= \int_M^N \frac{\log x}{(x+1)(N-x+1)^{1-a}} dx \\
&= \frac{1}{N^{1-a}} \int_{M/N}^{1} \frac{\log y + \log N}{(y+1/N)(1-y+1/N)^{1-a}} dy \\
&\sim \frac{\log^2 N}{N^{1-a}} + \frac{1}{N^{1-a}} \int_{M/N}^{1} \frac{\log y}{(y+1/N)(1-y+1/N)^{1-a}} dy \\
&\leq \frac{\log^2 N}{N^{1-a}}. \qquad (121)
\end{aligned}
$$

Now, we bound $I_2(N)$ from below. Notice that

$$
\begin{aligned}
& \frac{1}{N^{1-a}} \int_{M/N}^{1/2} \frac{\log y}{(y+1/N)(1-y+1/N)^{1-a}} dy \\
& \geq \frac{1}{(1/2+1/N)N^{1-a}} \int_{M/N}^{1/2} \frac{\log y}{(1-y+1/N)^{1-a}} dy
\end{aligned}
$$

$$\begin{aligned}
&\geq \frac{\log M - \log N}{(1/2 + 1/N)N^{1-a}} \int_{M/N}^{1/2} \frac{1}{(1-y+1/N)^{1-a}} dy \\
&= \frac{\log M - \log N}{a(1/2+1/N)N^{1-a}} \left[(1 - M/N + 1/N)^a - (1/2 + 1/N)^a \right] \\
&\approx \frac{\log M - \log N}{aN^{1-a}}. \qquad (122)
\end{aligned}$$

We also have

$$\begin{aligned}
&\frac{1}{N^{1-a}} \int_{1/2}^{1} \frac{\log y}{(y+1/N)(1-y+1/N)^{1-a}} dy \geq -\frac{\log 2}{N^{1-a}} \int_{1/2}^{1} \frac{1}{(1-y+1/N)^{1-a}} dy \\
&= -\frac{\log 2}{a(1/2+1/N)N^{1-a}} \left[(1/2+1/N)^a - (1/N)^a \right] \approx -\frac{\log 2}{aN^{1-a}}. \qquad (123)
\end{aligned}$$

It follows from (121), (122), and (123) that

$$\frac{\log^2 N}{N^{1-a}} + \frac{\log M - \log N}{aN^{1-a}} - \frac{\log 2}{aN^{1-a}} \leq I_2(N) \leq \frac{\log^2 N}{N^{1-a}}. \qquad (124)$$

For any fixed $k \geq 2$, we obtain the upper bound of $I_{k+1}(N)$.

$$\begin{aligned}
I_{k+1}(N) &= \int_{M}^{N} \frac{\log^k x}{(x+1)(N-x+1)^{1-a}} dx \\
&= \frac{1}{N^{1-a}} \int_{M/N}^{1} \frac{(\log y + \log N)^k}{(y+1/N)(1-y+1/N)^{1-a}} dy \\
&= \frac{1}{N^{1-a}} \int_{M/N}^{1} \frac{(\log y + \log N)^{k-1}(\log y + \log N)}{(y+1/N)(1-y+1/N)^{1-a}} dy \\
&= I_k(N) \log N + \frac{1}{N^{1-a}} \int_{M/N}^{1} \frac{(\log y + \log N)^{k-1} \log y}{(y+1/N)(1-y+1/N)^{1-a}} dy \\
&\leq I_k(N) \log N. \qquad (125)
\end{aligned}$$

Next, we seek the lower bound of $I_{k+1}(N)$. First of all,

$$\begin{aligned}
&\frac{1}{N^{1-a}} \int_{M/N}^{1/2} \frac{(\log y + \log N)^{k-1} \log y}{(y+1/N)(1-y+1/N)^{1-a}} dy \\
&\geq \frac{(\log M - \log N) \log^{k-1} M}{N^{1-a}} \int_{M/N}^{1/2} \frac{1}{(y+1/N)(1-y+1/N)^{1-a}} dy \\
&\geq \frac{(\log M - \log N) \log^{k-1} M}{(1/2+1/N)N^{1-a}} \int_{M/N}^{1/2} \frac{1}{(1-y+1/N)^{1-a}} dy \\
&= \frac{(\log M - \log N) \log^{k-1} M}{a(1/2+1/N)N^{1-a}} \left[(1 - M/N + 1/N)^a - (1/2+1/N)^a \right] \\
&\approx \frac{2(1-2^{-a})(\log M - \log N) \log^{k-1} M}{aN^{1-a}}. \qquad (126)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\frac{1}{N^{1-a}}\int_{1/2}^{1}\frac{(\log y+\log N)^{k-1}\log y}{(y+1/N)(1-y+1/N)^{1-a}}dy\\
&\quad\geq -\frac{(\log N-\log 2)^{k-1}\log 2}{a(1/2+1/N)N^{1-a}}\left[(1/2+1/N)^{a}-(1/N)^{a}\right]\\
&\quad\approx -\frac{2^{1-a}\log 2(\log N-\log 2)^{k-1}}{aN^{1-a}}.
\end{aligned}\tag{127}$$

It follows from (125), (126), and (127) that

$$\begin{aligned}
0&\leq I_k(N)\log N-I_{k+1}(N)\\
&\leq \frac{2(1-2^{-a})(\log N-\log M)\log^{k-1}M}{aN^{1-a}}+\frac{2^{1-a}\log 2(\log N-\log 2)^{k-1}}{aN^{1-a}}.
\end{aligned}\tag{128}$$

Combining (124) and (128), we have proved (120). It follows from (118) and (120) that

$$\left|J_{4k1}(N)-\frac{C_2B^{-(k+1)}(a,b)\log^{k}N}{N^{b}}\right|\leq\frac{C_2C_7B^{-(k+1)}(a,b)\log^{k-1}N}{N^{b}}.\tag{129}$$

Similarly, we can show that

$$\begin{aligned}
|J_{4k2}(N)|&\leq\frac{B(a,b)\Gamma(N+1)}{\Gamma(N+a+b)}\sum_{m=M}^{N}\frac{\Gamma(m+b)\Gamma(N-m+a)\log^{k-2+b}N}{m^{b}\Gamma(m+1)\Gamma(N-m+1)}\\
&\leq\frac{C_8\log^{k-1+b}N}{N^{b}},
\end{aligned}\tag{130}$$

where C_8 is a positive constant. Combining (117), (129), (130), we get (115). In the case of $b=1$, (116) can be proved in a similar manner. (115) and (116) imply (16) in Lemma 1. □

References

Abramowitz, M., Stegun, I., (1972). *Handbook of mathematical functions*, New York: Dover.

Barger, K., Bunge, J. (2010). Objective bayesian estimation of the number of species. *Bayesian Analysis*, *5*, 765–786.

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

Berger, J., Bernardo, J., Sun, D. (2012). Objective priors for discrete parameter spaces. *Journal of American Statistical Association*, *107*, 636–648.

Bolfarine, H., Leite, J., Rodrigues, J. (1992). On the estimation of the size of a finite and closed population. *Biometrical Journal*, *34*, 577–593.

Chao, M., Strawderman, W. (1972). Negative moments of positive random variables. *Journal of the American Statistical Association*, *67*, 429–431.

Erdelyi, A., Tricomi, F. (1951). The asymptotic expansion of a ratio of gamma functions. *Pacific Journal of Mathematics*, *1*, 133–142.

Farcomeni, A., Tardella, L. (2010). Reference bayesian methods for alternative recapture models with heterogeneity. *TEST*, *19*, 187–208.

George, E., Robert, C. (1992). Capture-recapture sampling via gibbs sampling. *Biometrika*, *79*, 677–683.

Jeffreys, H. (1961). *Theory of Probability*. London: Oxford University Press.

Lindsay, B., Roeder, K. (1987). A unified treatment of integer parameter models. *Journal of the American Statistical Association*, *82*, 758–764.

Otis, D., Burghal, K., White, G., Anderson, D. (1978). Statistical inference from capture-recapture data on closed animal populations. *Wildlife Monograph*, *62*, 1–135.

Raftery, A. (1988). Analysis of simple debugging model. *Applied Statistics*, *37*, 12–32.

Royle, A. (2008). Analysis of capture-recapture models with individual covariates using data augmentation. *Biometrics*, *65*, 267–274.

Seber, G., Huakau, J., Simmons, D. (2000). Capture-recapture, epidemiology and list mismatches: two lists. *Biometrics*, *56*, 1227–1232.

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

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

Wang, X., He, C., Sun, D. (2005). Bayesian inference on the patient population size given list mismatches. *Statistics in Medicine*, *24*, 249–267.

Wang, X., He, C., Sun, D. (2007). Bayesian population estimation for small sample capture-recapture data using noninformative priors. *Journal of Statistical Planning and Inference*, *137*, 1099–1118.