ESTIMATING POPULATION SIZE WITH OBJECTIVE BAYESIAN METHODS

A Dissertation

Presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

CHANG XU

Dr. Dongchu Sun, Dissertation Supervisor

JULY 2012

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

ESTIMATING POPULATION SIZE WITH
OBJECTIVE BAYESIAN METHODS

presented by Chang Xu,

a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Dongchu Sun

Dr. Chong He

Dr. Paul Speckman

Dr. Athanasios Micheas

Dr. Joshua Millspaugh

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my adviser, Dr. Dongchu Sun, for his guidance, encouragement and patience throughout my graduate study. This work would never have been done without his inspiration and enthusiasm.

I extend my gratitude to my co-advisor Dr. Chong He for her discussion and advices in simulations. I am also thankful to her for providing me the opportunity as a statistical intern in Missouri Department of Conservation.

I also express my appreciation to the other members of my advisory committee: Dr. Paul Speckman, Dr. Athanasios Micheas and Dr. Joshua Millspaugh. They are really nice and supportive all through my graduate study.

In addition, I would like to thank the Statistics Department faculty members, staff, and my fellow graduate students for making the department a strong one with a friendly atmosphere.

I am forever indebted to my families for their understanding and moral support throughout my time at school. Finally, I am grateful to my fiance Wenwen Fan. This work cannot be done without her patience and support.

TABLE OF CONTENTS

ACKNOWLEDGMENTS . . . . **ii**

LIST OF TABLES . . . . **v**

LIST OF FIGURES . . . . **vii**

ABSTRACT . . . . **xi**

CHAPTER . . . .

1 Introduction **1**

1.1 Overview 1

1.2 Outline of the Dissertation 2

2 Objective Bayesian Analysis for a Capture-recapture Model . . . **4**

2.1 Introduction 4

2.2 Reference prior of $\boldsymbol{\theta}$ 6

2.3 Reference prior of N 8

2.4 Simulation Study 13

2.4.1 Simulation Part 1 13

2.4.2 Simulation Part 2 15

2.5 Examples 17

2.5.1 Least chipmunk example 17

2.5.2 Turtle example 17

2.6 Concluding Remarks 18

3 Bayesian Analysis of a Capture-Recapture Model with Time Effect and Individual Covariates **29**

3.1 Introduction 29

3.2 Model Specification 31

3.2.1 Likelihood 31

3.2.2 Priors 32

3.2.3 The Posterior Propriety 33

3.2.4 Parameter Estimation 35

3.3 Softshell Turtle Example 36

3.3.1 Data Description and Preliminary Analysis 36

3.3.2 Results 37

4 Estimating Population Size with Exponential Failure under Type-I Censoring 41

4.1 Introduction 41

4.2 Hierarchical Model 43

4.2.1 Priors 43

4.2.2 Re-parameterization 51

4.3 Simulation Study 53

4.4 Concluding Remarks 55

APPENDIX 62

A Proof of Lemma 2.3.1 62

BIBLIOGRAPHY 96

VITA 100

LIST OF TABLES

Table Page

2.1 Least Chipmunk data by V. Reid (1975) 19

2.2 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). The first subscript represents the prior for N, and the second subscript represents the prior for $\boldsymbol{\theta}$ with U for Uniform prior and J for the Jeffreys prior. 19

2.3 Smooth Softshell Turtle data (2011) 20

2.4 Smooth Softshell Turtle 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). The first subscript represents the prior for N, and the second subscript represents the prior for $\boldsymbol{\theta}$ with U for Uniform prior and J for the Jeffreys prior. 20

3.1 Softshell Turtle capture history data at site No. 94 in Missouri River (2011). In total, 90 turtles were captured for at least once during 8 sampling occasions. In each sampling occasion, '1' means the turtle was captured and '0' means the turtle was not captured. Gender (Female/Male) and life stage (Adult/Juvenile) were recorded for each captured turtle. 39

3.2 Parameter estimates and DIC for models M_1, M_2, M_3, M_4. The numbers in parentheses are standard deviations for the estimates. 39

4.1 Estimates of N at $N = 50$. In each cell, the numbers outside and inside the parentheses are the estimate $\hat{N}$ and standard deviation of $\hat{N}$ respectively. 56

4.2 Estimates of N at $N = 100$. In each cell, the numbers outside and inside the parentheses are the estimate $\hat{N}$ and standard deviation of $\hat{N}$ respectively. 56

4.3 Estimates of N at $N = 200$. In each cell, the numbers outside and inside the parentheses are the estimate $\hat{N}$ and standard deviation of $\hat{N}$ respectively. 57

4.4 Estimates of N at $N = 500$. In each cell, the numbers outside and inside the parentheses are the estimate $\hat{N}$ and standard deviation of $\hat{N}$ respectively. 57

LIST OF FIGURES

Figure Page

2.1 Simulation results for k=3. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. 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. 21

2.2 Simulation results for k=4. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. 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. 22

2.3 Simulation results for k=5. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. 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. 23

2.4 Simulation results for k=6. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. 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. 24

2.5 Simulation results for k=7. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. 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. 25

2.6 Simulation results for k=8. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. 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. 26
2.7 Compare $\pi^A(N)$ with $\pi^R(N)$ for $k = 3, 4, 5$. In each graph, the x-axis denotes the true values of N from 50 to 800, and the y-axis denotes the bias. 27
2.8 Compare $\pi^A(N)$ with $\pi^R(N)$ for $k = 6, 7, 8$. In each graph, the x-axis denotes the true values of N from 50 to 800, and the y-axis denotes the bias. 28

3.1 Posterior probability mass function of N in the full model M_4. The solid line is the PMF under $\pi_A(N)$. The dashed line is the PMF under $\pi_U(N)$. The dotted line is the PMF under $\pi_J(N)$. 40

4.1 Coverage probabilities at $N = 50$ and $\gamma = 0.1, 0.25, 0.5, 0.75, 0.9$. The x axis represents the nominal coverages from 0.1 to 0.9 and 0.95, with 0.1 increment. The y axis represents the actual coverages. 58
4.2 Coverage probabilities at $N = 100$ and $\gamma = 0.1, 0.25, 0.5, 0.75, 0.9$. The x axis represents the nominal coverages from 0.1 to 0.9 and 0.95, with 0.1 increment. The y axis represents the actual coverages. 59
4.3 Coverage probabilities at $N = 200$ and $\gamma = 0.1, 0.25, 0.5, 0.75, 0.9$. The x axis represents the nominal coverages from 0.1 to 0.9 and 0.95, with 0.1 increment. The y axis represents the actual coverages. 60

4.4 Coverage probabilities at $N = 500$ and $\gamma = 0.1, 0.25, 0.5, 0.75, 0.9$. The x axis represents the nominal coverages from 0.1 to 0.9 and 0.95, with 0.1 increment. The y axis represents the actual coverages. 61

ABSTRACT

Bayesian inference of a discrete parameter, including population size, is sensitive to the choice of priors. In this dissertation I will develop objective priors for several population size parameters appeared in different models. Full Bayesian analysis for several datasets will be presented.

In the first part, a capture-recapture model (M_t model) is studied and the objective prior for the population size is developed. Simulation results demonstrate the advantages of the objective prior over the ad-hoc non-informative priors. Two real data examples are presented.

In the second part, another capture-recapture model (M_th model) is studied. The capture probabilities are modeled by both time effect and individual covariates via logit link function. A detailed analysis of Turtle data is presented to demonstrate our model and method.

In the third part, I consider the survival model with exponential failure rate and type-I censored failure times. The total number of failures is unknown due to censoring and needs to be estimated. The problem is attacked with a hierarchical Bayesian model. Simulation results are reported to compare several objective priors.

Chapter 1

Introduction

1.1 Overview

The need for estimating population size arises in many situations. For example in reliability experiments, one is interested in the unknown number of potential faults in a system. In ecology, the scientists study how the population sizes of species living together as a group change over time and space. In medical research, it is important to estimate the size of the population with certain disease.

Over the recent decades, there has been a large body of statistical literature devoted to this problem with both frequentist and Bayesian methods. However objective Bayesian methods have not been widely used to estimate population sizes, partially because developing objective prior for discrete parameter is a challenging problem and there is no unified approach for this. Bernardo & Smith (1994) showed the standard reference prior approach will always yield uniform prior if the parameter space is finite. Jeffreys (1961) recommended to use $1/N$ for the discrete parameter N. These two priors have been extensively studied and used in various models. Besides these two priors, Barger & Bunge (2008) suggested applying the Jeffreys-rule or reference approaches based on the discrete analogies of the Fisher information matrix (see Lindsay & Roeder 1987). More recently, Berger et al. (2011) proposed four different approaches to develop objective priors for discrete parameter spaces.

In this dissertation, we consider two models involving unknown population sizes. Several different approaches are used to develop the objective priors. Extensive simulations are conducted to compare the performances of various priors.

1.2 Outline of the Dissertation

The dissertation is organized as follows. In Chapter 2, we consider a simple capture-recapture model (M_t model) that has been widely used to monitor specie richness in ecological research. The parameter of interest is the population size. The nuisance parameters include a vector of independent capture probabilities. The conditional Jeffreys-rule prior for the capture probabilities and the marginal reference prior for the population size are developed sequentially. Simulation studies are conducted and the results are summarized. Two real data examples are also provided.

Chapter 3 extends the simple M_t model to the more applicable and complicated M_{th} model involving individual covariates. In this model, the capture probabilities are assumed to vary not only across the sampling occasions, but also the individual captured animals. Individual information, such as gender and life stage, is used to help estimate the capture probabilities and population size with better accuracy. The objective priors developed in Chapter 2 are applied in this model, and conditions for posterior propriety are given. This model has been applied to an ongoing research in Missouri River turtles.

Chapter 4 considers a survival model with exponential failure rates, and the failure times are observed prior to a pre-determined censoring time. This model has applications in software engineering and wildlife science. Due to censoring, only a proportion of potential failures will be observed, and the number of remaining failures remains unknowns and needs to be estimated. The nuisance parameter is the hazard rate of the exponential distribution. We treat the population size as a latent variable and impose a hierarchical structure on it. Standard non-informative priors are considered including a Jeffreys-rule prior and a one-at-a-time reference prior. The simulation results are presented.

In Chapter 4, several directions for future research are discussed.

Chapter 2

Objective Bayesian Analysis for a Capture-recapture Model

2.1 Introduction

Capture-recapture is a method to estimate population size. It was originally developed in ecological study, and has been utilized extensively in epidemiology and demographic research. A sequence of classic capture-recapture models were discussed in Otis et al. (1978), commonly referred as M_0, M_t, M_b, M_h, M_{tb}, M_{th}, M_{bh}, and M_{tbh}. For all these models, a closed population is assumed. Meanwhile, each model allows for different assumptions on capture probabilities. Bayesian approaches have been proved very useful in analyzing both classic capture-recapture models and more complexed, hierarchical models that incorporate spatial patterns and covariates (see Royle 2008, Wang et al. 2007, Basu & Ebrahimi 2001, George & Robert 1992, Bolfarine et al. 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 to selecting a prior for N is to impose a hierarchical structure and place priors on the hyper-parameters. For example, Raftery (1988a) and Smith (1991) recommended using a Poisson(λ) prior for N, and choosing a 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 & Robert (1992) studied the M_t model and computed Bayesian estimates via Gibbs sampling. They suggested to use $1/N$ or a Poisson prior on N. Bolfarine et al. (1992) used proper prior $1/(N+1)^2$. 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 the M_t model with fixed N and $\boldsymbol{\theta}$, the Bayesian estimate of N would decrease as the number of sampling occasions increases. In consequence, Bayesian estimates underestimate N for large k.

We will focus on the M_t model and develop the reference priors for both $\boldsymbol{\theta}$ and N. The M_t model is one of the simplest capture-recapture models described in Otis et al. (1978). In addition to the common "closed population" assumption, the M_t model requires (1) homogeneity among all units in the population (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 application in wildlife management, ecology, epidemiology and other areas. Secondly, this is the first paper that attempts to formally develop an likelihood-based objective prior for population size in a capture-recapture model. We start from a simple model and will continue working on more complexed ones. Thirdly, the formal approach of deriving reference prior is rather complicated even for a model as simple as M_t.

The paper is organized as follows. In section 2.2 the conditional Jeffreys prior for $\boldsymbol{\theta}$ given N will be derived from the conditional likelihood. In section 2.3 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 section 2.4 we will present and explain the simulation results and show that the reference priors have advantages over the commonly used non-informative priors. An analysis of a real data example is presented in section 2.5. Finally we will conclude the paper and discuss future work in section 2.6.

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

We assume 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 digit binary numbers, in 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

$$p_i(\boldsymbol{\theta}) = \prod_{j=1}^{k} \theta_j^{b_{ij}} (1 - \theta_j)^{1-b_{ij}}, \tag{2.1}$$

where b_{ij} is the j^{th} binary digit of the integer i. The likelihood function of N and $\boldsymbol{\theta}$ given $n_1, \ldots, n_L$ is

$$L_1(N, \boldsymbol{\theta} \mid n_1, \ldots, n_L) = \frac{N!}{n_0! n_1! \cdots n_L!} \prod_{i=0}^{L} p_i(\boldsymbol{\theta})^{n_i}. \tag{2.2}$$

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

Theorem 2.2.1. *The Jeffreys prior for $\boldsymbol{\theta}$ given N is the product of Beta$(\frac{1}{2}, \frac{1}{2})$, and given by*

$$\pi(\boldsymbol{\theta} \mid N) \propto \frac{1}{\prod_{i=1}^{k} \sqrt{\theta_i(1-\theta_i)}}. \tag{2.3}$$

Proof. The likelihood function (2.2) 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

$$\boldsymbol{e}_j \mid \boldsymbol{\theta} \sim \text{Multi}(1; p_1(\boldsymbol{\theta}), \ldots, p_L(\boldsymbol{\theta})). \tag{2.4}$$

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

$$\log p(\boldsymbol{e}_j \mid \boldsymbol{\theta}) = \sum_{i=0}^{L} e_{ji} \log p_i(\boldsymbol{\theta}). \tag{2.5}$$

For convenience, define

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

It is easy to see that, for any $r = 1, \ldots, k$,

$$\begin{aligned} \frac{\partial^2}{\partial \theta_r^2} \sum_{i=0}^{L} e_{ji} \log p_i(\boldsymbol{\theta}) &= \frac{\partial}{\partial \theta_r} \sum_{i \in \boldsymbol{C}^r} e_{ji} \frac{1}{\theta_r} - \frac{\partial}{\partial \theta_r} \sum_{i \in \boldsymbol{C}_r^c} \frac{1}{1-\theta_r} \\ &= -\sum_{i \in \boldsymbol{C}_r} e_{ji} \frac{1}{\theta_r^2} - \sum_{i \in \boldsymbol{C}_r^c} e_{ji} \frac{1}{(1-\theta_r)^2}. \end{aligned} \tag{2.6}$$

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

$$\mathrm{E}\left(\frac{\partial^2}{\partial\theta_r^2}\sum_{i=0}^{L} e_{ji}\log p_i(\boldsymbol{\theta})\right) = -\frac{1}{\theta_r}\sum_{i\in\boldsymbol{C}_r}\frac{p_i(\boldsymbol{\theta})}{\theta_r} - \frac{1}{(1-\theta_r)}\sum_{i\in\boldsymbol{C}_r^c}\frac{p_i(\boldsymbol{\theta})}{(1-\theta_r)} = -\frac{1}{\theta_r(1-\theta_r)}. \tag{2.7}$$

Also we have

$$\frac{\partial^2}{\partial\theta_s\theta_t}\sum_{i=0}^{L} e_{ji}\log p_i(\boldsymbol{\theta}) = 0. \tag{2.8}$$

Therefore the Fisher information matrix for $\boldsymbol{\theta}$ is

$$\mathrm{Diag}\left\{\theta_1^{-1}(1-\theta_1^{-1}), \ldots, \theta_k^{-1}(1-\theta_k^{-1})\right\},$$

yielding the result. □

2.3 Reference prior of N

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

$$\theta_j \stackrel{\text{iid}}{\sim} \mathrm{Beta}(a, b), \tag{2.9}$$

where a, b are two positive constants. Let $\pi_{a,b}(\boldsymbol{\theta})$ be the joint density of the vector $\boldsymbol{\theta}$. For convenience, we rewrite the likelihood function (2.2) in the following form:

$$L_2(N, \boldsymbol{\theta} \mid n_1, \ldots, n_L) = \frac{N!}{n_0!n_1!\cdots n_L!}\prod_{i=0}^{k}\theta_i^{m_i}(1-\theta_i)^{N-m_i}. \tag{2.10}$$

The equivalence of (2.2) and (2.10) can be verified by simple algebra. First we obtain the marginal likelihood of N by averaging $\boldsymbol{\theta}$ over its prior.

$$
\begin{aligned}
L_3(N \mid n_1, \ldots, n_L) &= \int L_2(N, \boldsymbol{\theta} \mid n_1, \ldots, n_L) \pi_{a,b}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int_0^1 \cdots \int_0^1 \frac{N!}{n_0! n_1! \cdots n_L!} \prod_{i=0}^{k} \theta_i^{m_i} (1-\theta_i)^{N-m_i} \theta_i^{a-1} (1-\theta_i)^{b-1} d\theta_1 \cdots d\theta_k \\
&= \frac{N!}{n_0! n_1! \cdots n_L!} \int_0^1 \cdots \int_0^1 \prod_{i=1}^{k} \theta_i^{m_i+a-1} (1-\theta_i)^{N-m_i+b-1} d\theta_1 \cdots d\theta_k \\
&= \frac{N!}{n_0! n_1! \cdots n_L!} \prod_{i=1}^{k} \frac{\Gamma(m_i + a)\Gamma(N - m_i + b)}{\Gamma(N + a + b)}. \qquad (2.11)
\end{aligned}
$$

Taking the second derivative of $\log L_3(N \mid n_1, \ldots, n_L)$ with respect to N, we have

$$
\begin{aligned}
& \frac{\partial^2}{\partial N^2} \log L_3(N \mid n_1, \ldots, n_L) \\
= \; & \frac{\partial^2}{\partial N^2} \Big\{ \log \Gamma(N+1) + \sum_{l=1}^{k} \log \Gamma(N - m_l + b) - k \log \Gamma(N + a + b) \\
& - \log \Gamma(N - n + 1) \Big\}. \qquad (2.12)
\end{aligned}
$$

Using the following property of trigamma function (Abramowitz & Stegun 1972):

$$
\frac{\partial^2}{\partial N^2} \log \Gamma(N) = \sum_{i=0}^{\infty} \frac{1}{(N+i)^2}, \qquad (2.13)
$$

we have

$$
\text{The RHS of (2.12)} \; = -J_1(N) + J_2(N) - J_3(N), \qquad (2.14)
$$

where

$$J_1(N) = \sum_{i=0}^{\infty} \frac{k}{(N+a+b+i)^2} - \sum_{i=0}^{\infty} \frac{1}{(N+1+i)^2}, \tag{2.15}$$

$$J_2(N) = \sum_{l=1}^{k} \sum_{i=0}^{\infty} \frac{1}{(N-m_l+b+i)^2}, \tag{2.16}$$

$$J_3(N) = \sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2}. \tag{2.17}$$

The Fisher information of N is then

$$I(N) = J_1(N) - \mathrm{E}(J_2(N)) + \mathrm{E}(J_3(N)), \tag{2.18}$$

where the expectation is taken under the marginal distribution of $(n_1, \ldots, n_L)$, given by (2.11). 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 $I(N)$.

Theorem 2.3.1. *Let $\pi^R(N)$ be the reference prior of N. For any fixed $k \geq 2$ and $0 < a \leq 1$,*

(1) *if $0 < b < 1$,*

$$\pi^R(N) \propto \frac{(\log N)^{\frac{k-1}{2}}}{(N+1)^{\frac{b}{2}}}; \tag{2.19}$$

(2) *if $b = 1$,*

$$\pi^R(N) \propto \frac{(\log N)^{\frac{k}{2}}}{(N+1)^{\frac{1}{2}}}. \tag{2.20}$$

Proof. We need the following lemma to prove this result.

Lemma 2.3.1. *For any fixed $k \geq 2$ and $0 < a, b \leq 1$, as $N \to \infty$, we have*

$$E(J_1(N)) \sim \frac{k}{N+a+b} - \frac{1}{N+1}. \tag{2.21}$$

For $0 < b < 1$*,*

$$\begin{aligned} E(J_2(N)) &\sim \frac{1}{N^b}, &(2.22)\\ E(J_3(N)) &\sim \frac{(\log N)^{k-1}}{(N+1)^b}. &(2.23)\end{aligned}$$

For $b = 1$*,*

$$\begin{aligned} E(J_2(N)) &\leq \frac{C_0 \log N}{N+1}, &(2.24)\\ E(J_3(N)) &\sim \frac{(\log N)^{k}}{N+1}, &(2.25)\end{aligned}$$

where C_0 *is a positive constant.*

The proof of Lemma 2.3.1 is fairly sophisticated and can be found in Appendix A. Lemma 2.3.1 yields Theorem 2.3.1 immediately. $\square$

One can easily show that under $\pi^R(N)$, the posterior distribution of N is proper and has moments unless the data are too extreme.

Corollary 2.3.1. *For any fixed* $k \geq 2$*,* $0 < a, b \leq 1$ *and non-negative integer* g*, under priors (2.19) and (2.20), the* g^{th} *posterior moment of* N *exists if there exists* $\epsilon > 0$ *such that*

$$\sum_{l=1}^{k} m_l - n + ka + b/2 \geq g + 1 + \epsilon. \tag{2.26}$$

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

$\pi_{a,b}(\boldsymbol{\theta})$.

$$
\begin{aligned}
L(N \mid \text{data}) = & \int L(N, \boldsymbol{\theta} \mid data)\pi_{a,b}(\boldsymbol{\theta})d\boldsymbol{\theta} \\
\propto & \int_0^1 \cdots \int_0^1 \frac{N!}{(N-n)!} \prod_{l=1}^{k} \theta_l^{m_l+a-1}(1-\theta_l)^{N-m_l+b-1} d\theta_1 \cdots d\theta_k \\
= & \frac{\Gamma(N+1)}{\Gamma(N-n+1)} \prod_{l=1}^{k} \frac{\Gamma(m_l+a)\Gamma(N-m_l+b)}{\Gamma(N+a+b)} \\
\propto & \frac{\Gamma(N+1)}{\Gamma(N-n+1)} \prod_{l=1}^{k} \frac{\Gamma(N-m_l+b)}{\Gamma(N+a+b)} \\
\sim & N^{-(ka-n+\sum_{l=1}^{k} m_l)}.
\end{aligned} \tag{2.27}
$$

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

$$
\begin{aligned}
\pi(N \mid \text{data}) & \propto L(N \mid \text{data})\pi^R(N) \\
& \sim (\log N)^{\frac{k-1}{2}} (1+N)^{-\frac{b}{2}} N^{-\left[ka-n+\sum_{l=1}^{k} m_l\right]} \\
& \sim (\log N)^{\frac{k-1}{2}} N^{-\left[ka+b/2+\sum_{l=1}^{k} m_l - n\right]}.
\end{aligned} \tag{2.28}
$$

For $b = 1$, similarly we have

$$
\pi(N \mid \text{data}) \sim (\log N)^{\frac{k}{2}} N^{-\left[ka+1/2+\sum_{l=1}^{k} m_l - n\right]}. \tag{2.29}
$$

Clearly, if condition (2.26) holds, for $0 < b < 1$ we have

$$
\mathrm{E}(N^g) = \sum_{N=n}^{\infty} N^g \pi(N \mid \text{data}) \leq \sum_{N=n}^{\infty} \frac{(\log N)^{\frac{k-1}{2}}}{N^{1+\epsilon}} < \infty; \tag{2.30}
$$

for $b = 1$,

$$
\mathrm{E}(N^g) = \sum_{N=n}^{\infty} N^g \pi(N \mid \text{data}) \leq \sum_{N=n}^{\infty} \frac{(\log N)^{\frac{k}{2}}}{N^{1+\epsilon}} < \infty. \tag{2.31}
$$

$\square$

This condition is very easy to satisfy in practice. For example, when $g = 0$, inequality (2.26) holds as long as there is at least one recapture.

2.4 Simulation Study

The simulation study consists of two parts. In the first part we compare the performances of the asymptotic reference prior under different (a, b)'s. In the second part, we simulate Bayes estimates based on an alternative approximation to the exact reference prior, which does not require large N. The details and results of the simulation study are summarized in the following subsections.

2.4.1 Simulation Part 1

For comparison, we assume that a and b equal and take values $0.25, 0.50, 0.75, 1.00$. Notice that when $a = b = 0.5$, it corresponds to the Jeffreys prior for $\boldsymbol{\theta}$. The number of sampling occasions k varies from 3 to 8. The rest of the simulation setups are adopted from Wang et al. (2007). The population size N is set to vary from 50 to 800 with 50 increment. 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, large capture probabilities.

The posterior means are estimated numerically when they exist. For each $(N, \boldsymbol{\theta}, k)$ combination, 2000 capture history data sets are generated according to likelihood (2.2). For the j^{th} dataset, the posterior mean, given it exists, is estimated by

$$\hat{N}_j = \sum_{N=n}^{V} \frac{NL(N \mid \text{data}_j)\pi^R(N)}{L(N \mid \text{data}_j)\pi^R(N)}, \tag{2.32}$$

where $V = 500,000$ (in some cases V is set to be $1,000,000$ to ensure convergence) and $L(N \mid \text{data})$ is given in (2.2). Therefore the estimated posterior mean is

$$\hat{N} = \frac{1}{2000} \sum_{j=1}^{2000} \hat{N}_j. \tag{2.33}$$

The estimated bias and mean square error can be calculated based on $\hat{N}_j$'s:

$$\begin{aligned} \widehat{\text{Bias}} &= \hat{N} - N, & (2.34) \\ \widehat{\text{MSE}} &= \frac{1}{2000} \sum_{j=1}^{2000} (\hat{N}_j - N)^2. & (2.35) \end{aligned}$$

We also compare 95% frequentist coverage probabilities and credible intervals of N. In general, the $100(1-\alpha)\%$ frequentist coverage probability of N is

$$P_\alpha = P(N \leq N_\alpha \mid \text{data}), \tag{2.36}$$

where N_α is the frequentist $100(1-\alpha)^{th}$ percentile. The two sided 95% credible intervals are obtained by taking the 2.5^{th} and 97.5^{th} sample percentiles from the 2000 estimates after sorting.

The results are presented in figures 2.6-2.6. Each figure corresponds to a sampling occasion k ranging from 3 to 8 and contains three panels. 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 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 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 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 $k \geq 5$, $a = b = 0.50$ (Jeffreys prior on $\boldsymbol{\theta}$) is the best choice, since the bias stays around zero. With large capture probabilities, $a = b = 0.5$ leads to the smallest bias when $k \geq 7$. In terms of coverage probability, while both $b = 0.25$ and $b = 0.50$ provide satisfactory results consistently, $b = 0.25$ leads to nearly exact coverage probability. The MSEs under four priors are very close to each other.

2.4.2 Simulation Part 2

Since $\pi^R(N)$ is the limiting rate of $\sqrt{I(N)}$ as $N \to \infty$, an alternative reference prior is needed. To avoid confusion, let Y^* be the simulated (fake) data and Y be the real data. Both Y^* and Y have the same density $f(\dot{|}N)$. We can define the Fisher information of N alternatively as

$$I^*(N) = \mathrm{E}\left(\frac{\partial}{\partial N}\log f(Y^* \mid N)\right)^2. \tag{2.37}$$

Therefore the alternative reference prior of N can be defined and approximated as

$$\pi^A(N) = \sqrt{\mathrm{E}\left(\frac{\partial}{\partial N}\log f(Y^* \mid N)\right)^2} \approx -\mathrm{E}\left(\frac{\partial}{\partial N}\log f(Y^* \mid N)\right). \tag{2.38}$$

The closed form of $\pi^A(N)$ is not available, so we resort to numerical methods. In order to compare $\pi^A(N)$ with $\pi^R(N)$ we use the same setup as in simulation part 1. For each $(N, \boldsymbol{\theta}, k)$ triplet, 500 samples are independently generated. For each sample Y_i, $i = 1, \cdots, 500$, the posterior mean of N given data Y_i based on $\pi^A(N)$ can be calculated as

$$
\begin{aligned}
\mathrm{E}(N \mid Y_i) &= \frac{\sum N f(Y_i \mid N)\pi_1^R(N)}{\int f(Y_i \mid N)\pi_1^R(N)} \\
&= \frac{-\sum N\left(\frac{\partial}{\partial N}\log f(Y^* \mid N)dY^*\right) f(Y_i \mid N)}{-\sum\left(\frac{\partial}{\partial N}\log f(Y^* \mid N) f(Y^* \mid N)dY^*\right) f(Y_i \mid N)} \\
&= \frac{\sum N \frac{\partial}{\partial N}\log f(Y^* \mid N) f(Y_i \mid N) f(Y^* \mid N)dY^*}{\sum \frac{\partial}{\partial N}\log f(Y^* \mid N) f(Y_i \mid N) f(Y^* \mid N)dY^*}. \qquad (2.39)
\end{aligned}
$$

Since the numerator can be viewed as the expected value of $N\frac{\partial}{\partial N}\log f(Y^* \mid N)$ with respect to the joint distribution of (Y^*, N). we use Monte Carlo integration to approximate it. The samples are drawn via Gibbs sampling since the full conditional distributions of N and $\boldsymbol{\theta}$ are known.

$$
\begin{aligned}
N \mid \cdot &\sim \mathrm{NegBin}\left(n, 1 - \prod_{j=1}^{k}(1-\theta_j)\right), \\
\theta_j \mid \cdot &\sim \mathrm{Beta}(m_j + a, N - m_j + b). \qquad (2.40)
\end{aligned}
$$

The full algorithm is presented below:

In cycle l $(l = 1, \ldots, M)$,

Step 1. Draw $\theta_j^{(l)}$ from $\mathrm{Beta}(m_j + a, N^{(l-1)} - m_j + b)$, for $j = 1, \ldots, k$,

Step 2. Draw $N^{(l)}$ from $\mathrm{NegBin}\left(n, 1 - \prod_{j=1}^{k}(1-\theta_j^{(l)})\right)$,

Step 3. Simulate "fake" data Y_l^* based on updated parameters $(N^{(l)}, \boldsymbol{\theta}^{(l)})$.

Step 4. Calculate $h(Y_l^*) = \frac{\partial}{\partial N^{(l)}}\log f(Y_l^* \mid N^{(l)})$.

In the end of M cycles, the numerator of (2.39) can be calculated by $\mathrm{E}(N \mid Y_i) \approx \frac{1}{M}\sum_{l=1}^{M} h(Y_l^*)$. The denominator of (2.39) can be approximated similarly. In this way we can obtain the frequentist properties of Bayes estimates under $\pi^A(N)$ by averaging over $\mathrm{E}(N \mid Y_i)$'s. Figures 2.6 and 2.6 compare the frequentist biases of Bayes estimates under $\pi^R(N)$ and $\pi^A(N)$ when $a = b = 0.5$. In each figure, x-axis represents N from

50 to 800 and y-axis represents the bias. The solid curve corresponds to $\pi^A(N)$ and the dashed curve corresponds to $\pi^R(N)$.

Interestingly, $\pi^A(N)$ dominates $\pi^R(N)$ when $k \leq 5$ and leads to nearly unbiased estimates, especially for small or moderate capture probabilities. When $k \geq 6$, like the commonly used non-informative priors, $\pi^A(N)$ produces Bayes estimates that underestimate the true parameter while $\pi^R(N)$ adjust for the effect of large k and provides nearly unbiased estimates.

2.5 Examples

2.5.1 Least chipmunk example

V. Reid (as reported in Otis et al. (1978) and re-analyzed by Wang et al. (2007)) did a capture-recapture experiment on Least Chipmunk *Eutamias minimus* in 1975. In this study, a 9×11 livetrapping grid was set up spaced 50 feet apart, and Least chipmunk were trapped once a day, for 6 consecutive days ($k = 6$). Otis et al. (1978) suggested to use the M_t model based on a discrimination procedure. The capture history data, after being processed, are given in table 2.6.

Since k is relatively large, we choose the prior $\pi^R(N)$ and Jeffrey prior for $\boldsymbol{\theta}$ ($a = b = 0.5$). We present our results, as well as Wang et al. (2007)'s, in Table 2.6. Clearly, although all five estimates are very close to each other, π^R produces the shortest 95% credible interval.

2.5.2 Turtle example

In the summer of 2011, a research team in the Department of Fisheries and Wildlife Sciences at University of Missouri conducted a mark-recapture study of smooth soft-shell 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 sufficient statistics are summarized in Table 2.6.

The estimates of N, 95% credible intervals, and standard deviations are summarized in Table 2.6. The marginal reference prior for $a = b = 0.5$ and maximum likelihood method results in close estimates for N. The maximum likelihood estimates for capture probabilities are $(0.048, 0.056, 0.072, 0.032, 0.092, 0.104, 0.088, 0.096)$, corresponding to the small capture probabilities case in the simulation.

2.6 Concluding Remarks

In conclusion, objective Bayesian approach is very useful in estimating population size even when no historical information is available. Based on theoretical and simulation results, the conditional Jeffreys prior ($a = b = 0.5$) for $\boldsymbol{\theta}$ and asymptotic reference prior for N provides the most accurate and robust Bayes estimates with $k \geq 5$. For $k = 3, 4$, we recommend to use the alternative reference prior $\pi^A(N)$ to significantly reduce the bias.

It is unknown to us whether $\pi^A(N)$ has the same order as $\pi^R(N)$ as $N \to \infty$. In the future we will continue working towards a unified reference prior for N. It will also be interesting and challenging to explore the frequentist properties of $\pi^R(N)$ analytically, so that we may explain how it adjusts for large k. Last but not least, it is interesting to explore why smaller b (such as $b = 0.25$) leads to nearly exact matching prior.

Table 2.1: 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					

Table 2.2: 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). The first subscript represents the prior for N, and the second subscript represents the prior for $\boldsymbol{\theta}$ with U for Uniform prior and J for the Jeffreys prior.

	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^R	51	3.38	$(46, 56)$

Table 2.3: 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 2.4: Smooth Softshell Turtle 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). The first subscript represents the prior for N, and the second subscript represents the prior for $\boldsymbol{\theta}$ with U for Uniform prior and J for the Jeffreys prior.

	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	$(182, 320)$
$\pi^R, a = b = 0.25$	256	37	$(197, 341)$
$\pi^R, a = b = 0.50$	247	34	$(191, 325)$
$\pi^R, a = b = 0.75$	238	32	$(187, 311)$
$\pi^R, a = b = 1.00$	231	29	$(183, 299)$

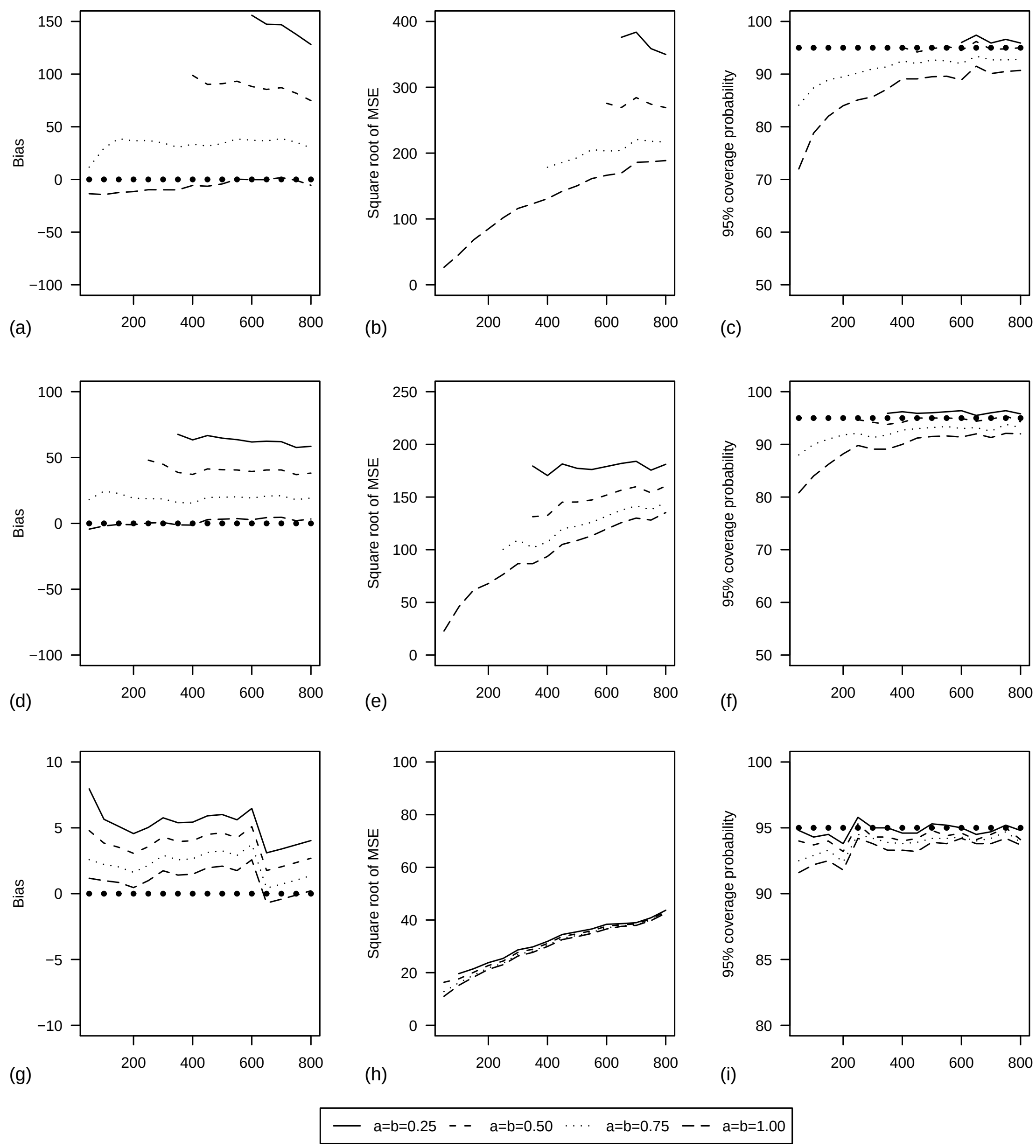

Figure 2.1: Simulation results for k=3. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. The top, middle, and bottom panel corresponds to the small, moderate, and large capture probabilities. In each panel, the left graph compares the bias, the middle graph compares 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.

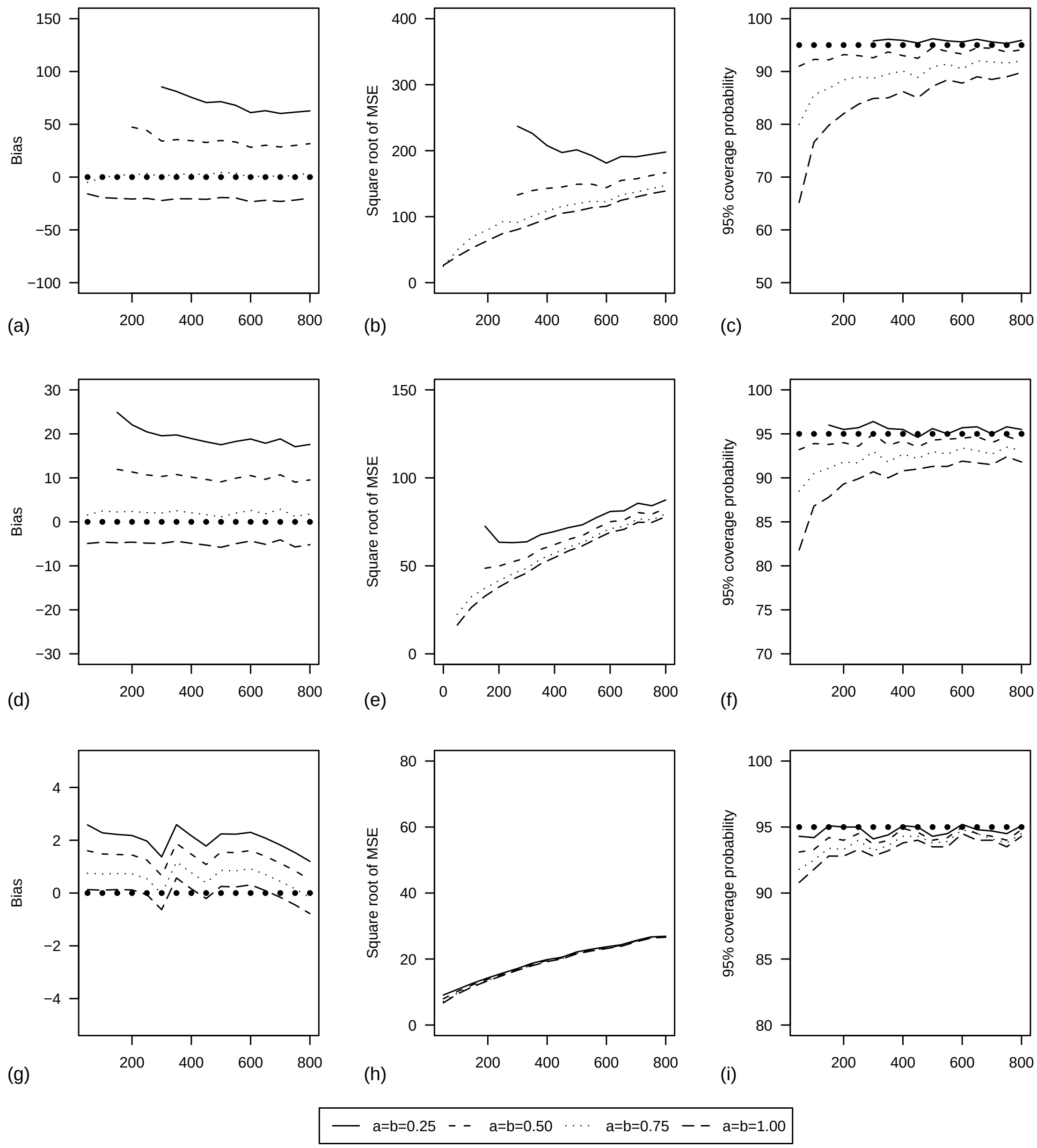

Figure 2.2: Simulation results for k=4. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. The top, middle, and bottom panel corresponds to the small, moderate, and large capture probabilities. In each panel, the left graph compares the bias, the middle graph compares 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.

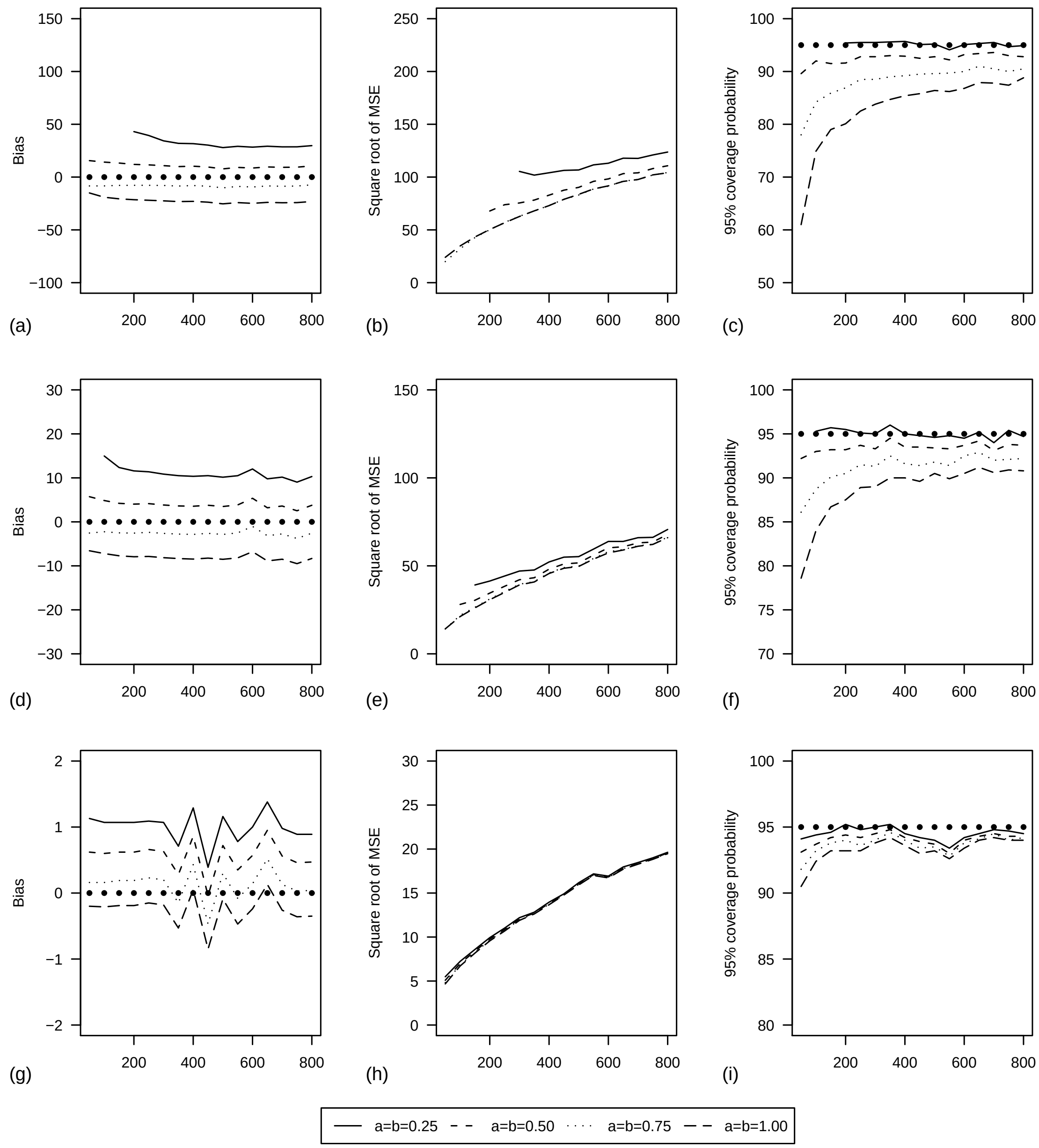

Figure 2.3: Simulation results for k=5. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. The top, middle, and bottom panel corresponds to the small, moderate, and large capture probabilities. In each panel, the left graph compares the bias, the middle graph compares 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.

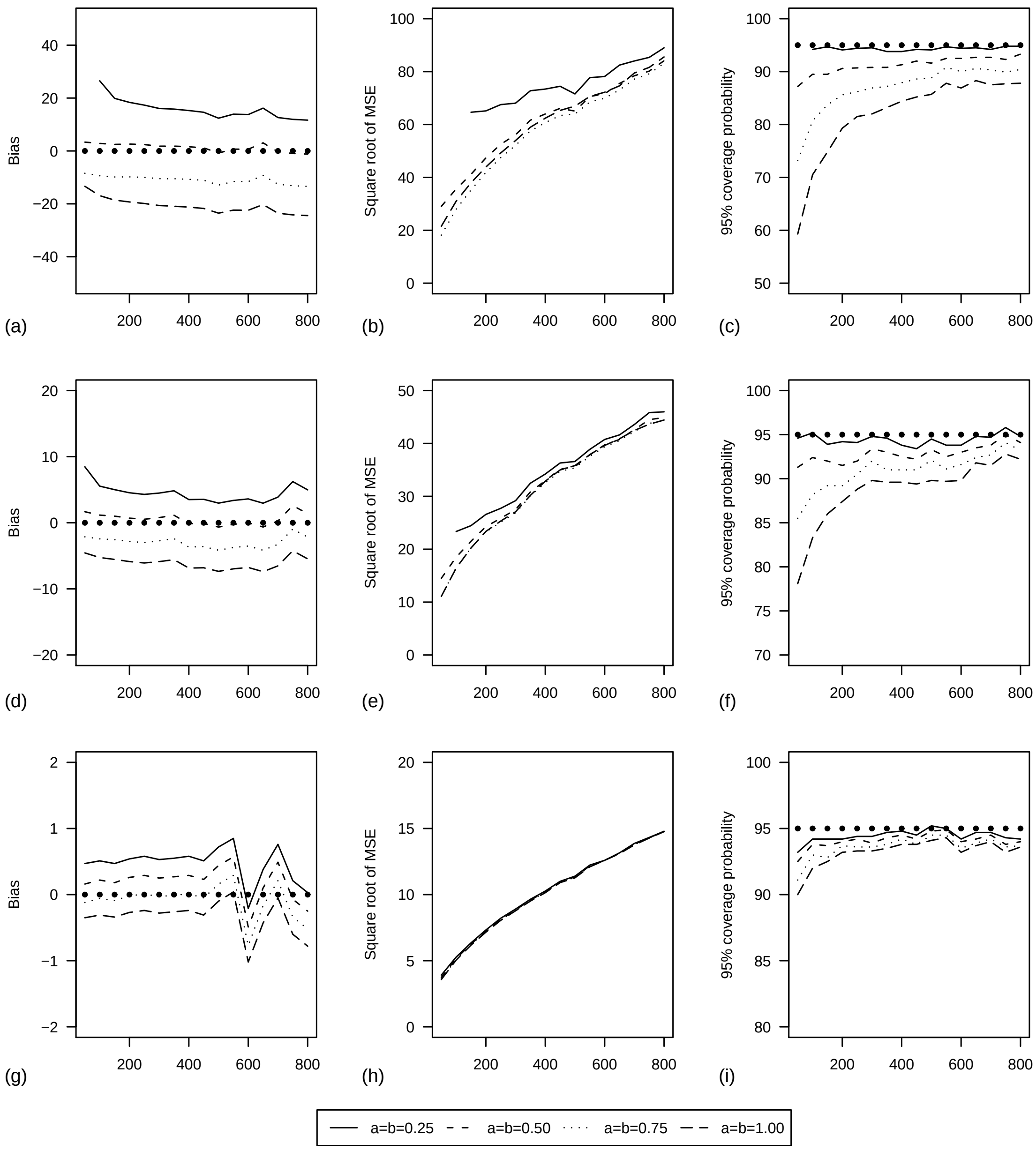

Figure 2.4: Simulation results for k=6. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. The top, middle, and bottom panel corresponds to the small, moderate, and large capture probabilities. In each panel, the left graph compares the bias, the middle graph compares 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.

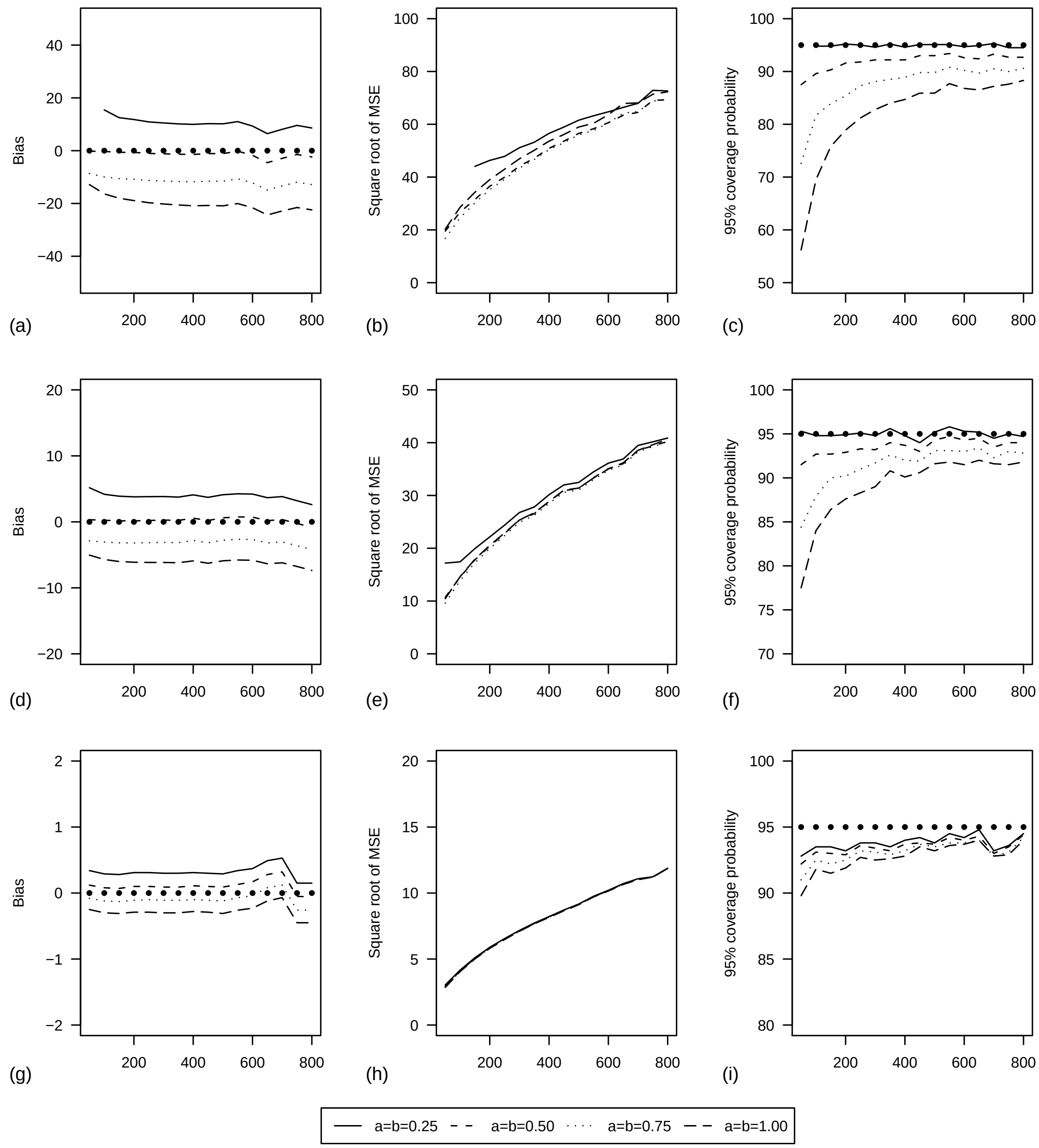

Figure 2.5: Simulation results for k=7. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. The top, middle, and bottom panel corresponds to the small, moderate, and large capture probabilities. In each panel, the left graph compares the bias, the middle graph compares 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.

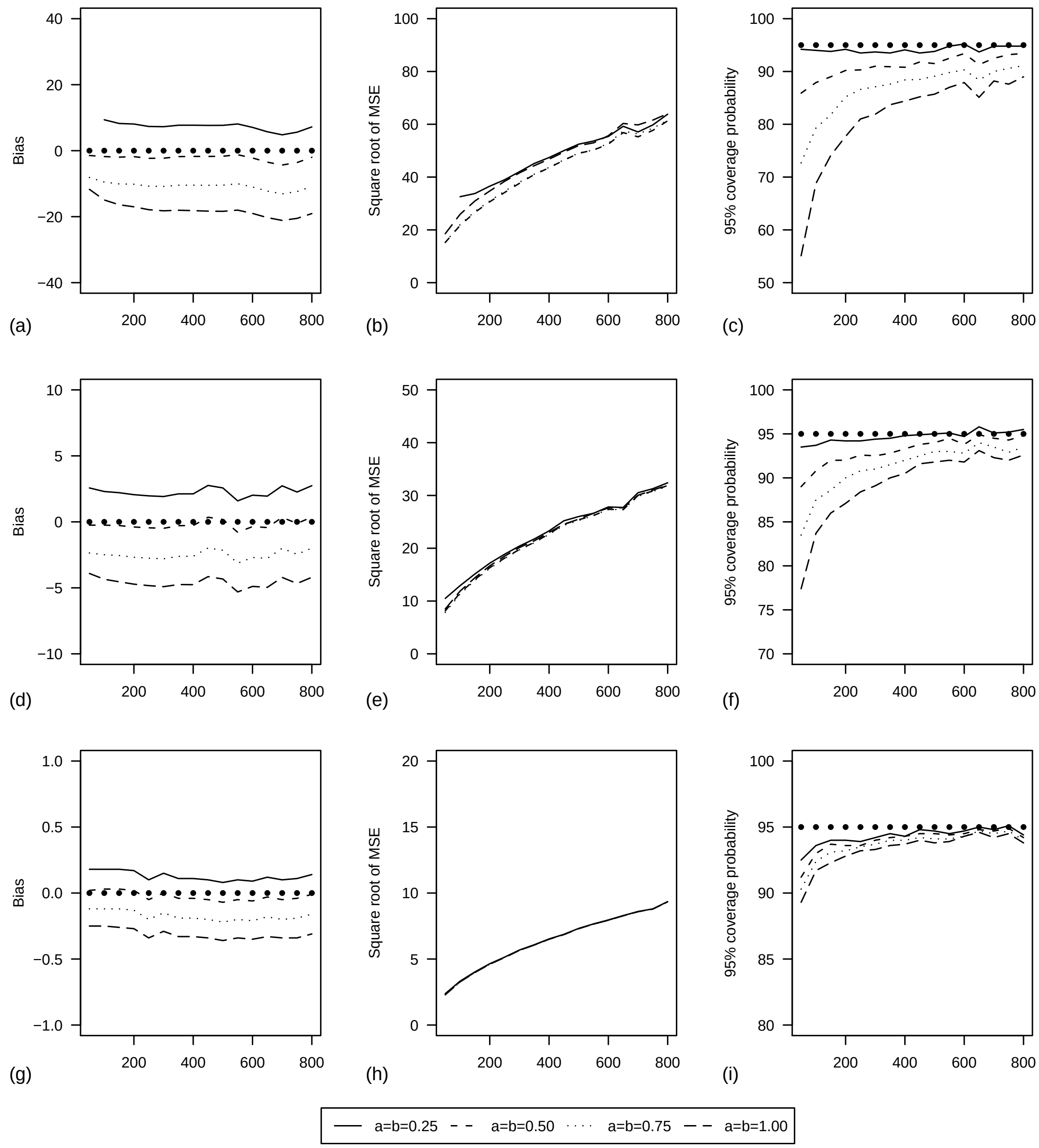

Figure 2.6: Simulation results for k=8. Four priors are compared, corresponding to $a = b = 0.25, 0.50, 0.75, \text{and} 1.00$. The top, middle, and bottom panel corresponds to the small, moderate, and large capture probabilities. In each panel, the left graph compares the bias, the middle graph compares 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.

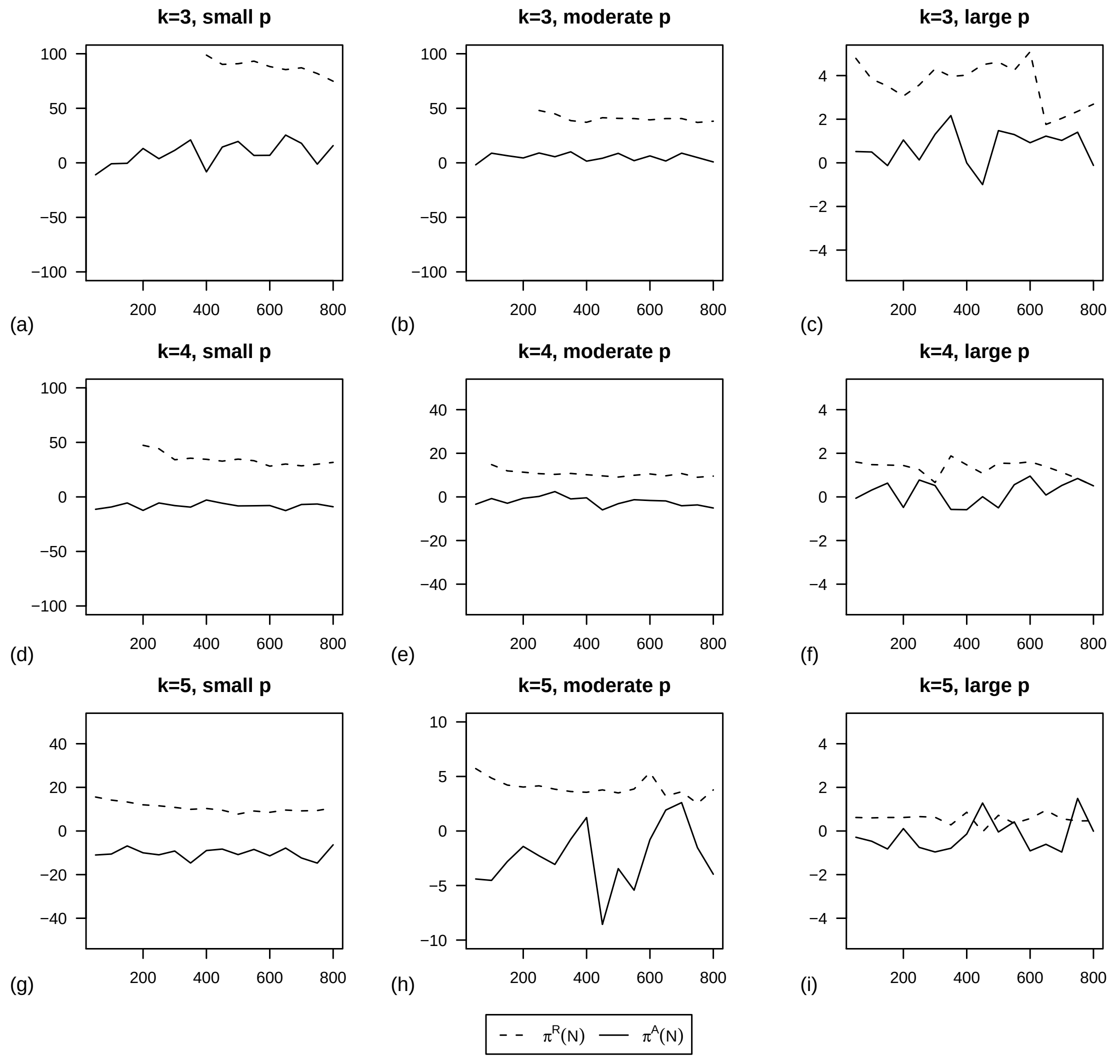

Figure 2.7: Compare $\pi^A(N)$ with $\pi^R(N)$ for $k = 3, 4, 5$. In each graph, the x-axis denotes the true values of N from 50 to 800, and the y-axis denotes the bias.

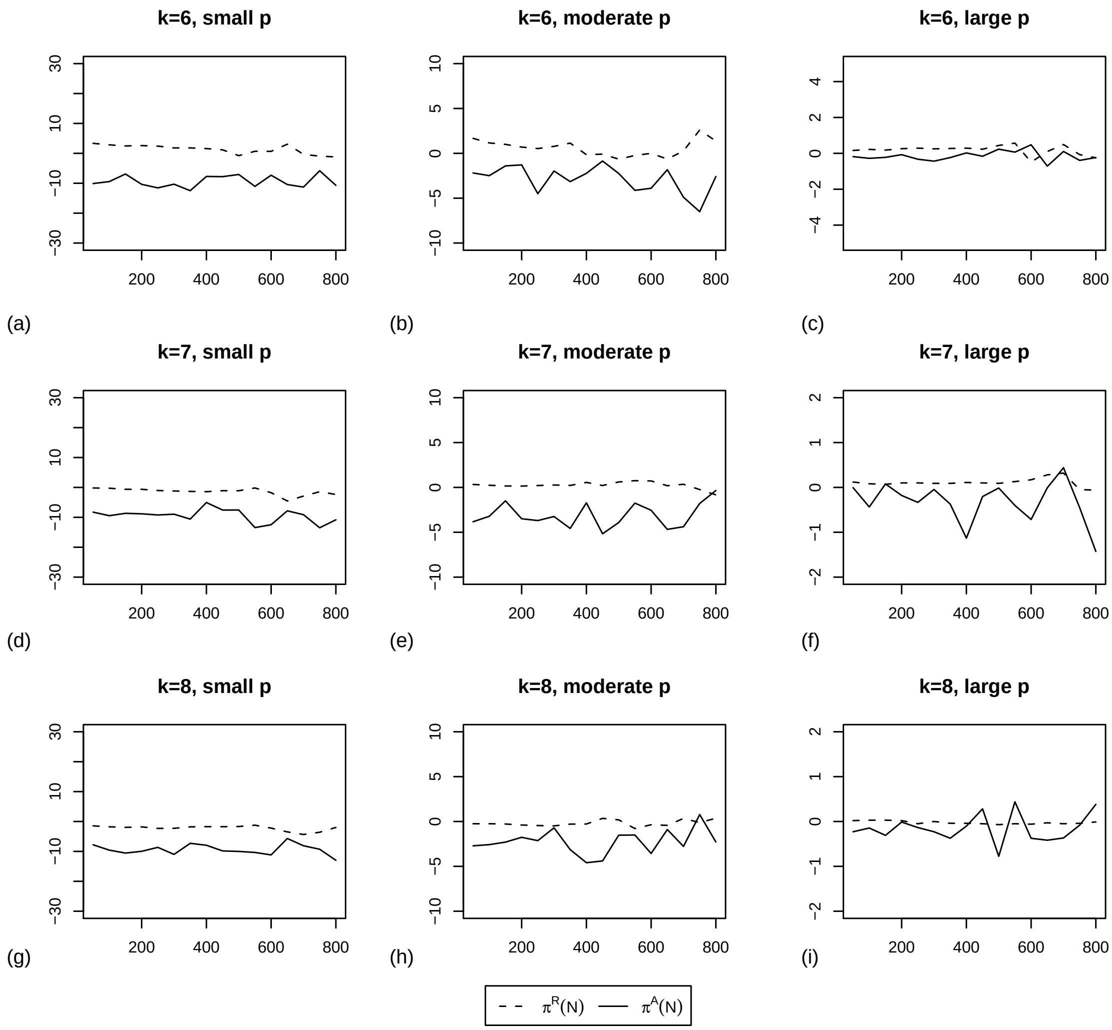

Figure 2.8: Compare $\pi^A(N)$ with $\pi^R(N)$ for $k = 6, 7, 8$. In each graph, the x-axis denotes the true values of N from 50 to 800, and the y-axis denotes the bias.

Chapter 3

Bayesian Analysis of a Capture-Recapture Model with Time Effect and Individual Covariates

3.1 Introduction

Capture-recapture experiments are designed to estimate an unknown population size N. Classic closed population capture-recapture models, commonly referred as M_0, M_t, M_b, M_h, M_{tb}, M_{th} , M_{bh}, and M_{tbh} by Otis et al. (1978), have been studied extensively by many authors. A number of methods, such as maximum likelihood, the method of moments, the jackknife, martingales, and Bayesian methods have been developed to analyze these models. A list of references can be found in Wang et al. (2007).

In practice, heterogeneity in capture probabilities is a very common, sometimes crucial, assumption. Failure to account for existing heterogeneity may result in underestimation of N (Otis et al. 1978, Dorazio & Royle 2003). When individual information is unavailable, several approaches have been developed to model heterogeneity. For example, the M_h model assumes that each member in the population has its own capture probability. N can be estimated with the jackknife estimator (Burnham & Overton 1978). Latent class models (Agresti 1994) and Beta-binomial mixture models (Dorazio & Royle 2003) are also widely used.

Often, individual data such as gender, life stage, and other physical measurements are collected during the experiment. It is tempting to involve these covariates into the model. Huggins (1989) studied capture-recapture models involving both individual and environmental covariates with a conditional likelihood approach. Stoklosa et al. (2011) concluded that the relative efficiency loss for using conditional likelihood was around 5% and 15% in most situations. Royle (2009) studied the M_h and M_{bh} models involving continuous individual covariates by using a data augmentation scheme. Pledger (2000) gave a unified linear-logistic framework based on various finite partitions of the individuals. This highly general framework includes all eight models in Otis et al. (1978).

In this article, we introduce a novel capture-recapture model that incorporates both individual information and time effects. Bayesian inference for this model is discussed in detail. The paper is organized as follows. In section 3.2, we introduce the model structure and present the full likelihood function. Several priors are discussed. Sufficient conditions for posterior propriety are given and a proof is also provided. We also develop an MCMC scheme to draw posterior samples. In section 3.3, we illustrate our method through the analysis of Softshell Turtle capture history data.

3.2 Model Specification

3.2.1 Likelihood

Typical capture-recapture data are represented by binary numbers x_{ij}, $i = 1, \ldots, N$ and $j = 1, \ldots, k$, where N is the unknown population size and k is the number of sampling occasions. $x_{ij} = 1$ means that the ith individual was captured during the jth sampling occasion, $x_{ij} = 0$ otherwise. In addition, we denote n as the number of individuals that were captured at least once, and p_{ij} as the unknown capture probability for the ith individual in the jth sampling occasion. The primary goal of the capture-recapture experiment is to estimate N.

Without the covariates, the model was named M_{th} by Otis et al. (1978). The density function of $\boldsymbol{X} = \{x_{ij}\}$ given N and $\boldsymbol{P} = \{p_{ij}\}$ can be written as

$$p(\boldsymbol{X} \mid N, \boldsymbol{P}) \propto \frac{N!}{(N-n)!} \prod_{i=1}^{N} \prod_{j=1}^{k} p_{ij}^{x_{ij}} (1 - p_{ij})^{1 - x_{ij}}. \tag{3.1}$$

In many situations, individual data such as gender, age, and other physical measurements for the captured animals are collected. We denote $\boldsymbol{z}_i$ as the covariates for the ith captured individual, where $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{it})'$ is a column vector of length t. In addition, we use the vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ to account for the time effect on capture probabilities. Therefore, the logit of p_{ij} can be modeled by $\boldsymbol{z}_i$ and μ_j as such:

$$\log \frac{p_{ij}}{1 - p_{ij}} = \mu_j + \boldsymbol{z}_i' \boldsymbol{\beta}, \quad i = 1, \ldots, n, \tag{3.2}$$

$$\log \frac{p_{ij}}{1 - p_{ij}} = \mu_j, \quad i = n+1, \ldots, N, \tag{3.3}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_t)'$ is the vector of regression coefficients. Combining (3.1), (3.2), and (3.3), we have the following likelihood function of parameters $(N, \boldsymbol{\beta}, \boldsymbol{\mu})$ given data $\boldsymbol{X}$:

$$L(\mathrm{N}, \boldsymbol{\beta}, \boldsymbol{\mu} \mid \boldsymbol{X}) \propto \frac{N!}{(N-n)!} \prod_{i=1}^{n} \prod_{j=1}^{k} \frac{e^{x_{ij}(\mu_j + \boldsymbol{z}_i' \boldsymbol{\beta})}}{1 + e^{\mu_j + \boldsymbol{z}_i' \boldsymbol{\beta}}} \prod_{j=1}^{k} \left(\frac{1}{1 + e^{\mu_j}} \right)^{N-n}. \tag{3.4}$$

3.2.2 Priors

Since historical information is often unavailable, we apply non-informative priors on $(N, \boldsymbol{\beta}, \boldsymbol{\mu})$. Commonly used non-informative priors for N include

$$\pi_U(N) = 1 \text{ and } \pi_J(N) = 1/N,$$

the uniform prior and the Jeffreys prior (see Smith 1991, Basu & Ebrahimi 2001, Wang et al. 2007). Recently, Xu et al. (2012) developed the parameter-based asymptotic objective prior under the M_t model:

$$\pi_A(N) = (\log N)^{\frac{k-1}{2}} / N^{\frac{1}{4}}.$$

They also illustrated by simulation that $\pi_A(N)$ lead to smaller bias than $\pi_U(N)$ and $\pi_J(N)$ when $k \geq 6$. We will consider these three priors for N in the real data example.

The ideal joint prior for regression coefficients $\boldsymbol{\beta}$ and intercepts $\boldsymbol{\mu}$ is the uniform prior in $\mathbb{R}^{k+t}$. However, posterior propriety under the uniform prior is unclear. Therefore, we consider the prior for $(\boldsymbol{\beta}, \boldsymbol{\mu})$ as $\pi(\boldsymbol{\beta}, \boldsymbol{\mu}) = 1_{\{\mu_1 \geq -M\}}$. This prior restricts μ_1 in the interval $[-M, \infty)$, M being a large positive number, and leaves other parameters uniform in $\mathbb{R}^{k+t-1}$.

3.2.3 The Posterior Propriety

Before presenting the main theorem, we need the following notation. For $i = 1, \dots, n$ and $j = 1, \dots, k$, define the $k + t$ dimensional vector

$$\boldsymbol{y}_{ij} = (0, \dots, 1, \dots, 0, z_{i1}, \dots, z_{it})^T,$$

where the 1 is at the jth position. The remaining definitions $(\mathcal{S}, \mathcal{F})$ and conditions (C1,C2) in this section were originally given by Silvapulle (1981) for a binomial response model, and then restated in slightly different forms by Speckman et al. (2009) for a general multinomial choice model. We adopt Speckman et al.'s version with modification. Define

$$\begin{aligned} \mathcal{S} &= \left\{ \sum_{i,j,x_{ij}=1} \lambda_{ij} \boldsymbol{y}_{ij}, \quad \lambda_{ij} \geq 0 \right\}, \\ \mathcal{F} &= \left\{ \sum_{i,j,x_{ij}=0} \lambda_{ij} \boldsymbol{y}_{ij}, \quad \lambda_{ij} \geq 0 \right\}. \end{aligned} \tag{3.5}$$

The following conditions are essential for the propriety of posterior in a number of multinomial choice models. If a dataset satisfies both conditions, it is said to be 'overlapped'.

(C1) The rank of $\boldsymbol{Y} = (\boldsymbol{y}_{11}, \dots, \boldsymbol{y}_{nk})$ is $k + t$.

(C2) Either $\mathcal{S} \cap \mathcal{F} = \emptyset$ or one of $\mathcal{S}, \mathcal{F} = \mathbb{R}^{k+t}$.

In the following theorem, we give a sufficient condition for posterior propriety under the prior

$$\pi(N, \boldsymbol{\beta}, \boldsymbol{\mu}) = 1_{\{\mu_1 \geq -M\}}. \tag{3.6}$$

Theorem 3.2.1. *Under the likelihood (3.4) and the prior (3.6), the joint posterior of $(N, \boldsymbol{\beta}, \boldsymbol{\mu})$ is proper if C1 and C2 are satisfied.*

Proof. Let

$$\gamma(\boldsymbol{\beta}, \boldsymbol{\mu}) = \prod_{j=1}^{k} \prod_{i=1}^{n} \frac{e^{x_{ij}(\mu_j + \boldsymbol{z}_i' \boldsymbol{\beta})}}{1 + e^{\mu_j + \boldsymbol{z}_i' \boldsymbol{\beta}}}.$$

The marginal posterior density of $(\boldsymbol{\beta}, \boldsymbol{\mu})$ given $\boldsymbol{X}$ can be bounded as

$$\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{\mu} \mid \boldsymbol{X}) &\propto \sum_{N=n}^{\infty} L(N, \boldsymbol{\beta}, \boldsymbol{\mu} \mid \boldsymbol{X}) \pi(N, \boldsymbol{\beta}, \boldsymbol{\mu}) \\
&= \sum_{N=n}^{\infty} \frac{N!}{(N-n)!} \prod_{j=1}^{k} \left(\frac{1}{1 + e^{\mu_j}} \right)^{N-n} \gamma(\boldsymbol{\beta}, \boldsymbol{\mu}) 1_{\{\mu_1 \geq -M\}} \\
&= n! \left(1 - \prod_{j=1}^{k} \frac{1}{1 + e^{\mu_j}} \right)^{-(n+1)} \gamma(\boldsymbol{\beta}, \boldsymbol{\mu}) 1_{\{\mu_1 \geq -M\}} \\
&\leq n! \left(1 - \frac{1}{1 + e^{\mu_1}} \right)^{-(n+1)} \gamma(\boldsymbol{\beta}, \boldsymbol{\mu}) 1_{\{\mu_1 \geq -M\}} \\
&\leq C \gamma(\boldsymbol{\beta}, \boldsymbol{\mu}) 1_{\{\mu_1 \geq -M\}},
\end{aligned}$$

where $C = n! \left(1 - \frac{1}{1+e^{-M}}\right)^{-(n+1)}$. It follows from Speckman et al. (2009) that when (C1) and (C2) are satisfied, $\gamma(\boldsymbol{\beta}, \boldsymbol{\mu})$ can be bounded by

$$\gamma(\boldsymbol{\beta}, \boldsymbol{\mu}) \leq H(-C_1 \|\boldsymbol{\theta}\|). \tag{3.7}$$

On the right hand side of (3.7), C_1 is a positive constant, $\| \cdot \|$ is the Euclidean norm, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\mu}')'$, and $H(t) = P(\xi_1 - \xi_2 \leq t)$, where ξ_1 and ξ_2 follow iid the Type I extreme value distribution (Gumbel distribution). Similar as in Speckman et al. (2009), by using a polar transformation, we get

$$\begin{aligned}
\int_{\mathbb{R}^{k+t}} p(\boldsymbol{\beta}, \boldsymbol{\mu} \mid \boldsymbol{X}) d\boldsymbol{\beta} d\boldsymbol{\mu} &\leq C \int_{\mathbb{R}^{k+t}} H(-C_1 \|\boldsymbol{\theta}\|) 1_{\{\mu_1 \geq -M\}} d\boldsymbol{\beta} d\boldsymbol{\mu} \\
&\leq C_2 \int_0^{\infty} r^{k+t-1} H(-r) dr = C_2 \int_0^{\infty} r^{k+t-1} \int_{-\infty}^{-r} dH(s) dr \\
&= C_2 \int_{-\infty}^{0} \int_0^{-s} r^{k+t-1} dr dH(s) \leq C_3 \int_{-\infty}^{\infty} |s|^{k+t} dH(s).
\end{aligned}$$

Since ξ_1 has a $(k+t)$th finite moment, the right hand side is finite.

Note that $\mathbb{R}^{k+t}$ with Lebesgue measure and $\mathbb{N}$ with counting measure are both σ-finite. Therefore, by Fubini's Theorem, we can exchange the order of summation and integration and get

$$\begin{aligned}\sum_{N=n}^{\infty} p(N \mid \boldsymbol{X}) &= \sum_{N=n}^{\infty} \left\{ \int_{\mathbb{R}^{k+t}} L(N, \boldsymbol{\beta}, \boldsymbol{\mu} \mid \boldsymbol{X}) \pi(\boldsymbol{\beta}, \boldsymbol{\mu}) d\boldsymbol{\beta} d\boldsymbol{\mu} \right\} \pi_U(N) \\ &= \int_{\mathbb{R}^{k+t}} \left\{ \sum_{N=n}^{\infty} L(N, \boldsymbol{\beta}, \boldsymbol{\mu} \mid \boldsymbol{X}) \pi_U(N) \right\} \pi(\boldsymbol{\beta}, \boldsymbol{\mu}) d\boldsymbol{\beta} d\boldsymbol{\mu} < \infty.\end{aligned}$$

This proves Theorem 3.2.1. $\square$

Remark 3.2.1. *Since the tails of $\pi_J(N) = 1/N$ and $\pi_A(N) = (\log N)^{\frac{k-1}{2}}/N^{\frac{1}{4}}$ are smaller than 1, the posterior is proper under these two priors, given C1 and C2 are satisfied.*

3.2.4 Parameter Estimation

A Gibbs sampling scheme is designed to generate posterior samples of $(N, \boldsymbol{\beta}, \boldsymbol{\mu})$. The full conditional densities for $(\boldsymbol{\beta}, \boldsymbol{\mu})$ are given by

$$\begin{aligned}\pi(\boldsymbol{\beta} \mid N, \boldsymbol{\mu}, \boldsymbol{X}) &\propto \prod_{j=1}^{k} \prod_{i=1}^{n} \frac{e^{x_{ij}(\mu_j + \boldsymbol{z}_i' \boldsymbol{\beta})}}{1 + e^{\mu_j + \boldsymbol{z}_i' \boldsymbol{\beta}}}, \\ \pi(\boldsymbol{\mu} \mid N, \boldsymbol{\beta}, \boldsymbol{X}) &\propto \frac{e^{x_{ij}(\mu_j + \boldsymbol{z}_i' \boldsymbol{\beta})}}{1 + e^{\mu_j + \boldsymbol{z}_i' \boldsymbol{\beta}}} \prod_{j=1}^{k} \left(\frac{1}{1 + e^{\mu_j}} \right)^{N-n} 1_{\{\mu_1 \geq -M\}}.\end{aligned}$$

In each cycle, $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ can be sampled from their full conditional distributions by using the Metropolis algorithm with a normal proposal density.

The sampling method for N depends on its prior. Under $\pi_U(N)$ and $\pi_J(N)$, the full conditional distribution $\pi(N \mid \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{X})$ follows a negative binomial distribution, which is easy to sample from. Under $\pi_A(N)$, $\pi(N \mid \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{X})$ has the form

$$\pi(N \mid \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{X}) \propto \frac{(\log N)^{(k-1)/2} N!}{N^{1/4}(N-n)!} \left(\prod_{j=1}^{k} \frac{1}{1+e^{\mu_j}} \right)^{N-n}, \quad \text{for } N \geq n. \tag{3.8}$$

We use the ratio of two uniforms method to sample N from (3.8). The method was originally introduced by Kinderman & Monahan (1977) to generate continuous random variables from non-standard densities. To be specific, for any integrable function $f(\cdot)$, we generate (U, V) uniformly from the region $0 < U < \sqrt{f(V/U)}$. The ratio V/U then has density $f(x)/\int f(x)dx$. To adjust for discrete parameter N, we define step function $\pi^*(\cdot)$ with domain $[n, \infty)$, where $\pi^*(x) = \pi(N \mid \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{X})$ for $x \in [N, N+1)$. $\pi^*(\cdot)$ can be viewed as the limit of a sequence of integrable continuous functions. We then apply a slightly modified ratio of two uniforms method. First, let $b_1 = max_{x \geq n} \sqrt{\pi^*(x)}$ and $b_2 = max_{x \geq n} x \sqrt{\pi^*(x)}$. Generate $U \sim Unif(0, b_1)$ and $V \sim Unif(0, b_2)$. It is easy to show that $(0, b_1) \times (0, b_2)$ covers the region $0 < U < \sqrt{\pi^*(V/U)}$. Second, accept $N_{new} = [V/U]$ if $V/U \geq n$ and $0 < U < \sqrt{\pi^*(V/U)}$, where $[x]$ is the largest integer that is smaller than x.

3.3 Softshell Turtle Example

3.3.1 Data Description and Preliminary Analysis

A Capture-recapture experiment on Softshell Turtle was conducted by researchers from the Department of Wildlife and Fisheries at University of Missouri in the summer of 2011. Softshell Turtles were commercially harvested at a number of sites in three rivers (Gasconade River, Missouri River, and Osage River). We use the capture history data from site 94 in Missouri River (94 indicates the river mile) to demonstrate our model. At this site, 90 turtles were captured at least once during 8 trapping nights. Each turtle, after being captured, was marked with a unique PTI number, so that it could be identified in future captures. The gender (Female or Male) and life stage (Adult or Juvenile) information for each captured turtle was collected. The data structure is summarized in Table 3.1.

We will use the model described in section 3.2 to analyze this dataset. The model can be justified by some preliminary analyses. First, notice that the numbers of captures in 8 trapping nights are $(11, 15, 16, 9, 14, 18, 21, 23)$. To explain the variation, it is reasonable to assume that the capture probabilities vary across sampling occasions. Second, suppose Y_i is the number of sampling occasions during which the ith turtle was captured. The averages of Y_i's for adult and juvenile turtles are 1.31 and 2.08 respectively, indicating that juvenile turtles are in much higher risk of being captured than adult turtles. The same quantities for female and male turtles are 1.44 and 1.37, indicating male turtles are slightly harder to capture. Therefore it is also proper to assume that gender and life stage have influence on capture probabilities.

The dataset satisfies conditions C1 and C2, so the inference is based on proper posterior distribution. In fact, C1 and C2 are true as long as during one of the sampling occasions, in one of the gender by life stage combinations, there exists at least two distinct capture histories.

3.3.2 Results

We use the algorithm described in section 3.2.4 to sample $(N, \boldsymbol{\beta}, \boldsymbol{\mu})$. Three priors π_U, π_J, and π_A are considered for N. For $(\boldsymbol{\beta}, \boldsymbol{\mu})$, we use $\pi(\boldsymbol{\beta}, \boldsymbol{\mu}) = 1_{\{\mu_1 \geq -M\}}$, where M is set to be 10^{20}. Parameters are estimated with posterior means. We consider four different models on p_{ij} as below. For $i = 1, \ldots, n$ and $j = 1, \ldots, k$,

$$\begin{aligned}
M_1: &\quad p_{ij} = p_j,\\
M_2: &\quad \log \frac{p_{ij}}{1-p_{ij}} = \mu_j + \beta_{1i}\text{Gender}_i,\\
M_3: &\quad \log \frac{p_{ij}}{1-p_{ij}} = \mu_j + \beta_{2i}\text{Stage}_i,\\
M_4: &\quad \log \frac{p_{ij}}{1-p_{ij}} = \mu_j + \beta_{1i}\text{Gender}_i + \beta_{2i}\text{Stage}_i.
\end{aligned}$$

In the models, Gender and Stage are binary categorical variables. Gender equals 1 for female and 0 for male. Stage equals 1 for juvenile and 0 for adult. DIC and pD for these models are also calculated. The posterior means of N under π_U, π_J, π_A are $177, 175, 177$ respectively. Although the estimates are very close, we recommend to use π_A or π_U because Xu et al. (2012) showed by simulation that N might be underestimated under π_J in M_t models when k is large. The posterior probability mass functions of N in model M_4 are plotted under the three priors. In each model, 10^4 posterior samples were drawn, and the first 10^3 samples were discarded. Estimates of $(N, \boldsymbol{\beta})$ and DIC values under $\pi_A(N)$ are summarized in Table 2.

Note that $\hat{\beta}_1$ and $\hat{\beta}_2$ are significantly higher than 0 in M_2, M_3, and M_4. It seems that female and juvenile turtles are at higher risk of being captured than male and adult turtles. The estimates of N under M_1, M_2, M_3, and M_4 are $153, 174, 164$, and 177, respectively. Otis et al. (1978) pointed out that ignoring heterogeneity would cause underestimation of N. It seems that that N is underestimated in model M_1. Relatively higher DIC value also indicates that M_1 is not proper for this dataset. Since both gender and life stage affect the capture probabilities, it indicates that M_4 fits the data best among the 4 models. The lowest DIC score confirms this.

Table 3.1: Softshell Turtle capture history data at site No. 94 in Missouri River (2011). In total, 90 turtles were captured for at least once during 8 sampling occasions. In each sampling occasion, '1' means the turtle was captured and '0' means the turtle was not captured. Gender (Female/Male) and life stage (Adult/Juvenile) were recorded for each captured turtle.

		Sampling occasions					
	Turtle	1	2	$\cdots$	8	Gender	Life Stage
Captured	1	0	1	$\cdots$	0	Male	Adult
	2	1	0	$\cdots$	0	Female	Adult
	$\vdots$			x_{ij}		$\vdots$	$\vdots$
	90	0	1	$\cdots$	1	Female	Juvenile
Uncaptured	$N-90$	0	0	$\cdots$	0	N.A.	N.A.

Table 3.2: Parameter estimates and DIC for models M_1, M_2, M_3, M_4. The numbers in parentheses are standard deviations for the estimates.

Model	$\hat{N}$	$\hat{\beta}_1$ (Gender)	$\hat{\beta}_2$ (Stage)	DIC
M_1	153(18)			-15.9
M_2	174(17)	0.87(0.19)		-42.2
M_3	164(15)		1.23(0.25)	-47.2
M_4	177(17)	0.62(0.22)	0.78(0.31)	-52.4

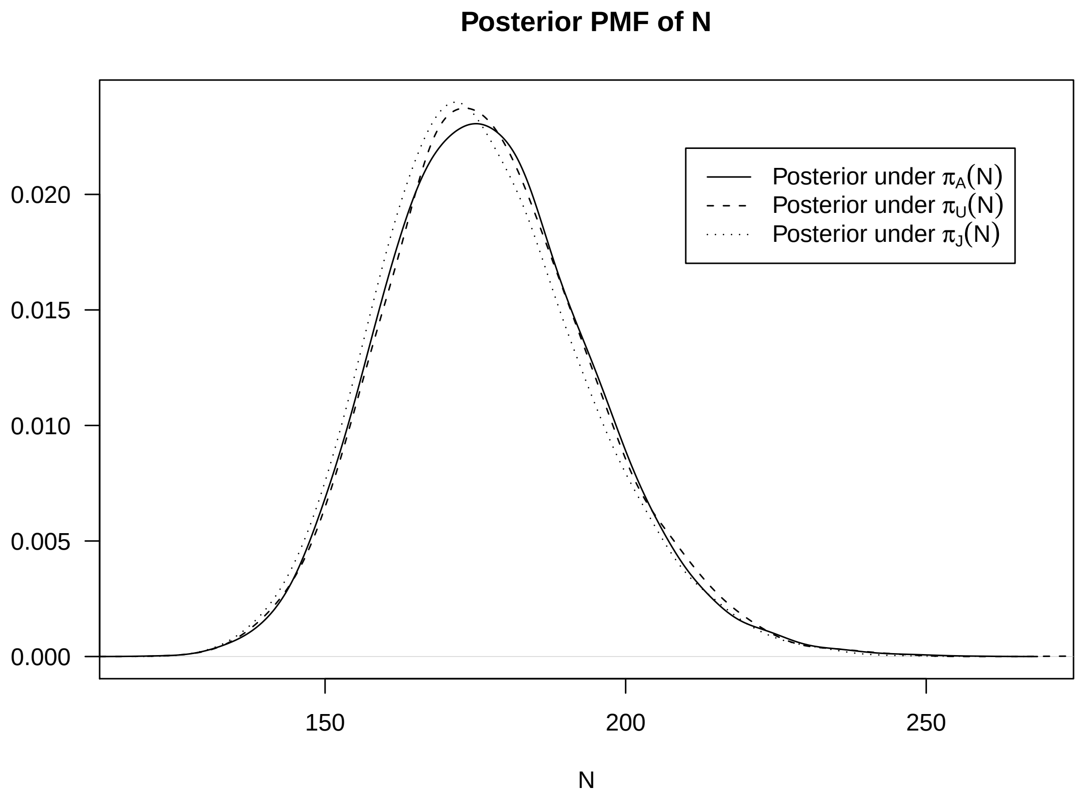

Figure 3.1: Posterior probability mass function of N in the full model M_4. The solid line is the PMF under $\pi_A(N)$. The dashed line is the PMF under $\pi_U(N)$. The dotted line is the PMF under $\pi_J(N)$.

Chapter 4

Estimating Population Size with Exponential Failure under Type-I Censoring

4.1 Introduction

In general, let us consider a system containing an unknown number N of faults, and each fault could cause a failure. The failure time of each fault is assumed to follow an i.i.d. exponential distribution, whose density is given by

$$p(t_j \mid \beta) = \beta e^{-\beta t}, \quad j = 1, \ldots, N. \tag{4.1}$$

Due to time or budget limits, the system is only observed until a fixed time T, and the type-I censored failure times are noted as $0 < t_1 < t_2 < \cdots < t_n < T$, where n is the random number of failures that occurred prior to time T. Our goal is to estimate N as well as the nuisance parameter β.

It is a very challenging task to estimate N. Many commonly used estimation methods fail in this problem. For instance, Blumenthal & Marcus (1975) showed that when $R > \frac{N+1}{2N}$, the maximum likelihood estimator of N became infinite. The authors modified the maximum likelihood estimator by an approximation with stochastically order $O(\frac{1}{\sqrt{N}})$. Goudie & Goldie (1981) showed that an unbiased estimate of N based on the sufficient statistics does not exist. Moreover, they proved that the moment estimator does not exist either when $\frac{1}{nT}\sum_{j=1}^{n} t_j > \frac{1}{2}$. Meanwhile, several less common frequentist methods appears work, including the conditional maximum likelihood estimator (CMLE) of Blumenthal & Marcus (1975) and the harmonic mean estimator (HME) of Joe & Reid (1985). Joe & Reid (1985) suggested to estimate N by the harmonic mean of two end points of the likelihood function. More specifically, for a given constant $c \in (0, 1]$, find a lower endpoint n_1 and an upper endpoint n_2, where

$$n_1 = \inf\{n \geq r : f(\boldsymbol{t} \mid n, R) \geq f(\boldsymbol{t} \mid \hat{N}_{MLE}, R)\}, \tag{4.2}$$

$$n_2 = \sup\{n \geq r : f(\boldsymbol{t} \mid n, R) \geq f(\boldsymbol{t} \mid \hat{N}_{MLE}, R)\}. \tag{4.3}$$

Let $\tilde{N}$ be the closest integer to the harmonic mean of n_1 and n_2, namely, $2/(n_1^{-1}+n_2^{-1})$. The harmonic mean estimator exists in the case that $\hat{N}_{MLE} = \infty$.

Bayesian methods provide very useful tools in estimating discrete parameters. For example, Blumenthal & Marcus (1975) suggested using a gamma prior for β and an uniform prior for N. Raftery (1988) treated N as a latent variable with a Poisson distribution and used non-informative priors on the hyper-parameters. Berger et al. (2009b) summarized four approaches to develop objective priors on general discrete parameters. In this paper, while assuming Poisson distribution on N, we will develop Jeffreys and reference priors on other parameters based on different parameterizations. Simulation studies are done to compare the frequentist properties of various priors.

4.2 Hierarchical Model

4.2.1 Priors

To make the model unit-less, we let $\delta = \beta T$ and $r_j = t_j/T$. The joint density for the transformed failure times $\boldsymbol{r} = (r_1, r_2, \cdots, r_n)$ given (N, δ) is then

$$f(\boldsymbol{r} \mid N, \delta) = \frac{N!}{(N-n)!}\delta^n e^{-\delta(N-n+R)}, \tag{4.4}$$

where $R = \sum_{j=1}^{n} r_j$ and $0 < R < n$. Clearly the sufficient statistics of (N, δ) are (R, n). The conditional distribution of R given δ does not depend on N and its density was given by Goudie & Goldie (1981) :

$$p(R \mid n, \delta) = \delta^n e^{-\delta R}(1 - e^{-\delta})^{-n} g_{n+1}(R+1), \quad 0 < R < n, \tag{4.5}$$

where

$$g_j(\alpha) = \frac{1}{(j-2)!}\sum_{k=1}^{[\alpha]}(-1)^{k-1}\begin{pmatrix} j-1 \\ k-1 \end{pmatrix}(\alpha - k)^{j-2}, \quad 1 < \alpha < j. \tag{4.6}$$

Consequently, n follows a binomial distribution, and its probability mass function can be written as

$$p(n \mid N, \delta) = \frac{N!}{n!(N-n)!}(1 - e^{-\delta})^n e^{-\delta(N-n)}, \quad n = 0, 1, \cdots, N. \tag{4.7}$$

The joint density of (R, n) is then

$$p(n, R \mid N, \delta) = \frac{N!}{n!(N-n)!}\delta^n e^{-\delta(N-n+R)} g_{n+1}(R+1), \tag{4.8}$$

for $0 \leq R \leq n \leq N$.

It is both justifiable and practical to treat N as a latent variable with a Poisson(λ) distribution. An interpretation of the hierarchical structure was given in Scholz (1986) and Raftery (1988) by treating the failure times as event times from a non-homogeneous Poisson process. From a practical point of view, since the Poisson distribution is conjugate with the likelihood function given in (4.4), we can sum over N and work on the continuous parameters (δ, λ). The likelihood function of (δ, λ) is given by

$$
\begin{aligned}
L(\delta, \lambda \mid \boldsymbol{r}) &= \sum_{N=n}^{\infty} f(\boldsymbol{r} \mid \delta, N)\pi(N \mid \lambda) \\
&= \sum_{N=n}^{\infty} \frac{N!}{(N-n)!}\delta^n e^{-\delta(N-n+R)}\frac{\lambda^N}{N!}e^{-\lambda} \\
&= \sum_{N=n}^{\infty} \frac{(\lambda e^{-\delta})^{N-n}}{(N-n)!}(\delta\lambda)^n e^{-\delta R-\lambda} \\
&\propto (\delta\lambda)^n \exp(-\delta R - \lambda(1-e^{-\delta})). \qquad (4.9)
\end{aligned}
$$

We consider a family of ad-hoc non-informative priors on (δ, λ),

$$\pi_{a,b}(\delta, \lambda) = \frac{1}{\delta^a \lambda^b}, \quad \delta > 0, \lambda > 0, \qquad (4.10)$$

where a and b are nonnegative constants.

Proposition 4.2.1. *Under prior* $\pi_{a,b}(\delta, \lambda)$*,*

(a) *The marginal posterior distribution of* N *is proper if and only if* $0 \leq a < b$*.*

(b) *The* k*th posterior moment of* N *exists if* $b - a > k$*.*

Proof. Under $\pi_{a,b}(\delta, \lambda)$, the posterior probability mass function of N can be calculated as

$$\begin{aligned}
f(N \mid \boldsymbol{r}, \delta, \lambda) &\propto \int_0^\infty \int_0^\infty f(\boldsymbol{r} \mid N, \delta)\pi(N \mid \lambda)\pi_{a,b}(\delta, \lambda) d\delta d\lambda \\
&= \int_0^\infty \int_0^\infty \frac{N!}{(N-n)!}\delta^n e^{-\delta(N-n+R)} \frac{\lambda^N e^{-\lambda}}{N!} \frac{1}{\delta^a \lambda^b} d\delta d\lambda \\
&= \frac{\Gamma(N-b+1)}{\Gamma(N-n+1)} \int_0^\infty \delta^{n-a} e^{-\delta(N-n+R)} d\delta \\
&= \frac{\Gamma(N-b+1)\Gamma(n-a+1)}{\Gamma(N-n+1)\Gamma(N-n+R)^{n-a+1}} \\
&\sim \frac{1}{N^{b-a+1}}. \qquad (4.11)
\end{aligned}$$

The proposition is proved by (4.11). □

We consider two members of this family of priors: $\pi_{1,2}(\delta, \lambda)$ and $\pi_{0,1}(\delta, \lambda)$. According to Proposition 4.2.1, both priors will lead to proper posterior distributions of N, with infinite moments. Raftery (1988) defined $\rho = \beta\lambda$ and assumed $\pi(\rho, \beta) = \rho^{-2}$. In fact, Raftery's prior is equivalent to $\pi_{1,2}(\delta, \lambda)$ after transformation. One can argue that $\pi_{1,2}$ is not 'diffuse' enough in the sense that the marginal prior for N is $1/N(N+1)$, which is a proper prior. For this reason, we take an alternative prior, $\pi_{0,1}$, into consideration. The marginal prior for N based on $\pi_{0,1}$ is $1/N$, one of the most commonly used priors for population sizes parameters. The drawback of these ad-hoc priors is that the selection of (a, b) seems arbitrary and unjustified. Moreover, although (a, b) may be chosen to get a diffuse prior as well as a proper posterior distribution, the prior selected may not achieve desirable frequentist properties such as invariance over transformation, frequentist matching, or maximum missing information (see Jeffreys 1961, Mukerjee & Ghosh 1997, Berger et al. 2009b).

As a commonly used objective prior, the Jeffreys prior is invariant to transformation and has a tendency to produce a proper posterior. It is straightforward to develop the Jeffreys prior, as shown in the following proposition.

Proposition 4.2.2.

(a) *The Jeffreys' prior for* (δ, λ) *is*

$$\pi_1^J(\delta, \lambda) = \sqrt{\frac{(1-e^{-\delta})^2}{\delta^2} - e^{-\delta}}. \tag{4.12}$$

(b) *Under* $\pi_1^J(\delta, \lambda)$*, the posterior distribution of* N *is proper but has no finite moments.*

Proof. We starting from deriving Jeffreys' prior and proving part (a). It follows from (4.7) and $\pi(N \mid \lambda)$ that

$$\begin{aligned} \mathrm{E}(n \mid \delta, \lambda) &= \mathrm{E}(\mathrm{E}(n \mid N, \delta) \mid \lambda) \\ &= \mathrm{E}(N(1-e^{-\delta}) \mid \lambda) = \lambda(1-e^{-\delta}). \end{aligned} \tag{4.13}$$

Then the Fisher information matrix of (δ, λ) can be shown to be

$$I(\delta, \lambda) = -\mathrm{E}\begin{pmatrix} -\frac{n}{\delta^2} + \lambda & e^{-\delta} \\ e^{-\delta} & \frac{n}{\lambda^2} \end{pmatrix} = \begin{pmatrix} \lambda\left(\frac{1-e^{-\delta}}{\delta^2}\right) - e^{-\delta} & e^{-\delta} \\ e^{-\delta} & \frac{1-e^{-\delta}}{\lambda} \end{pmatrix}. \tag{4.14}$$

Now we prove Part (b). For convenience, define $\pi_1^J(N \mid \boldsymbol{r})$ to be the posterior probability mass function of N.

$$\begin{aligned} \pi_1^J(N \mid \boldsymbol{r}) &= \int_0^\infty \int_0^\infty f(\boldsymbol{r} \mid \delta, N)\pi(N \mid \lambda)\pi_1^J(\lambda, \delta) d\lambda d\delta \\ &= \int_0^\infty \int_0^\infty \frac{N!}{(N-n)!}\delta^n e^{-\delta(N-n+R)}\sqrt{\frac{(1-e^{-\delta})^2}{\delta^2} - e^{-\delta}}\frac{\lambda^N}{N!}e^{-\lambda} d\lambda d\delta \\ &= \int_0^\infty \frac{N!}{(N-n)!}\delta^n e^{-\delta(N-n+R)}\sqrt{\frac{(1-e^{-\delta})^2}{\delta^2} - e^{-\delta}} d\delta. \end{aligned} \tag{4.15}$$

Define

$$J_1 = \int_0^\infty \sum_{N=n}^\infty \frac{N!}{(N-n)!}\delta^n e^{-\delta(N-n+R)}\sqrt{\frac{(1-e^{-\delta})^2}{\delta^2} - e^{-\delta}}d\delta$$
$$= n!\int_0^\infty \frac{\delta^n e^{-\delta R}}{(1-e^{-\delta})^{n+1}}\sqrt{\frac{(1-e^{-\delta})^2}{\delta^2} - e^{-\delta}}d\delta. \tag{4.16}$$

For fixed $\epsilon > 0$, we consider the integral from 0 to ϵ:

$$n!\int_0^\epsilon \frac{\delta^n e^{-\delta R}}{(1-e^{-\delta})^{n+1}}\sqrt{\frac{(1-e^{-\delta})^2}{\delta^2} - e^{-\delta}}d\delta \le n!\int_0^\epsilon \frac{\delta^n}{(1-e^{-\delta})^n}\sqrt{\frac{1}{\delta^2} - \frac{e^{-\delta}}{(1-e^{-\delta})^2}}d\delta. \tag{4.17}$$

Since $\lim_{\delta\to 0}\frac{\delta}{1-e^{-\delta}} = 1$, there exists a small $\epsilon > 0$, such that for fixed $\eta > 0$,

$$\frac{\delta^n}{(1-e^{-\delta})^n} < 1+\eta. \tag{4.18}$$

Therefore

$$\text{The RHS of (4.17)} \le \frac{(1+\eta)n!}{T^n}\int_0^\epsilon g(\delta)d\delta, \tag{4.19}$$

where

$$g(\delta) = \sqrt{\frac{1}{\delta^2} - \frac{e^{-\delta}}{(1-e^{-\delta})^2}}$$
$$= \sqrt{\frac{1}{\delta^2} - \frac{1}{e^{-\delta}+e^{\delta}-2}} = \sqrt{\frac{1}{\delta^2} - \frac{1}{\delta^2 + 2\delta^4/4! + 2\delta^6/6!i + \cdots}}$$
$$\le \sqrt{\frac{1}{\delta^2} - \frac{1}{\delta^2+\delta^4/12}} \le \frac{1}{\sqrt{12}}. \tag{4.20}$$

Consequently,

$$\text{The RHS of (4.17)} \le \frac{(1+\eta)n!\epsilon}{\sqrt{12}T^n} < \infty. \tag{4.21}$$

On the other hand,

$$
\begin{aligned}
& n! \int_{1/\epsilon}^{\infty} \frac{\delta^n e^{-\delta R}}{(1-e^{-\delta})^{n+1}} \sqrt{\frac{(1-e^{-\delta})^2}{\delta^2} - e^{-\delta}} d\delta \\
\leq \;& \frac{n!}{(1-e^{-\epsilon})^{n+1}} \int_{1/\epsilon}^{\infty} \delta^{n-1} e^{-\delta R} d\delta < \infty. \qquad (4.22)
\end{aligned}
$$

It follows from (4.21) and (4.22) that $J_1 < \infty$. By Fubini's theorem, we can exchange the order of summation and integration and prove that the posterior distribution of N is proper,

$$
\sum_{N=n}^{\infty} \pi_1^J(N \mid \boldsymbol{r}) = J_1 < \infty. \qquad (4.23)
$$

However we can show similarly that the posterior moments of N do not exist:

$$
\begin{aligned}
& \mathrm{E}_1^J(N+1 \mid \boldsymbol{r}) = \sum_{N=n}^{\infty} (N+1) \pi_1^J(N \mid \boldsymbol{r}) \\
\geq \;& \int_0^{\epsilon} \frac{(n+1)!\delta^n}{(1-e^{\delta})^{n+1}} \sqrt{\frac{(1-e^{-\delta})^2}{\delta^2} - e^{-\delta}} d\delta \\
\approx \;& \frac{(n+1)!}{\sqrt{12}} \int_0^{\epsilon} \frac{1}{1-e^{\delta}} d\delta \\
\geq \;& \frac{(n+1)!}{\sqrt{12}} \int_0^{\epsilon} \frac{1}{\delta} d\delta = \infty. \qquad (4.24)
\end{aligned}
$$

□

As an alternative objective prior, the reference prior has gained rising popularity in recent years (see Bernardo 1979, Berger & Bernardo 1992, Berger et al. 2009b). The joint reference prior for two parameters depends on the relative order. In this case, since δ is one of the parameters of interest while λ is the hyper-parameter, δ is of higher importance than λ. The following proposition gives the reference prior and examines its posterior propriety.

Proposition 4.2.3.

(a) *The joint reference prior for (δ, λ) is*

$$\pi_1^R(\delta, \lambda) = \sqrt{\frac{1}{\lambda}\left(\frac{1-e^{-\delta}}{\delta^2} - \frac{e^{-\delta}}{1-e^{-\delta}}\right)}. \tag{4.25}$$

(b) *Under $\pi_1^R(\delta, \lambda)$, the posterior distribution of N is proper but has no finite moments.*

Proof. Based on the two-stage algorithm in Berger & Bernardo (1992), the conditional reference prior for λ given δ is

$$\pi_1^R(\lambda \mid \delta) = \sqrt{I_{\lambda\lambda}} \propto \frac{1}{\sqrt{\lambda}}, \tag{4.26}$$

and the marginal prior for δ is

$$\pi_1^R(\delta) = \exp\left[\frac{1}{2}\int \pi_1^R(\lambda \mid \delta) \log\frac{|I|}{|I_{\lambda\lambda}|} d\lambda\right]. \tag{4.27}$$

Note that

$$\frac{|I|}{|I_{\lambda\lambda}|} = \frac{(1-e^{-\delta})\left(\frac{1-e^{-\delta}}{\delta^2} - e^{-\delta}\right) - e^{-2\delta}}{(1-e^{-\delta})/\lambda} = \lambda\left(\frac{1-e^{-\delta}}{\delta^2} - \frac{e^{-\delta}}{1-e^{-\delta}}\right), \tag{4.28}$$

so we have

$$\pi_1^R(\delta) = \exp\left[\frac{1}{2}\int \frac{1}{\sqrt{\lambda}} \log \lambda \left(\frac{1-e^{-\delta}}{\delta^2} - \frac{e^{-\delta}}{1-e^{-\delta}}\right) d\lambda\right] \propto \sqrt{\frac{1-e^{-\delta}}{\delta^2} - \frac{e^{-\delta}}{1-e^{-\delta}}}. \tag{4.29}$$

Therefore the joint reference prior for (δ, λ) is the product of (4.26) and (4.29), which is given in (4.25). Denote $\pi_1^R(N \mid \boldsymbol{r})$ as the posterior probability mass function of N.

It is easy to show that

$$
\begin{aligned}
\pi_1^R(N \mid \boldsymbol{r}) &= \frac{\Gamma(N+1/2)}{\Gamma(N+1)} \int_0^\infty \frac{N!}{(N-n)!} \delta^n e^{-\delta(N-n_R)} \sqrt{\frac{1-e^{-\delta}}{\delta^2} - \frac{e^{-\delta}}{1-e^{-\delta}}} d\delta \\
&\leq \int_0^\infty \frac{N!}{(N-n)!} \delta^n e^{-\delta(N-n_R)} \sqrt{\frac{1-e^{-\delta}}{\delta^2} - \frac{e^{-\delta}}{1-e^{-\delta}}} d\delta. \qquad (4.30)
\end{aligned}
$$

Define

$$
\begin{aligned}
J_2 &= \int_0^\infty \sum_{N=n}^\infty \frac{N!}{(N-n)!} \delta^n e^{-\delta(N-n_R)} \sqrt{\frac{1-e^{-\delta}}{\delta^2} - \frac{e^{-\delta}}{1-e^{-\delta}}} d\delta \\
&= \int_0^\infty \frac{\delta^n e^{-\delta R}}{(1-e^{-\delta})^{n+3/2}} \sqrt{\frac{(1-e^{-\delta})^2}{\delta^2} - e^{-\delta}} d\delta. \qquad (4.31)
\end{aligned}
$$

Similar as in the proof of Proposition 4.2.2, for small $\epsilon > 0$, there exists $\eta > 0$ such that

$$
\int_0^\epsilon \frac{\delta^n e^{-\delta R}}{(1-e^{-\delta})^{n+3/2}} \sqrt{\frac{(1-e^{-\delta})^2}{\delta^2} - e^{-\delta}} d\delta \leq \frac{1+\eta}{\sqrt{12}} \int_0^\epsilon \frac{1}{\sqrt{\delta}} d\delta < \infty. \qquad (4.32)
$$

On the other hand,

$$
\int_{1/\epsilon}^\infty \frac{\delta^n e^{-\delta R}}{(1-e^{-\delta})^{n+3/2}} \sqrt{\frac{(1-e^{-\delta})^2}{\delta^2} - e^{-\delta}} d\delta \leq \frac{1}{(1-e^{-\delta})^{n+2/3}} \int_{1/\epsilon}^\infty \delta^{n-1} e^{-\delta R} d\delta < \infty. \qquad (4.33)
$$

It follows from (4.32) and (4.33) that $J_2 < \infty$. According to Fubini's theorem,

$$
\sum_{N=n}^\infty \pi_1^R(N \mid \boldsymbol{r}) = J_2 < \infty. \qquad (4.34)
$$

Finally, the expected value of $N+1$ is infinite since

$$
\begin{aligned}
&\mathrm{E}(N+1 \mid \boldsymbol{r}) = \\
&\quad \sum_{N=n}^\infty \frac{\Gamma(N+1/2)}{\Gamma(N)} \int_0^\infty \frac{N!}{(N-n)!} \delta^n e^{-\delta(N-n_R)} \sqrt{\frac{1-e^{-\delta}}{\delta^2} - \frac{e^{-\delta}}{1-e^{-\delta}}} d\delta = \infty \qquad (4.35)
\end{aligned}
$$

$\square$

4.2.2 Re-parameterization

We define $\theta = \lambda(1-e^{\delta})$ and consider the transformed parameters (δ, θ). There are two advantages of this re-parameterization. First of all, the new parameter θ is meaningful and equals the conditional expected value of n. To see this, we first notice that given N and δ, n follows a binomial distribution with mean $N(1-e^{-\delta})$. Also, N is assumed to be Poisson distributed, so we have

$$\mathrm{E}(n \mid \lambda, \delta) = \mathrm{E}(N(1-e^{\delta})) = \lambda(1-e^{\delta}) = \theta. \tag{4.36}$$

Secondly, the Fisher information matrix of (δ, θ) is diagonal. The Jeffreys-rule prior is invariant to transformation, so we derive the reference priors in the next proposition.

Proposition 4.2.4. *The reference prior for (δ, θ) is invariant to the transformation $(\delta, \lambda) \to (\delta, \theta)$ and does not depend on the inferential order of δ and θ. The prior, denoted by $\pi_2^R(\delta, \theta)$, is given by*

$$\pi_2^R(\delta, \theta) \propto \sqrt{\frac{1}{\theta}\left[\frac{1}{\delta^2} - \frac{e^{\delta}}{(e^{\delta}-1)^2}\right]}. \tag{4.37}$$

Proof. The likelihood of (δ, θ) is

$$\begin{aligned}
L(\delta, \theta \mid \boldsymbol{r}) &= \sum_{N=n}^{\infty} \frac{N!}{(N-n)!}\delta^n \exp(-\delta(N-n+R))\frac{(\frac{\theta}{1-e^{-\delta}})^N}{N!}\exp\Big(-\frac{\delta}{1-e^{-\delta}}\Big) \\
&= \sum_{N=n}^{\infty} \frac{(\frac{\theta e^{-\delta}}{1-e^{-\delta}})^N}{(N-n)!}\delta^n \exp\Big(-\delta(R-n) - \frac{\theta}{1-e^{-\delta}}\Big) \\
&= \Big(\frac{\theta e^{-\delta}}{1-e^{-\delta}}\Big)^n \exp\Big(\frac{\theta e^{-\delta}}{1-e^{-\delta}}\Big)\delta^n \exp\Big(-\delta(R-n) - \frac{\theta}{1-e^{-\delta}}\Big) \\
&= \Big(\frac{\theta\delta}{1-e^{-\delta}}\Big)^n e^{-\delta R-\theta}.
\end{aligned} \tag{4.38}$$

Since $\mathrm{E}(n \mid \delta, \theta) = \theta$, we can calculate the Fisher information matrix of (δ, θ) as

$$I(\delta, \theta) = \begin{pmatrix} \theta\left(\frac{1}{\delta^2} - \frac{e^\delta}{(e^\delta - 1)^2}\right) & 0 \\ 0 & \frac{1}{\theta} \end{pmatrix}. \tag{4.39}$$

Based on the two-stage algorithm in Berger & Bernardo (1992), the conditional reference prior for θ given δ is

$$\pi_2^R(\theta \mid \delta) = \sqrt{|I_{\theta\theta}|} = \frac{1}{\sqrt{\theta}}, \tag{4.40}$$

and the marginal reference prior for δ is

$$\pi_2^R(\delta) = \exp\left[\frac{1}{2}\int \pi_2^R(\theta \mid \delta) \log \frac{|I|}{|I_{\theta\theta}|} d\theta\right] \propto \sqrt{\frac{1}{\delta^2} - \frac{e^\delta}{(e^\delta - 1)^2}}. \tag{4.41}$$

Combining (4.40) and (4.29), the reference prior for (δ, θ), assuming δ is of higher importance, is

$$\pi_2^R(\delta, \theta) \propto \sqrt{\frac{1}{\theta}\left[\frac{1}{\delta^2} - \frac{e^\delta}{(e^\delta - 1)^2}\right]}. \tag{4.42}$$

One can easily show that the reference prior for (θ, δ), assuming θ is of more interest, remains the same as $\pi_2^R(\delta, \theta)$. Moreover, the reference prior is identical to $\pi_1^R(\delta, \lambda)$ after transformation since

$$\begin{aligned} \pi_2^R(\delta, \lambda) &= \pi_2^R(\delta, \theta(\lambda, \delta))\frac{\partial(\delta, \theta)}{\partial(\delta, \lambda)} \\ &= \sqrt{\frac{1}{\lambda(1 - e^{-\delta})}\left[\frac{1}{\delta^2} - \frac{e^\delta}{(e^\delta - 1)^2}\right]}(1 - e^{-\delta}) \\ &= \sqrt{\frac{1}{\lambda}\left(\frac{1 - e^{-\delta}}{\delta^2} - \frac{e^{-\delta}}{1 - e^{-\delta}}\right)} = \pi_1^R(\delta, \lambda). \end{aligned} \tag{4.43}$$

$\square$

4.3 Simulation Study

The purpose of this simulation study is to compare the performances of four non-informative priors, $\pi_1^J(\delta,\lambda)$, $\pi_1^R(\delta,\lambda)$, $\pi_{1,2}(\delta,\lambda)$, and $\pi_{0,1}(\delta,\lambda)$. In the study, we set $N = 50, 100, 200, 500$ and $\gamma = 0.90, 0.75, 0.50, 0.25, 0.10$, where $\gamma = 1 - e^{-\delta}$ represents the proportion of failures detected prior to the censoring time. For each (N,γ) combination, 200 sets of failure times with censoring time at 1 are generated. For each dataset D_i, $i = 1,\dots,200$, we obtain 5000 posterior samples of (N,δ) via Gibbs sampling.

Under prior π_1^J, the conditional posterior distribution of $N-n$ follows

$$N - n \mid \delta, D_i \sim \text{negbin}(n+1, 1-e^{-\delta}). \tag{4.44}$$

The conditional density of δ is not a known density. So we sample δ with the R function 'arms()', which implements the adaptive rejection Metropolis sampling methods, proposed by Gilks et al. (1995).

Under π_1^R, the conditional posterior distribution of $N-n$ follows a generalize negative binomial distribution:

$$N - n \mid \delta, D_i \sim \text{negbin}(n+0.5, 1-e^{-\delta}). \tag{4.45}$$

The posterior samples of δ are also obtained by 'arms()' function.

The latter two priors give rise to known conditional distributions for both N and δ. Specifically, under $\pi_{1,2}$, we have

$$\begin{aligned} N - n \mid \delta, D_i &\sim \text{negbin}(n-1, 1-e^{-\delta}), \\ \delta \mid N, D_i &\sim \text{gamma}(n, N-n+R). \end{aligned} \tag{4.46}$$

Under $\pi_{0,1}$, we have

$$\begin{aligned} N - n \mid \delta, D_i &\sim \text{negbin}(n, 1 - e^{-\delta}), \\ \delta \mid N, D_i &\sim \text{gamma}(n+1, N-n+R). \end{aligned} \tag{4.47}$$

Since the posterior expectation of N does not exist under all four priors, we use the posterior median to estimate N. The estimates, as well as the standard deviations, are presented in Tables 4.1-4.4. In each table, the numbers outside the parenthesizes are the estimates while the numbers inside are standard deviations. We also examine the coverage probabilities of the one-sided credible intervals for N under the four priors with the nominal coverages ranging from 0.1 to 0.95. The results are plotted in Figures 4.1- 4.4.

The following conclusions can be drawn from the simulation results. First of all, for larger γ, the estimates of N are more accurate, and the actual coverage probabilities are closer to the nominal values. The explanation is straightforward. Since γ measures the proportion of failures detected prior to censoring, larger γ means more data, hence more accurate inferences. In fact, we observe in tables 4.1-4.4 that the true values of N are severely underestimated when $\gamma = 0.1, 0.25$. It is because the two parameters, N and δ, are confounding factors to each other. When γ is small, the data do not carry sufficient information to make precise inferences. For similar reasons, we obtain more precise estimates and coverage probabilities for large N. Secondly, the four priors perform similarly under all circumstances. However we observe that $\pi_{1,2}$ leads to slightly larger estimates, hence smaller biases than the other three, especially when $\gamma \leq 0.5$. For γ larger than 0.5, $\pi_{0,1}$ provides slightly better estimates. Moreover, whereas $\pi_{1,2}$ provides the closest coverage probabilities to nominal values for small γ among the four priors, it has the most inflated coverage probabilities for large γ. On the other hand, $\pi_{0,1}$ has the closest coverage probabilities for large γ, but the most underestimated coverage probabilities for small γ. The Jeffreys and reference priors π_1^J and π_1^R provide moderate estimates and coverage probabilities that are in between of the latter two.

4.4 Concluding Remarks

In a nutshell, we studied the survival model with exponential failure rate and type I censoring, and estimated the population size parameter N with objective Bayesian methods. We also compared four different objective priors. In practice, if the investigator can roughly estimate the proportion of failures occurred prior to censoring time from historical data or experience, he or she may refer to the simulation results and select the appropriate prior accordingly. If the proportion of failures occurred γ is unclear, one may estimate δ from its marginal likelihood function (4.5) and then calculate $\hat{\gamma}$. Alternatively, one can use the Jeffreys or reference priors because they are provide moderate and relatively stable estimates under various γ's compared to the other two priors.

Table 4.1: Estimates of N at $N = 50$. In each cell, the numbers outside and inside the parentheses are the estimate $\hat{N}$ and standard deviation of $\hat{N}$ respectively.

	$\gamma = 0.1$	$\gamma = 0.25$	$\gamma = 0.5$	$\gamma = 0.75$	$\gamma = 0.9$
$\pi_1^J(\delta, \lambda)$	9(6)	24(13)	51(22)	54(15)	51(5)
$\pi_1^R(\delta, \lambda)$	9(6)	24(13)	51(25)	54(13)	51(5)
$\pi_{1,2}(\delta, \lambda)$	9(6)	26(13)	55(27)	56(17)	51(5)
$\pi_{0,1}(\delta, \lambda)$	8(6)	23(12)	50(26)	54(16)	51(5)

Table 4.2: Estimates of N at $N = 100$. In each cell, the numbers outside and inside the parentheses are the estimate $\hat{N}$ and standard deviation of $\hat{N}$ respectively.

	$\gamma = 0.1$	$\gamma = 0.25$	$\gamma = 0.5$	$\gamma = 0.75$	$\gamma = 0.9$
$\pi_1^J(\delta, \lambda)$	21(11)	64(30)	113(49)	107(29)	102(7)
$\pi_1^R(\delta, \lambda)$	21(11)	63(30)	111(45)	106(24)	102(7)
$\pi_{1,2}(\delta, \lambda)$	22(11)	68(31)	119(52)	108(26)	102(8)
$\pi_{0,1}(\delta, \lambda)$	20(11)	62(31)	109(44)	105(24)	101(7)

Table 4.3: Estimates of N at $N = 200$. In each cell, the numbers outside and inside the parentheses are the estimate $\hat{N}$ and standard deviation of $\hat{N}$ respectively.

	$\gamma = 0.1$	$\gamma = 0.25$	$\gamma = 0.5$	$\gamma = 0.75$	$\gamma = 0.9$
$\pi_1^J(\delta, \lambda)$	53(27)	145(65)	238(119)	207(28)	202(10)
$\pi_1^R(\delta, \lambda)$	51(25)	148(71)	232(105)	207(29)	202(10)
$\pi_{1,2}(\delta, \lambda)$	57(29)	156(70)	247(130)	208(29)	202(10)
$\pi_{0,1}(\delta, \lambda)$	51(25)	145(79)	229(98)	206(28)	202(10)

Table 4.4: Estimates of N at $N = 500$. In each cell, the numbers outside and inside the parentheses are the estimate $\hat{N}$ and standard deviation of $\hat{N}$ respectively.

	$\gamma = 0.1$	$\gamma = 0.25$	$\gamma = 0.5$	$\gamma = 0.75$	$\gamma = 0.9$
$\pi_1^J(\delta, \lambda)$	195(102)	512(325)	566(194)	507(37)	502(14)
$\pi_1^R(\delta, \lambda)$	192(100)	556(648)	563(183)	507(37)	502(14)
$\pi_{1,2}(\delta, \lambda)$	208(114)	501(265)	581(211)	509(37)	502(14)
$\pi_{0,1}(\delta, \lambda)$	194(104)	490(300)	555(170)	507(36)	501(14)

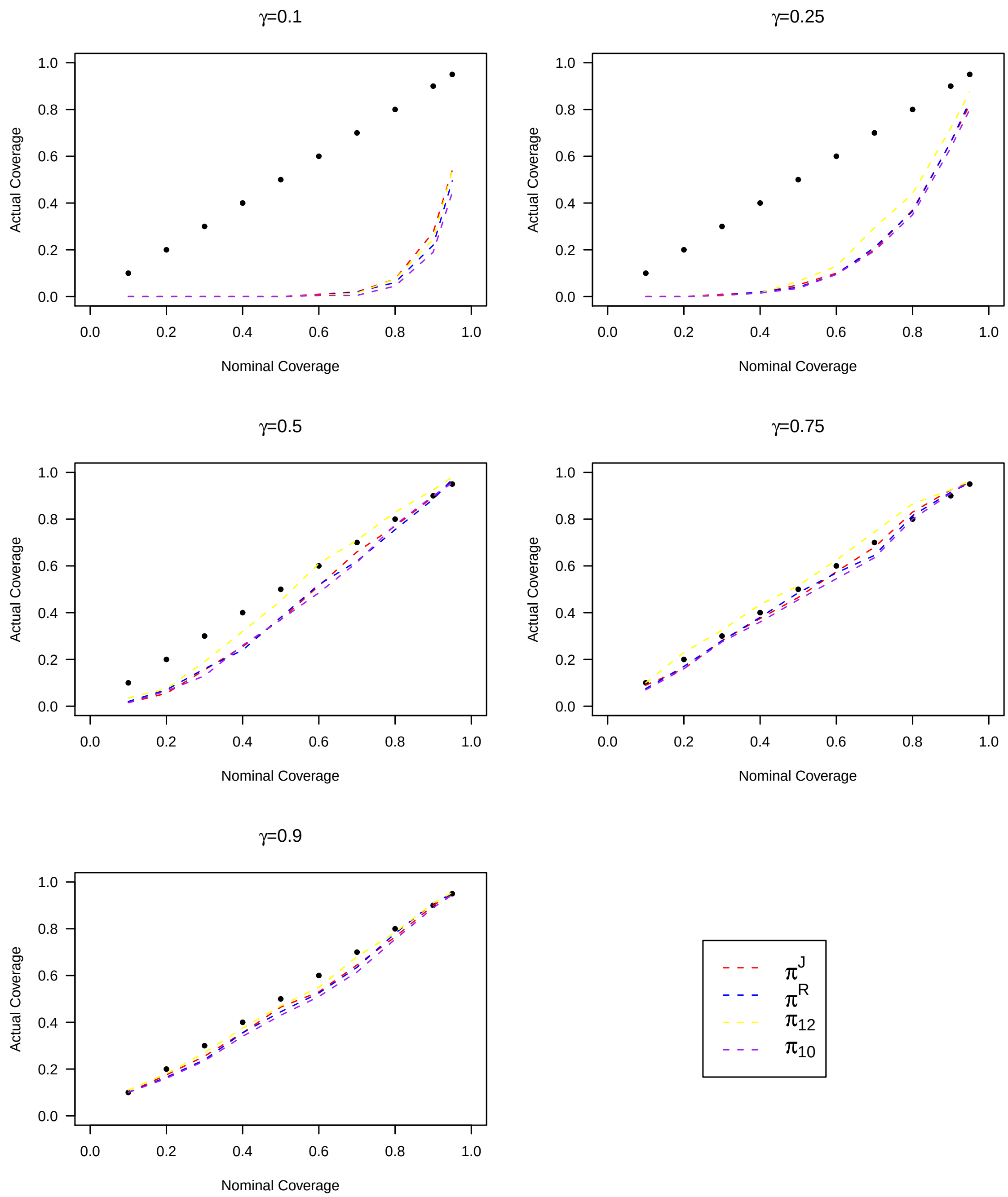

Figure 4.1: Coverage probabilities at $N = 50$ and $\gamma = 0.1, 0.25, 0.5, 0.75, 0.9$. The x axis represents the nominal coverages from 0.1 to 0.9 and 0.95, with 0.1 increment. The y axis represents the actual coverages.

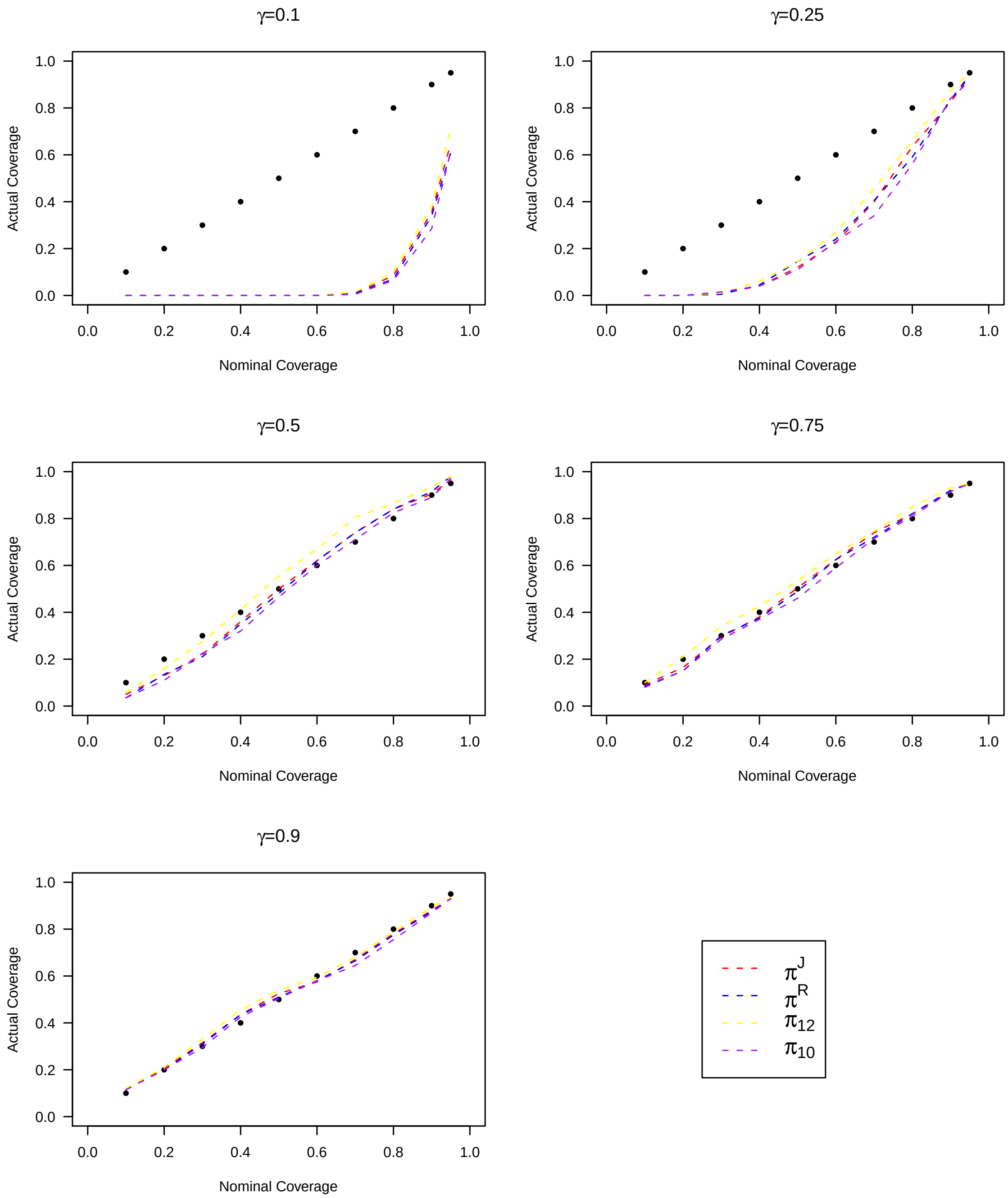

Figure 4.2: Coverage probabilities at $N = 100$ and $\gamma = 0.1, 0.25, 0.5, 0.75, 0.9$. The x axis represents the nominal coverages from 0.1 to 0.9 and 0.95, with 0.1 increment. The y axis represents the actual coverages.

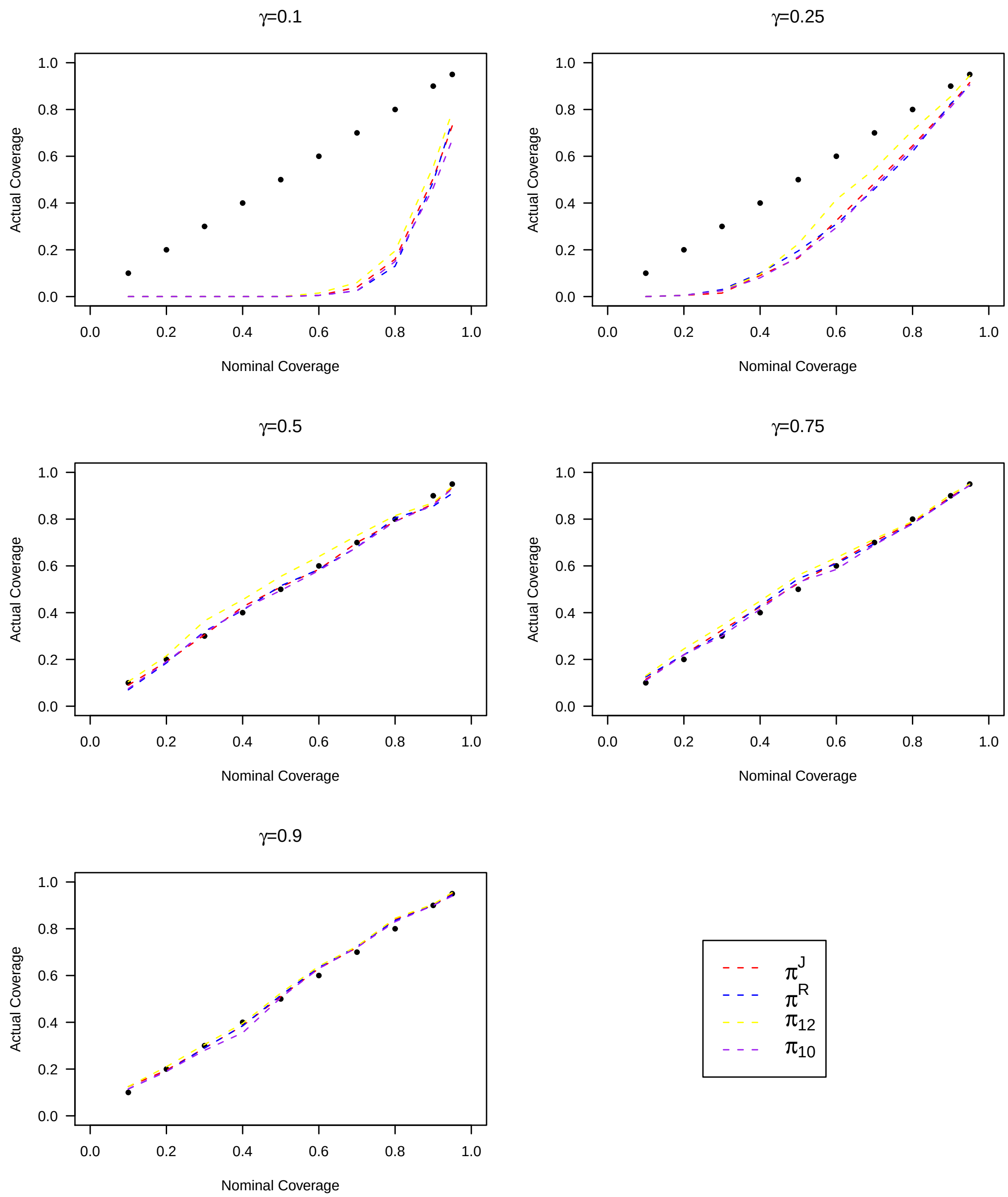

Figure 4.3: Coverage probabilities at $N = 200$ and $\gamma = 0.1, 0.25, 0.5, 0.75, 0.9$. The x axis represents the nominal coverages from 0.1 to 0.9 and 0.95, with 0.1 increment. The y axis represents the actual coverages.

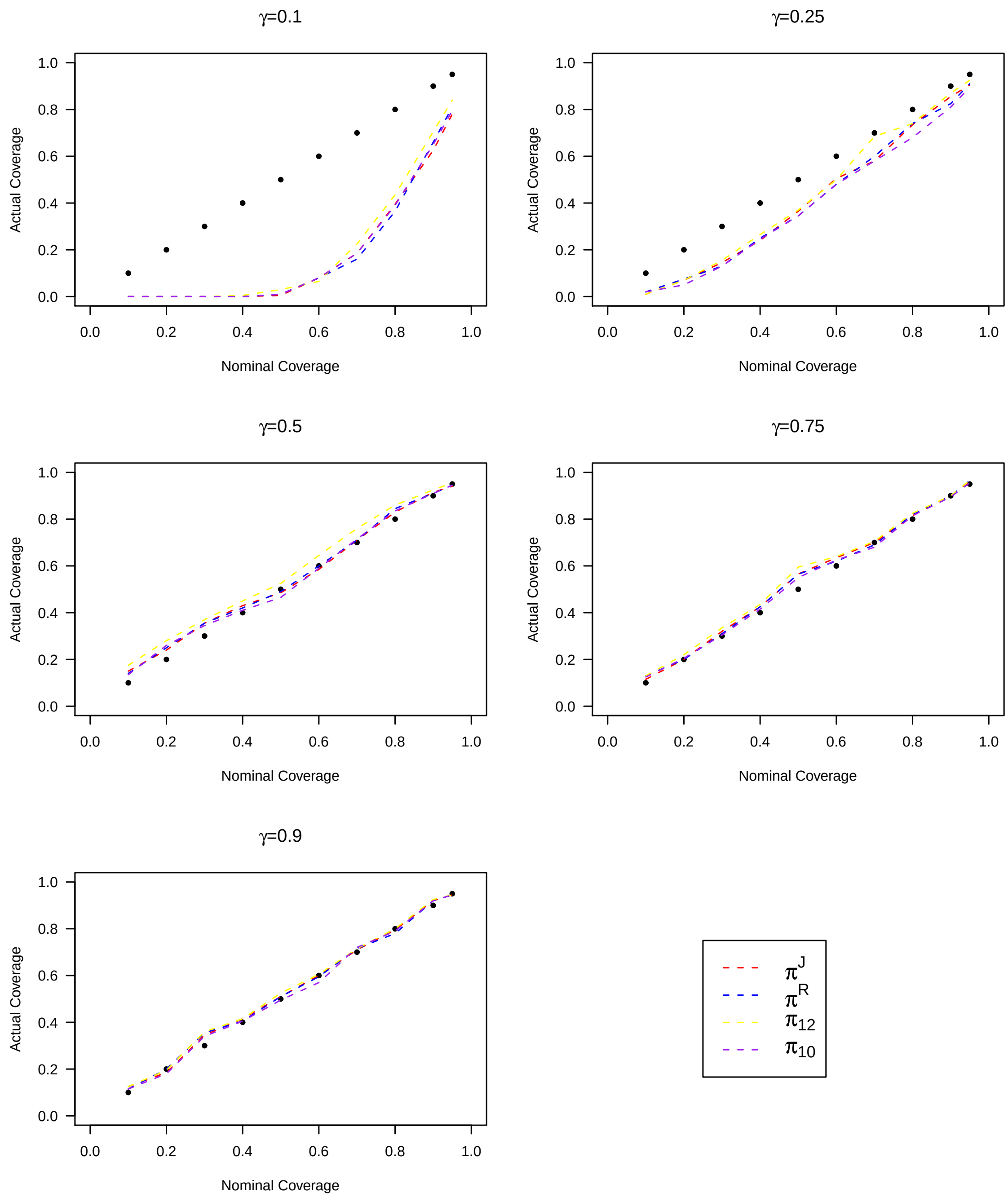

Figure 4.4: Coverage probabilities at $N = 500$ and $\gamma = 0.1, 0.25, 0.5, 0.75, 0.9$. The x axis represents the nominal coverages from 0.1 to 0.9 and 0.95, with 0.1 increment. The y axis represents the actual coverages.

Appendix A

Proof of Lemma 2.3.1

We start from proving (2.21). Since $\frac{1}{(N+1+i)^2}$ is a decreasing function of i, we can approximate the partial sum with integration and bound the error terms as the following.

$$0 \leq \sum_{i=0}^{\infty} \frac{k}{(N+a+b+i)^2} - \int_0^{\infty} \frac{k}{(N+a+b+x)^2} dx \leq \frac{k}{(N+a+b)^2}, \quad \text{(A.1)}$$

$$0 \leq \sum_{i=0}^{\infty} \frac{1}{(N+1+i)^2} - \int_0^{\infty} \frac{1}{(N+1+x)^2} dx \leq \frac{1}{(N+1)^2}. \quad \text{(A.2)}$$

Based on the facts that

$$\int_0^{\infty} \frac{k}{(N+a+b+x)^2} dx = \frac{k}{N+a+b}, \quad \text{(A.3)}$$

$$\int_0^{\infty} \frac{1}{(N+1+x)^2} dx = \frac{1}{N+1}, \quad \text{(A.4)}$$

we have

$$\begin{aligned}
\frac{k}{N+a+b} &\leq \sum_{i=0}^{\infty} \frac{k}{(N+a+b+i)^2} \leq \frac{k}{N+a+b} + \frac{k}{(N+a+b)^2}, && \text{(A.5)}\\
\frac{1}{N+1} &\leq \sum_{i=0}^{\infty} \frac{1}{(N+1+i)^2} \leq \frac{1}{N+1} + \frac{1}{(N+1)^2}. && \text{(A.6)}
\end{aligned}$$

We can bound $J_1(N)$ by

$$\frac{k}{N+a+b} - \frac{1}{N+1} - \frac{1}{(N+1)^2} \leq J_1(N) \leq \frac{k}{N+a+b} - \frac{1}{N+1} + \frac{k}{(N+a+b)^2}. \quad \text{(A.7)}$$

$J_1(N)$ does not depend on the data, so $\mathrm{E}(J_1(N)) = J_1(N)$. The inequality (A.7) yields (2.21) immediately.

To prove equation (2.22), we decompose $J_2(N)$ as the following,

$$J_2(N) = J_{21}(N) + J_{22}(N), \quad \text{(A.8)}$$

where

$$\begin{aligned}
J_{21}(N) &= \sum_{l=1}^{k} \frac{1}{N - m_l + 1},\\
J_{22}(N) &= \sum_{l=1}^{k} \sum_{i=0}^{\infty} \frac{1}{(N - m_l + a + b + i)^2} - \sum_{l=1}^{k} \frac{1}{N - m_l + 1}.
\end{aligned}$$

Interestingly, these two terms are of the same order as $N \to \infty$, which is presented in the following proposition.

Proposition A.0.1.

(a) *For fixed $k \geq 2$, $0 < a \leq 1$ and $0 < b < 1$, as $N \to \infty$,*

$$\begin{aligned}
E(J_{21}(N)) &\sim \frac{1}{N^b}, && \text{(A.9)}\\
E(J_{22}(N)) &\sim \frac{1}{N^b}. && \text{(A.10)}
\end{aligned}$$

(b) *If $b=1$, for fixed $k \geq 2$ and $0 < a \leq 1$, as $N \to \infty$,*

$$E(J_{21}(N)) \sim \frac{\log(N+a)}{N+1}, \tag{A.11}$$

$$\frac{\Gamma(a+b)}{b^2\Gamma(a)(N+a)^b} \leq E(J_{22}(N)) \leq \frac{C_1 \log(N+a)}{N+1}, \tag{A.12}$$

where C_1 is a fixed positive constant.

Proof. For all $1 \leq l \leq k$,

$$m_l \mid N, \theta_l \sim \mathrm{Bin}(N, 1-\theta_l).$$

It follows from Chao & Strawderman (1972) that

$$\begin{aligned}
\frac{1}{k}\mathrm{E}(J_{21}(N)) &= \mathrm{E}\left[\mathrm{E}\left(\frac{1}{N-m_1+1} \mid \theta_1\right)\right] \\
&= \int_0^1 \frac{1-\theta_1^{N+1}}{(N+1)(1-\theta_1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta_1^{a-1}(1-\theta_1)^{b-1} d\theta_1 \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)(N+1)} \sum_{j=0}^{N} \int_0^1 \theta_1^{j+a-1}(1-\theta_1)^{b-1} d\theta_1 \\
&= \frac{\Gamma(a+b)}{\Gamma(a)(N+1)} \sum_{j=0}^{N} \frac{\Gamma(j+a)}{\Gamma(j+a+b)}.
\end{aligned} \tag{A.13}$$

We have the decomposition

$$\frac{1}{k}\mathrm{E}\left(J_{21}(N)\right) = J_{211}(N) + J_{212}(N) + J_{213}(N), \tag{A.14}$$

where

$$
\begin{aligned}
J_{211}(N) &= \frac{\Gamma(a+b)}{\Gamma(a)(N+1)} \int_0^N \frac{1}{(x+a)^b} dx, && \text{(A.15)} \\
J_{212}(N) &= \frac{\Gamma(a+b)}{\Gamma(a)(N+1)} \left(\sum_{j=0}^{N} \frac{1}{(j+a)^b} - \int_0^N \frac{1}{(x+a)^b} dx \right), && \text{(A.16)} \\
J_{213}(N) &= \frac{\Gamma(a+b)}{\Gamma(a)(N+1)} \sum_{j=0}^{N} \left(\frac{\Gamma(j+a)}{\Gamma(j+a+b)} - \frac{1}{(j+a)^b} \right). && \text{(A.17)}
\end{aligned}
$$

Now we study the limiting behaviors of $J_{211}(N)$, $J_{212}(N)$, and $J_{213}(N)$,respectively. First of all, we can find the order of $J_{211}(N)$ by solving the integration directly. If $0 \leq b < 1$,

$$
J_{211}(N) = \frac{\Gamma(a+b)}{(1-b)\Gamma(a)(N+1)} \left((N+a)^{1-b} - a^{1-b} \right) \sim \frac{1}{N^b}. \tag{A.18}
$$

If $b = 1$,

$$
J_{211}(N) = \frac{\Gamma(a+b)}{(1-b)\Gamma(a)(N+1)} \log \frac{N+a}{a} \sim \frac{\log(N+a)}{N+1}. \tag{A.19}
$$

Next we show that $J_{212}(N)$ is asymptotically smaller than N^{-1}. By using the following

$$
\frac{1}{(1+a)^b} < \int_{j-1}^{j} \frac{1}{(x+a)^b} dx < \frac{1}{(j-1+a)^b}, \quad j = 1, \cdots, N, \tag{A.20}
$$

we get

$$
\sum_{j=1}^{N} \frac{1}{(j+a)^b} < \int_0^N \frac{1}{(x+a)^b} dx < \sum_{j=0}^{N-1} \frac{1}{(j+a)^b}, \tag{A.21}
$$

which is equivalent to

$$
\frac{1}{(N+a)^b} < \sum_{j=0}^{N} \frac{1}{(j+a)^b} - \int_0^N \frac{1}{(x+a)^b} dx < \frac{1}{a^b}. \tag{A.22}
$$

So we can bound $J_{212}(N)$ as

$$\frac{\Gamma(a+b)}{\Gamma(a)(N+1)(N+a)^b} < J_{212}(N) < \frac{\Gamma(a+b)}{\Gamma(a)a^b(N+1)} \tag{A.23}$$

Now we prove that $J_{213}(N)$ has the order of N^{-1}. It follows from Erdelyi & Tricomi (1951) that for fixed $\alpha, \beta > 0$, as $z \to \infty$,

$$\frac{\Gamma(z+\alpha)}{\Gamma(z+\beta)} = z^{\alpha-\beta}\left(1 + \frac{(\alpha-\beta)(\alpha+\beta-1)}{2z} + O(\frac{1}{z^2})\right). \tag{A.24}$$

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

$$\frac{\Gamma(j+a)}{\Gamma(j+a+b)} = \frac{1}{(j+a)^b}\left(1 + \frac{b(1-b)}{2(j+a)} + J^*_{213}(j)\right), \tag{A.25}$$

where $J^*_{213}(j) \sim O\left((j+a)^{-2}\right)$, as $j \to \infty$. Thus we can rewrite $J_{213}(N)$ as

$$J_{213}(N) = \frac{\Gamma(a+b)}{\Gamma(a)(N+1)}\left(\frac{b(1-b)}{2}\sum_{j=0}^{N}\frac{1}{(j+a)^{1+b}} + \sum_{j=0}^{N}\frac{J^*_{213}(j)}{(j+a)^b}\right).$$

The facts that

$$\sum_{j=0}^{\infty}\frac{1}{(j+a)^{1+b}} < \infty \quad \text{and} \quad \sum_{j=0}^{\infty}\frac{J^*_{213}(j)}{(j+a)^b} < \infty$$

ensure that as $N \to \infty$,

$$J_{213}(N) \sim \frac{1}{N}. \tag{A.26}$$

Applying (A.18), (A.19), (A.23), and (A.26) to (A.14) will prove (A.9) and (A.11).

To study $\mathrm{E}(J_{22}(N))$, we bound $J_{22}(N)$ from both sides. On one hand,

$$
\begin{aligned}
J_{22}(N) &\leq \sum_{l=1}^{k}\left[\frac{1}{(N-m_l+b)^2}+\int_0^\infty \frac{1}{(N-m_l+b+x)^2}dx\right]-\sum_{l=1}^{k}\frac{1}{N-m_l+1} \\
&= \sum_{l=1}^{k}\frac{1}{(N-m_l+b)^2}+\sum_{l=1}^{k}\left(\frac{1}{N-m_l+b}-\frac{1}{N-m_l+1}\right) \\
&= \sum_{l=1}^{k}\frac{1}{(N-m_l+b)^2}+\sum_{l=1}^{k}\frac{1-b}{(N-m_l+b)(N-m_l+1)} \\
&\leq \sum_{l=1}^{k}\frac{1}{(N-m_l+b)^2}+\sum_{l=1}^{k}\frac{1-b}{(N-m_l+b)^2} \\
&= \sum_{l=1}^{k}\frac{2-b}{(N-m_l+b)^2}. \qquad \text{(A.27)}
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
J_{22}(N) &\geq \sum_{l=1}^{k}\left(\int_0^\infty \frac{1}{(N-m_l+b+x)^2}dx-\frac{1}{N-m_l+1}\right) \\
&= \sum_{l=1}^{k}\frac{1-b}{(N-m_l+b)(N-m_l+1)} \geq \sum_{l=1}^{k}\frac{b(1-b)}{(N-m_l+b)^2}. \qquad \text{(A.28)}
\end{aligned}
$$

Let

$$
J_{221}(N)=\mathrm{E}\left(\frac{1}{(N-m_1+b)^2}\right). \qquad \text{(A.29)}
$$

(A.27) and (A.28) imply that

$$
b(1-b)J_{221}(N) \leq \mathrm{E}(J_{22}(N)) \leq (2-b)J_{221}(N). \qquad \text{(A.30)}
$$

Now we study the limit of $J_{221}(N)$. In fact,

$$
\begin{aligned}
J_{221}(N) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \frac{1}{(N-m_l+b)^2}\sum_{m_l=0}^{N}\binom{N}{m_l}\theta_l^{m_l+a-1}(1-\theta_l)^{N-m_l+b-1}d\theta_l \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\sum_{i=0}^{N}\frac{\Gamma(N+1)\Gamma(i+a)\Gamma(N-i+b)}{(N-i+b)^2\Gamma(i+1)\Gamma(N-i+1)\Gamma(N+a+b)}, \qquad \text{(A.31)}
\end{aligned}
$$

where m_l is replaced with i for simplicity. Since $J_{221}(N)$ is a finite summation of positive components, it is greater than its last term,

$$J_{221}(N) \geq \frac{\Gamma(a+b)\Gamma(N+a)}{b^2\Gamma(a)\Gamma(N+a+b)} \approx \frac{\Gamma(a+b)}{b^2\Gamma(a)(N+a)^b}. \tag{A.32}$$

Now we bound $J_{221}(N)$ from above. According to (A.9) and (A.11) , as $N \to \infty$,

$$J_{221}(N) \leq \frac{1}{b^2}\mathrm{E}\left(\frac{1}{(N-m_1+1)}\right) \sim \begin{cases} \frac{1}{N^b} & \text{if } 0 < b < 1, \\ \frac{\log(N+a)}{N+1} & \text{if } b = 1. \end{cases} \tag{A.33}$$

If follows from (A.32) and (A.33) that, as $N \to \infty$, if $0 < b < 1$,

$$J_{221}(N) \sim N^{-b}. \tag{A.34}$$

If $b = 1$,

$$\frac{\Gamma(a+b)}{b^2\Gamma(a)(N+a)^b} \leq J_{221}(N) \leq \frac{C_1 \log(N+a)}{N+1} \tag{A.35}$$

Combining (A.30), (A.34), and (A.35), we get (A.10) and (A.12). □

Applying Proposition A.0.1 to (A.8), we get (2.22).

Induction method is applied to study $\mathrm{E}(J_3(N))$. When $k = 2$ we have

$$n \mid N, \theta_1, \theta_2 \sim \mathrm{Bin}\left(N, 1-(1-\theta_1)(1-\theta_2)\right). \tag{A.36}$$

For $0 < b < 1$ and $b = 1$, the limit of $\mathrm{E}_2(J_3(N))$ will be studied separately. Let

$\delta_i = 1 - \theta_i$, $i = 1, 2$. For $0 < b < 1$, the expectation of $J_3(N)$ can be expressed as

$$
\begin{aligned}
\mathrm{E}_2(J_3(N)) = & \sum_{n=0}^{N} \left(\sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \right) \frac{\Gamma(N+1)}{\Gamma(n+1)\Gamma(N-n+1)} \\
& \int_0^1 \int_0^1 (1-\delta_1\delta_2)^n (\delta_1\delta_2)^{N-n} B^{-2}(a,b)(\delta_1\delta_2)^{b-1}(1-\delta_1)^{a-1}(1-\delta_2)^{a-1} d\delta_1 d\delta_2 \\
= & \sum_{n=0}^{N} \left(\sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \right) \frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(n+1)\Gamma(N-n+1)} \\
& \int_0^1 \int_0^1 (1-\delta_2+\delta_2(1-\delta_1))^n (\delta_2\delta_2)^{N-n+b-1}(1-\delta_1)^{a-1}(1-\delta_2)^{a-1} d\delta_1 d\delta_2 \\
= & \sum_{n=0}^{N} \left(\sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \right) \frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(n+1)\Gamma(N-n+1)} \sum_{j=0}^{n} \frac{\Gamma(n+1)}{\Gamma(j+1)\Gamma(n-j+1)} \\
& \int_0^1 \int_0^1 \delta_1^{N-n+b-1}(1-\delta_1)^{n-j+a-1}\delta_2^{N-j+b-1}(1-\delta_2)^{j+a-1} d\delta_1 d\delta_2 \\
= & \frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=0}^{N} \left(\sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \right) \frac{\Gamma(N-n+b)}{\Gamma(N-n+1)} J_3^*(n), \qquad \text{(A.37)}
\end{aligned}
$$

where

$$
J_3^*(n) = \sum_{j=0}^{n} \frac{\Gamma(j+a)\Gamma(n-j+a)\Gamma(N-j+b)}{\Gamma(j+1)\Gamma(n-j+1)\Gamma(N-j+a+b)}. \qquad \text{(A.38)}
$$

We can decompose $J_3^*(n)$ by defining

$$
\begin{aligned}
J_{31}^*(n) &= \int_0^n \frac{1}{(x+1)^{1-a}(n-x+1)^{1-a}(N-x+1)^a} dx, && \text{(A.39)} \\
J_{32}^*(n) &= \sum_{j=0}^{n} \frac{1}{(j+1)^{1-a}(n-j+1)^{1-a}(N-j+1)^a} - J_{31}^*(n), && \text{(A.40)} \\
J_{33}^*(n) &= J^*(n) - J_{32}^*(n). && \text{(A.41)}
\end{aligned}
$$

Then we have

$$
J_3^*(n) = J_{31}^*(n) + J_{32}^*(n) + J_{33}^*(n) \qquad \text{(A.42)}
$$

and consequently

$$
\mathrm{E}_2(J_3(N)) = J_{31}(N) + J_{32}(N) + J_{33}(N), \qquad \text{(A.43)}
$$

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

$$J_{3l}(N) = B^{-2}(a,b)\frac{\Gamma(N+1)}{\Gamma(N+a+b)}\sum_{n=0}^{N}\left(\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\right)\frac{\Gamma(N-n+b)}{\Gamma(N-n+1)}J_{3l}^{*}(n). \tag{A.44}$$

Proposition A.0.2. *For $k = 2$ and $0 < a \le 1$, as $N \to \infty$,*

(1) *for $0 < b < 1$, there exists positive constant C_2 such that*

$$\frac{C_2 B^{-2}(a,b)\log(N+1)}{N^b} \le J_{31}(N) \le \frac{C_2 B^{-2}(a,b)\log(N+1)}{N^b} + \frac{(\log N)^b}{N^b}. \tag{A.45}$$

(2) *For $0 < b \le 1$,*

$$J_{32}(N) \le \frac{1}{N^b}. \tag{A.46}$$

(3) *For $0 < b < 1$,*

$$J_{33}(N) \le \frac{1}{N^b}. \tag{A.47}$$

For $b = 1$,

$$J_{33}(N) \le \frac{\log N}{N}. \tag{A.48}$$

Proof. In order to prove (A.45), we decompose $J_{31}(N)$ into two parts. Define

$$\begin{aligned} J_{311}(N) &= \frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)}\sum_{n=0}^{N-\log N}\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\frac{\Gamma(N-n+b)}{\Gamma(N-n+1)}J_{31}^{*}(n), \\ J_{312}(N) &= \frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)}\sum_{n=N-\log N}^{N}\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\frac{\Gamma(N-n+b)}{\Gamma(N-n+1)}J_{31}^{*}(n). \end{aligned} \tag{A.49}$$

Then we have

$$J_{31}(N) = J_{311}(N) + J_{312}(N). \tag{A.50}$$

Fact A.0.1. *For $0 < b < 1$, as $N \to \infty$,*

$$J_{311}(N) \leq \frac{(\log N)^b}{N^b}, \tag{A.51}$$

$$J_{312}(N) \sim \frac{\log N}{N^b}. \tag{A.52}$$

Proof. Let $y = x/n$ and we can bound $J_{31}^*(n)$ by

$$\begin{aligned} J_{31}^*(n) &= \frac{1}{n^{1-a}} \int_0^1 \frac{1}{(y+\frac{1}{n})^{1-a}(1-y+\frac{1}{n})^{1-a}(1-y+\frac{1}{n}+\frac{N-n}{n})^a} dy \\ &\leq \frac{1}{n^{1-a}} \int_0^1 \frac{1}{(y+\frac{1}{n})^{1-a}(1-y+\frac{1}{n})} dy \equiv J_{311}^*(n) + J_{312}^*(n), \end{aligned} \tag{A.53}$$

where

$$J_{311}^*(n) = \frac{\log(n+1)}{(n+1)^{1-a}}, \tag{A.54}$$

$$J_{312}^*(n) = \frac{1}{n^{1-a}} \int_0^1 \frac{1}{(y+\frac{1}{n})^{1-a}(1-y+\frac{1}{n})} dy - J_{311}^*(n). \tag{A.55}$$

Let

$$J_{3111}(N) = \frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=0}^{N-\log N} \sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \frac{\Gamma(N-n+b)}{\Gamma(N-n+1)} J_{311}^*(n),$$

$$J_{3112}(N) = \frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=0}^{N-\log N} \sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \frac{\Gamma(N-n+b)}{\Gamma(N-n+1)} J_{312}^*(n).$$

We study $J_{3111}(N)$ and $J_{3112}(N)$ separately. First we can bound $J_{3111}(N)$ from above by

$$\begin{aligned} J_{3111}(N) &\leq \frac{2B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=0}^{N-\log N} \frac{1}{N-n+1} \frac{\Gamma(N-n+b)}{\Gamma(N-n+1)} \frac{\log(n+1)}{(n+1)^{1-a}} \\ &= \frac{2B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=0}^{N-\log N} \frac{\Gamma(N-n+b)}{\Gamma(N-n+2)} \frac{\log(n+1)}{(n+1)^{1-a}} \\ &= J_{31111}(N) + J_{31112}(N) + J_{31113}(N), \end{aligned} \tag{A.56}$$

where

$$
\begin{aligned}
J_{31111}(N) &= 2B^{-2}(a,b)\frac{\Gamma(N+1)}{\Gamma(N+a+b)}\int_0^{N-\log N}\frac{\log(x+1)}{(x+1)^{1-a}(N-x+1)^{2-b}}dx,\\
J_{31112}(N) &= 2B^{-2}(a,b)\frac{\Gamma(N+1)}{\Gamma(N+a+b)}\left[\sum_{n=0}^{N-\log N}\frac{\log(n+1)}{(n+1)^{1-a}(N-n+1)^{2-b}}\right.\\
&\quad \left. -\int_0^{N-\log N}\frac{\log(x+1)}{(x+1)^{1-a}(N-x+1)^{2-b}}dx\right],\\
J_{31113}(N) &= 2B^{-2}(a,b)\frac{\Gamma(N+1)}{\Gamma(N+a+b)}\sum_{n=0}^{N-\log N}\left[\frac{\Gamma(N-n+b)}{\Gamma(N-n+2)}\frac{\log(n+1)}{(n+1)^{1-a}}\right.\\
&\quad \left. -\frac{\log(n+1)}{(n+1)^{1-a}(N-n+1)^{2-b}}\right].
\end{aligned}
\tag{A.57}
$$

Now we show that the order of $J_{31111}(N)$ cannot exceed $(\log N)^b/N^b$.

$$
\begin{aligned}
J_{31111}(N) &\le \frac{2B^{-2}(a,b)\Gamma(N+1)\log(N-\log N)}{\Gamma(N+a+b)}\int_0^{N-\log N}\frac{1}{(x+1)^{1-a}(N-x+1)^{2-b}}dx\\
&\le \frac{2B^{-2}(a,b)\Gamma(N+1)\log(N-\log N)}{(N-\log N)^{2-a-b}\Gamma(N+a+b)}\int_0^1\frac{1}{(y+\frac{1}{N-\log N})^{1-a}(\frac{N+1}{N-\log N}-y)^{2-b}}dy.
\end{aligned}
\tag{A.58}
$$

To study the integral, first we observe that

$$
\begin{aligned}
&\int_0^{1/2}\frac{1}{(y+\frac{1}{N-\log N})^{1-a}(\frac{N+1}{N-\log N}-y)^{2-b}}dy\\
\le\ & \frac{1}{\left(\frac{N+1}{N-\log N}-\frac{1}{2}\right)^{2-b}}\int_0^{1/2}\frac{1}{(y+\frac{1}{N-\log N})^{1-a}}dy\\
=\ & \frac{1}{a\left(\frac{N+1}{N-\log N}-\frac{1}{2}\right)^{2-b}}\left[\left(\frac{1}{2}+\frac{1}{N-\log N}\right)^a-\left(\frac{1}{N-\log N}\right)^a\right]\\
\sim\ & \frac{2^{2-a-b}}{a}.
\end{aligned}
\tag{A.59}
$$

On the other hand, since $0 < b < 1$, we have

$$\begin{aligned}
& \int_{1/2}^{1} \frac{1}{(y+\frac{1}{N-\log N})^{1-a}(\frac{N+1}{N-\log N}-y)^{2-b}}dy \\
\leq\ & \frac{1}{\left(\frac{1}{N-\log N}+\frac{1}{2}\right)^{1-a}} \int_{1/2}^{1} \frac{1}{\left(\frac{N+1}{N-\log N}-y\right)^{2-b}}dy \\
\leq\ & \frac{2^{1-a}}{1-b}\left[\left(\frac{1+\log N}{N-\log N}\right)^{b-1} - \left(\frac{N+1}{N-\log N}-\frac{1}{2}\right)^{b-1}\right] \\
\sim\ & \frac{2^{1-a}N^{1-b}}{(1-b)(\log N)^{1-b}}. && \text{(A.60)}
\end{aligned}$$

It follows from (A.59) and (A.60) that as $N \to \infty$,

$$\begin{aligned}
J_{31111}(N) \ \leq\ & \frac{2B^{-2}(a,b)\Gamma(N+1)\log(N-\log N)}{(N-\log N)^{2-a-b}\Gamma(N+a+b)} \frac{2^{1-a}N^{1-b}}{(1-b)(\log N)^{1-b}} \\
\sim\ & \frac{2^{2-a}B^{-2}(a,b)\log^b N}{(1-b)N^b}. && \text{(A.61)}
\end{aligned}$$

For $J_{31112}(N)$, define

$$f(x) = \frac{\log(x+1)}{(x+1)^{1-a}(N-x+1)^{2-b}}. \tag{A.62}$$

Since $f(x)$ is an increasing function, we have

$$\begin{aligned}
J_{31112}(N) =\ & \frac{2B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)}\left[\sum_{n=0}^{N-\log N} f(n) - \int_0^{N-\log N} f(x)dx\right] \\
\leq\ & \frac{2B^{-2}(a,b)f(N-\log N)\Gamma(N+1)}{\Gamma(N+a+b)} = \frac{2B^{-2}(a,b)\Gamma(N+1)\log(N+1)}{(N+1)^{1-a}\Gamma(N+a+b)(\log N+1)^{2-b}} \\
\approx\ & \frac{2B^{-2}(a,b)}{N^b\log^{1-b}N}. && \text{(A.63)}
\end{aligned}$$

Applying the results of Erdelyi & Tricomi (1951), we have

$$\begin{aligned}
& J_{31113}(N) \\
\leq\ & \frac{(1-b)(2-b)B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=0}^{N-\log N} \left[-\frac{\log(n+1)}{(n+1)^{1-a}(N-n+1)^{3-b}} \right] \\
\leq\ & \frac{(1-b)(2-b)}{2} (J_{31111}(N) + J_{31112}(N)) \\
\leq\ & \frac{2^{2-a}B^{-2}(a,b)\log^b N}{(1-b)N^b} + \frac{2B^{-2}(a,b)}{N^b \log^{1-b} N}.
\end{aligned} \tag{A.64}$$

Combining (A.56), (A.61), (A.63), and (A.64), we get

$$J_{3111}(N) \leq \frac{2^{3-a}B^{-2}(a,b)\log^b N}{(1-b)N^b}, \tag{A.65}$$

Next we show that the order of $J_{3112}(N)$ will not exceed N^{-b}. Notice that

$$\begin{aligned}
0 \leq & \int_0^{1/2} \frac{1}{(y+1/n)^{1-a}(1-y+1/n)} dy \\
\leq\ & \frac{1}{1/2+1/n} \int_0^{1/2} \frac{1}{(y+1/n)^{1-a}} dy \leq \frac{2}{a}\left(\frac{1}{2}+\frac{1}{n}\right)^a \\
\sim\ & \frac{2^{1-a}}{a}.
\end{aligned} \tag{A.66}$$

On the other hand,

$$\begin{aligned}
\int_{1/2}^1 \frac{1}{(y+1/n)^{1-a}(1-y+1/n)} dy \geq & \frac{1}{(1+1/n)^{1-a}} \int_{1/2}^1 \frac{1}{1-y+1/n} dy \\
= & \frac{1}{(1+1/n)^{1-a}} \log(1+n/2) \sim \log(1+n/2).
\end{aligned} \tag{A.67}$$

Similarly, we have

$$\int_{1/2}^1 \frac{1}{(y+1/n)^{1-a}(1-y+1/n)} dy \leq 2^{1-a}\log(1+n/2). \tag{A.68}$$

It follows from (A.66), (A.67), and (A.68) that

$$-\log 2 \le \int_0^1 \frac{1}{(y+1/n)^{1-a}(1-y+1/n)}dy - \log n \le 2^{1-a}(1-\log 2)/a. \qquad \text{(A.69)}$$

Consequently, we let $C_3 = \min\{\log 2, 2^{1-a}(1-\log 2)/a\}$ and bound $J^*_{312}(n)$ by

$$|J^*_{312}(n)| \le \frac{C_3}{(n+1)^{1-a}}. \qquad \text{(A.70)}$$

By the similar approach we used to approximate the ratio of gamma functions, we can bound $J_{3112}(N)$ as the following:

$$\begin{aligned}
|J_{3112}(N)| &\le C_3 B^{-2}(a,b)\frac{\Gamma(N+1)}{\Gamma(N+a+b)}\sum_{n=0}^{N}\frac{\Gamma(N-n+b)}{(n+1)^{1-a}\Gamma(N-n+2)} \\
&\approx C_3 B^{-2}(a,b)\frac{\Gamma(N+1)}{\Gamma(N+a+b)}\int_0^N \frac{1}{(x+1)^{1-a}(N-x+1)^{2-b}}dx \\
&= \frac{C_3 B^{-2}(a,b)\Gamma(N+1)}{N^{2-a-b}\Gamma(N+a+b)}\int_0^1 \frac{1}{(y+1/N)^{1-a}(1-y+1/N)^{2-b}}dx \\
&\sim \frac{C_3 B^{-2}(a,b)}{N^b}. \qquad \text{(A.71)}
\end{aligned}$$

It follows from (A.53), (A.65), and (A.71) that (A.51) holds.

Now we find the order of $J_{312}(N)$. First we can bound $J_{312}(N)$ from above. Since it has already been shown that $J^*_{31}(n) \le \frac{\log(n+1)}{(n+1)^{1-a}}$, it is easy to see that

$$\begin{aligned}
&J_{312}(N) \\
\le\ & \frac{B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)}\sum_{n=N-\log N}^{N}\left(\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\right)\frac{\log(n+1)\Gamma(N-n+b)}{(n+a)^{1-a}\Gamma(N-n+1)} \\
\le\ & \frac{B^{-2}(a,b)\log(N+1)\Gamma(N+1)}{(N-\log N+1)^{1-a}\Gamma(N+a+b)}H(N), \qquad \text{(A.72)}
\end{aligned}$$

where

$$H(N) = \sum_{n=N-\log N}^{N}\left(\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\right)\frac{\Gamma(N-n+b)}{\Gamma(N-n+1)}. \qquad \text{(A.73)}$$

We can also find a lower bound of $J_{312}(N)$. Notice that

$$
\begin{aligned}
J_{31}^*(N) \geq \frac{1}{n^{1-a}} \int_0^1 & \frac{1}{\left(y+\frac{1}{n}+\frac{N-n}{n}\right)^{1-a}\left(1-y+\frac{1}{n}+\frac{N-n}{n}\right)} dy \\
\geq & \frac{1}{n^{1-a}\left(1+\frac{1}{n}+\frac{N-n}{n}\right)^{1-a}} \int_0^1 \frac{1}{\left(1-y+\frac{1}{n}+\frac{N-n}{n}\right)} dy \\
= & \frac{1}{(N+1)^{1-a}} (\log n - \log(N-n+1)) \\
\geq & \frac{1}{(N+1)^{1-a}} (\log(N-\log N) - \log(1+\log N)). \qquad \text{(A.74)}
\end{aligned}
$$

Consequently, we have

$$
J_{312}(N) \geq B^{-2}(a,b) \frac{[\log(N-\log N) - \log(1+\log N)]\Gamma(N+1)}{(N+1)^{1-a}\Gamma(N+a+b)} H(N). \qquad \text{(A.75)}
$$

Let $m = N - n$ and we have

$$
\begin{aligned}
H(N) &= \sum_{m=0}^{\log N} \left(\sum_{i=0}^{\infty} \frac{1}{(m+1+i)^2} \right) \frac{\Gamma(m+b)}{\Gamma(m+1)} \\
&\leq \sum_{m=0}^{\log N} \frac{2\Gamma(m+b)}{(m+1)\Gamma(m+1)} \leq \sum_{m=0}^{\infty} \frac{2\Gamma(m+b)}{\Gamma(m+2)} \sim \sum_{m=0}^{\infty} \frac{1}{m^{2-b}} < \infty. \qquad \text{(A.76)}
\end{aligned}
$$

Since $H(N)$ is an increasing function of N, the limit of $H(N)$ as $N \to \infty$ exists and we assume

$$
\lim_{N \to \infty} H(N) = C_4, \qquad \text{(A.77)}
$$

where C_4 is a positive constant. Combining (A.72) and (A.75), we get

$$
\begin{aligned}
C_4 B^{-2}(a,b) & \frac{[\log(N-\log N) - \log(1+\log N)]\Gamma(N+1)}{(N+1)^{1-a}\Gamma(N+a+b)} \leq J_{312}(N) \\
\leq & \; C_4 B^{-2}(a,b) \frac{\log(N+1)\Gamma(N+1)}{(N-\log N+1)^{1-a}\Gamma(N+a+b)}, \qquad \text{(A.78)}
\end{aligned}
$$

which yields (A.52) immediately. $\square$

Fact A.0.1 leads to (A.45) immediately.

Next we prove (A.46). For $0 \le x \le n$, define $g(x)$ as

$$g(x) = \frac{1}{(x+1)^{1-a}(n-x+1)^{1-a}(N-x+1)^{a}}. \tag{A.79}$$

Then $J_{32}^*(N)$ can be expressed as

$$J_{32}^*(N) = \sum_{j=0}^{n} g(j) - \int_0^n g(x)dx. \tag{A.80}$$

We take derivative to $\log g(x)$ and get

$$\frac{d}{dx}\log g(x) = \frac{-(2-a)x^2 + [2(1-a)(N+1)+n]x + a - 1 + 1/(n+1)}{(x+1)(n-x+1)(N-x+1)}. \tag{A.81}$$

For sufficiently large n, we can show that

$$\begin{aligned} \frac{d}{dx}\log g(0) &< 0, \\ \frac{d}{dx}\log g(n) &> 0. \end{aligned} \tag{A.82}$$

There exists $0 < j_0 < n$, such that $g(x)$ decreases when $0 \le x \le j_0$ and increases when $j_0 + 1 \le x \le n$. Consequently we have

$$0 \le \sum_{j=0}^{j_0} g(j) - \int_0^{j_0} g(x)dx \le g(0) = \frac{1}{(n+1)^{1-a}(N+1)^{a}}, \tag{A.83}$$

$$0 \le \sum_{j=j_0+1}^{n} g(j) - \int_{j_0}^{n} g(x)dx \le g(n) = \frac{1}{(n+1)^{1-a}(N-n+1)^{a}}. \tag{A.84}$$

Combining the inequalities above, we have

$$0 \le J_{32}^*(N) \le \frac{1}{(n+1)^{1-a}}\left[\frac{1}{(N+1)^{a}} + \frac{1}{(N-n+1)^{a}}\right]. \tag{A.85}$$

It follows from (A.44) that

$$
\begin{aligned}
J_{32}(N) \leq B^{-2}(a,b)\frac{\Gamma(N+1)}{\Gamma(N+a+b)}\sum_{n=0}^{N}\left(\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\right)\frac{\Gamma(N-n+b)}{\Gamma(N-n+1)} \\
\frac{1}{(n+1)^{1-a}}\left[\frac{1}{(N+1)^a}+\frac{1}{(N-n+1)^a}\right] \\
\leq \quad B^{-2}(a,b)\frac{\Gamma(N+1)}{\Gamma(N+a+b)}\sum_{n=0}^{N}\frac{2}{N-n+1}\frac{\Gamma(N-n+b)}{\Gamma(N-n+1)}\frac{2}{(n+1)^{1-a}(N-n+1)} \\
= \quad 4B^{-2}(a,b)\frac{\Gamma(N+1)}{\Gamma(N+a+b)}\sum_{n=0}^{N}\frac{\Gamma(N-n+b)}{(n+a)^{1-a}(N-n+1)^a\Gamma(N-n+2)}. \qquad \text{(A.86)}
\end{aligned}
$$

We use a similar method to approximate the ratio of gamma functions and get

$$
\begin{aligned}
J_{32}(N) \quad &\leq \quad \frac{4B^{-2}(a,b)\Gamma(N+1)}{N^{2-b}\Gamma(N+a+b)}\int_0^1 \frac{1}{(y+1/N)^{1-a}(1-y+1/N)^{2+a-b}}dy \\
&\sim \quad \frac{4B^{-2}(a,b)N^{1+a-b}\Gamma(N+1)}{(1+a-b)N^{2-b}\Gamma(N+a+b)} \sim N^{-b}. \qquad \text{(A.87)}
\end{aligned}
$$

Now we prove (A.47) and (A.48) and complete the proof of the proposition. It follows from Erdelyi & Tricomi (1951) that

$$
\begin{aligned}
\frac{\Gamma(j+a)}{\Gamma(j+1)} &= \frac{1}{(j+1)^{1-a}}\left[1+\frac{a(1-a)}{2(j+a)}+T_1(j)\right], \qquad \text{(A.88)} \\
\frac{\Gamma(n-j+a)}{\Gamma(n-j+1)} &= \frac{1}{(n-j+1)^{1-a}}\left[1+\frac{a(1-a)}{2(n-j+a)}+T_2(j)\right], \qquad \text{(A.89)} \\
\frac{\Gamma(N-j+a)}{\Gamma(N-j+1)} &= \frac{1}{(N-j+1)^{1-a}}\left[1+\frac{a(1+a-2b)}{2(N-j+a)}+T_3(j)\right], \qquad \text{(A.90)}
\end{aligned}
$$

where

$$
\begin{aligned}
T_1(j) &= O\left(\frac{1}{(j+1)^2}\right), \quad j\to\infty. \\
T_2(j) &= O\left(\frac{1}{(n-j+1)^2}\right), \quad n-j\to\infty. \\
T_3(j) &= O\left(\frac{1}{(N-j+1)^2}\right), \quad N-j\to\infty.
\end{aligned}
$$

We can rewrite $J^*_{33}(n)$ as

$$J^*_{33}(n) = \sum_{j=0}^{n} \frac{1}{(j+1)^{1-a}(n-j+1)^{1-a}(N-j+1)^a} \left[1 + \frac{a(1-a)}{2(j+a)} + T_1(j)\right]$$
$$\left[1 + \frac{a(1-a)}{2(n-j+a)} + T_2(j)\right] \left[1 + \frac{a(1+a-2b)}{2(N-j+a)} + T_3(j)\right]. \qquad \text{(A.91)}$$

The leading terms of $J^*_{33}(n)$ are

$$J^*_{331}(n) = \frac{a(1-a)}{2(j+1)^{2-a}(n-j+1)^{1-a}(N-j+1)^a}, \qquad \text{(A.92)}$$
$$J^*_{332}(n) = \frac{a(1-a)}{2(j+1)^{1-a}(n-j+1)^{2-a}(N-j+1)^a}, \qquad \text{(A.93)}$$
$$J^*_{333}(n) = \frac{a(1+a-2b)}{2(j+1)^{1-a}(n-j+1)^{1-a}(N-j+1)^{1+a}}. \qquad \text{(A.94)}$$

We now show that all these terms are asymptotically smaller than $1/n^{1-a}$. Firstly,

$$\begin{aligned} J^*_{331}(n) &\leq \sum_{j=0}^{n} \frac{a(1-a)}{2(j+1)^{2-a}(n-j+1)} \\ &\approx \frac{a(1-a)}{2} \int_0^n \frac{1}{(x+1)^{2-a}(n-x+1)} dx \\ &= \frac{a(1-a)}{2n^{2-a}} \int_0^1 \frac{1}{(y+1/n)^{2-a}(1-y+1/n)} dy \\ &\sim 1/n. \end{aligned} \qquad \text{(A.95)}$$

Secondly,

$$\begin{aligned} J^*_{332}(n) &\leq \sum_{j=0}^{n} \frac{a(1-a)}{2(j+1)^{1-a}(n-j+1)^2} \\ &\approx \frac{a(1-a)}{2} \int_0^n \frac{1}{(x+1)^{1-a}(n-x+1)^2} dx \\ &= \frac{a(1-a)}{2n^{2-a}} \int_0^1 \frac{1}{(y+1/n)^{1-a}(1-y+1/n)^2} dy \\ &\sim 1/n^{1-a}. \end{aligned} \qquad \text{(A.96)}$$

Finally,

$$J_{333}^*(n) \le J_{332}^*(n) \sim 1/n^{1-a}. \tag{A.97}$$

It follows from (A.95), (A.96), and (A.97) that as $n \to \infty$,

$$J_{33}^*(n) \le 1/n^{1-a}. \tag{A.98}$$

Therefore we have

$$\begin{aligned} J_{33}(N) &\le \sum_{n=0}^{N} \left(\sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \right) \frac{B^{-2}(a,b)\Gamma(N+1)\Gamma(N-n+b)}{(n+1)^{1-a}\Gamma(N+a+b)\Gamma(N-n+1)} \\ &\le \frac{2B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=0}^{N} \frac{\Gamma(N-n+b)}{(n+1)^{1-a}\Gamma(N-n+2)}. \end{aligned} \tag{A.99}$$

The right hand side of (A.99) can be approximated by

$$\begin{aligned} &\text{The RHS of } (A.99) \\ \approx\ & \frac{2B^{-2}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \int_0^N \frac{1}{(x+1)^{1-a}(N-x+1)^{2-b}} dx \\ =\ & \frac{2B^{-2}(a,b)\Gamma(N+1)}{N^{2-a-b}\Gamma(N+a+b)} \int_0^1 \frac{1}{(y+1/N)^{1-a}(1-y+1/N)^{2-b}} dy \\ \sim\ & \begin{cases} \frac{1}{N^b}, & \text{if } 0 < b < 1; \\ \frac{\log N}{N}, & \text{if } b = 1. \end{cases} \end{aligned} \tag{A.100}$$

The proof is completed. $\square$

For $b = 1$, we use a different approach to study $\mathrm{E}_2(J_3(N))$. Notice $J_3(N)$ can be bounded by :

$$\frac{1}{N-n+1} \le J_3(N) \le \frac{1}{N-n+1} + \frac{1}{(N-n+1)^2}. \tag{A.101}$$

Proposition A.0.3. *For $k = 2$, $b = 1$, and fixed $0 < a \le 1$, sending N to infinity,* *we have*

$$
\begin{aligned}
E\left(\frac{1}{N-n+1}\right) &\geq \frac{B^{-2}(a,1)\log^2 N}{2(N+1)} && \text{(A.102)}\\
E\left(\frac{1}{N-n+1}\right) &\leq \frac{B^{-2}(a,1)\log^2 N}{2(N+1)} + \frac{B^{-2}(a,1)(\log N)^{3/2}}{a(N+1)}, && \text{(A.103)}\\
E\left(\frac{1}{(N-n+1)^2}\right) &\leq \frac{2B^{-2}(a,1)\log(N+1)}{N}. && \text{(A.104)}
\end{aligned}
$$

Proof. We first prove (A.102) and (A.103). For $l = 4, 5, 6$, define

$$
J_{3l}(N) = \frac{B^{-2}(a,1)}{N+1}\sum_{j=0}^{N}\frac{\Gamma(j+1)}{\Gamma(j+a+1)}J^*_{3lj}, \tag{A.105}
$$

and

$$
\begin{aligned}
J^*_{34j} &= \int_0^j \frac{1}{(x+a)^{1-a}(j-x+a)}dx,\\
J^*_{35j} &= \left[\sum_{i=0}^{j}\frac{1}{(i+a)^{1-a}(j-i+a)} - \int_0^j \frac{1}{(x+a)^{1-a}(j-x+a)}dx\right],\\
J^*_{36j} &= \left[\sum_{i=0}^{j}\frac{\Gamma(i+a)\Gamma(j-i+a)}{\Gamma(i+1)\Gamma(j-i+a+1)} - \sum_{i=0}^{j}\frac{1}{(i+a)^{1-a}(j-i+a)}\right].
\end{aligned}
$$

The priors of θ_1 and θ_2 are independent Beta$(a, 1)$. It follows from Chao & Strawderman (1972) that

$$
\begin{aligned}
\mathrm{E}\left(\frac{1}{N-n+1}\right) &= \mathrm{E}\left(\mathrm{E}\left(\frac{1}{N-n+1} \mid \boldsymbol{\theta}\right)\right) \\
&= B^{-2}(a,1)\int_0^1\int_0^1 \frac{1-\left(1-\prod_{i=1}^2(1-\theta_i)\right)^{N+1}}{(N+1)\prod_{i=1}^2(1-\theta_i)}\prod_{i=1}^2 \theta_i^{a-1}d\theta_1 d\theta_2 \\
&= B^{-2}(a,1)\int_0^1\int_0^1 \frac{1-\left(1-\prod_{i=1}^2\delta_i\right)^{N+1}}{(N+1)\prod_{i=1}^2\delta_i}\prod_{i=1}^2 (1-\delta_i)^{a-1}d\delta_1 d\delta_2 \\
&= \frac{B^{-2}(a,1)}{N+1}\sum_{j=0}^N\int_0^1\int_0^1 (1-\delta_2+\delta_2(1-\delta_1))^j\prod_{i=1}^2(1-\delta_i)^{a-1}d\delta_1 d\delta_2 \\
&= \frac{B^{-2}(a,1)}{N+1}\sum_{j=0}^N\frac{\Gamma(j+1)}{\Gamma(j+a+1)}\sum_{i=0}^j\frac{\Gamma(i+a)\Gamma(j-i+a)}{\Gamma(i+1)\Gamma(j-i+a+1)} \\
&= J_{34}(N)+J_{35}(N)+J_{36}(N), \qquad \text{(A.106)}
\end{aligned}
$$

The orders of $J_{34}(N)$, $J_{35}(N)$ and $J_{36}(N)$ are summarized in the following Fact.

Fact A.0.2. *For fixed $0 < a \leq 1$, sending N to infinity, we have*

$$
\begin{aligned}
&\frac{B^{-2}(a,1)\log^2 N}{2(N+1)} \leq J_{34}(N) \leq \frac{B^{-2}(a,1)\log^2 N}{2(N+1)} + \frac{B^{-2}(a,1)(\log N)^{3/2}}{a(N+1)}, && \text{(A.107)} \\
&J_{35}(N) \leq \frac{B^{-2}(a,1)\log N}{N+1}, && \text{(A.108)} \\
&J_{36}(N) \leq \frac{B^{-2}(a,1)\log N}{N+1}. && \text{(A.109)}
\end{aligned}
$$

Proof. We first prove (A.107). Let $y = x/j$ and $J^*_{34}(j)$ can be written as

$$
J_{34j} = \frac{1}{j^{1-a}}\int_0^1 \frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy. \qquad \text{(A.110)}
$$

We can show that integral has order $\log j$ for large j. Let $\epsilon = (\log j)^{-1/2}$. First, we integrate from 0 to $1-\epsilon$.

$$\begin{aligned}\int_0^{1-\epsilon} \frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy &\le \frac{1}{\epsilon+a/j}\int_0^{1-\epsilon}\frac{1}{(y+a/j)^{1-a}}dy \\ &= \frac{1}{a(\epsilon+a/j)}\left[(1-\epsilon+a/j)^a-(a/j)^a\right] \le \frac{(1-\epsilon+a/j)^a}{a(\epsilon+a/j)}. \end{aligned} \tag{A.111}$$

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

$$\begin{aligned}&\int_{1-\epsilon}^{1} \frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy \\ \le\ & \frac{1}{(1-\epsilon+a/j)^{1-a}}\int_{1-\epsilon}^{1}\frac{1}{1-y+a/j}dy = \frac{1}{(1-\epsilon+a/j)^{1-a}}\log\frac{\epsilon j+a}{a} \end{aligned}\tag{A.112}$$

On the other hand,

$$\begin{aligned}&\int_{1-\epsilon}^{1} \frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy \\ \ge\ & \frac{1}{(1+a/j)^{1-a}}\int_{1-\epsilon}^{1}\frac{1}{1-a+a/j}dy = \frac{1}{(1+a/j)^{1-a}}\log\frac{\epsilon j+a}{a}. \end{aligned}\tag{A.113}$$

It follows from (A.111), (A.112) and (A.113) that

$$\begin{aligned}\frac{1}{(1+a/j)^{1-a}}\log\frac{\epsilon j+a}{a} &\le \int_0^1 \frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy \\ &\le \frac{1}{(1-\epsilon+a/j)^{1-a}}\log\frac{\epsilon j+a}{a}+\frac{(1-\epsilon+a/j)^a}{\epsilon(\epsilon+a/j)}. \end{aligned}\tag{A.114}$$

We notice that when j is sufficiently large,

$$\frac{\log \epsilon j}{\log j} \le \frac{\log(\epsilon j+a)/a}{\log j} \le \frac{2\log\epsilon j}{\log j} \tag{A.115}$$

and

$$\lim_{j\to\infty}\frac{\log\epsilon j}{\log j} = 1+\lim_{j\to\infty}\frac{\log(\log j)^{-1/2}}{\log j} = 1-\lim_{j\to\infty}\frac{\log(\log j)}{2\log j} = 1. \tag{A.116}$$

As a consequence, we have

$$\lim_{j\to\infty}\frac{1}{(1+a/j)^{1-a}}\frac{\log(\epsilon j+a)/a}{\log j} = 1, \tag{A.117}$$

$$\lim_{j\to\infty}\frac{1}{(1-\epsilon+a/j)^{1-a}}\frac{\log(\epsilon j+a)/a}{\log j} = 1. \tag{A.118}$$

It is obvious that

$$\frac{(1-\epsilon+a/j)^a}{a(\epsilon+a/j)\log j}\sim\frac{1}{a(\log j)^{1/2}}. \tag{A.119}$$

It follows from (A.114), (A.117), (A.118), and (A.119) that

$$\log j\le\int_0^1\frac{1}{(y+a/j)^{1-a}(1-y+a/j)}dy\le\log j+\frac{1}{a}(\log j)^{1/2}. \tag{A.120}$$

Combining (A.120) and (A.110), we have

$$\frac{\log j}{j^{1-a}}\le J_{34j}\le\frac{\log j}{j^{1-a}}+\frac{(\log j)^{1/2}}{aj^{1-a}}. \tag{A.121}$$

We can decompose $J_{34}(N)$ as

$$J_{34}(N)=J_{341}(N)+J_{342}(N)+J_{343}(N), \tag{A.122}$$

where

$$J_{341}(N) = \frac{B^{-2}(a,1)}{N+1}\int_1^N\frac{\log x}{x}dx, \tag{A.123}$$

$$J_{342}(N) = \frac{B^{-2}(a,1)}{N+1}\left[\sum_{j=1}^N\frac{\log j}{j}-\int_1^N\frac{\log x}{x}dx\right], \tag{A.124}$$

$$J_{343}(N) = \frac{B^{-2}(a,1)}{N+1}\sum_{j=1}^N\left[\frac{\Gamma(j+1)}{\Gamma(j+a+1)}J_{34}^*(j)-\frac{\log j}{j}\right]. \tag{A.125}$$

By L'hospital's rule, we can show that

$$\lim_{N\to\infty} \frac{J_{341}(N)}{B^{-2}(a,1)\log^2 N/(N+1)} = \lim_{N\to\infty} \frac{\int_1^N \frac{\log x}{x} dx}{\log^2 N} = \lim_{N\to\infty} \frac{\frac{1}{N}\log N}{\frac{2}{N}\log N} = \frac{1}{2}. \quad \text{(A.126)}$$

To study the limit of $J_{342}(N)$, let us define

$$g_1(x) = \frac{\log x}{x}, \quad 1 \le x \le N. \quad \text{(A.127)}$$

Then

$$\begin{aligned} J_{342}(N) &= \frac{B^{-2}(a,1)}{N+1}\left[\sum_{j=1}^{N} g_1(j) - \int_1^N g_1(x)dx\right] \\ &= \frac{B^{-2}(a,1)}{N+1}\left(\sum_{j=1}^{2} g_1(j) - \int_1^3 g_1(x)dx + \sum_{j=3}^{N} g_1(j) - \int_3^N g_1(x)dx\right). \end{aligned} \quad \text{(A.128)}$$

Since $g_1(\cdot)$ increases when $1 \le x \le e$ and decreases when $e < x \le N$, the first term has an order of $\frac{1}{N+1}$. For the second term, since $g_1(\cdot)$ decreases from 3 to N, the following inequality holds:

$$\sum_{j=4}^{N} g_1(j) \le \int_3^N g_1(x)dx \le \sum_{j=3}^{N-1} g_1(j). \quad \text{(A.129)}$$

This is equivalent to

$$\frac{\log N}{N+1} = g_1(N) \le \sum_{j=3}^{N} g_1(j) - \int_3^N g_1(x)dx \le g_1(3) = \frac{\log 3}{4}. \quad \text{(A.130)}$$

It follows from (A.128) and (A.130) that as $N \to \infty$,

$$J_{342}(N) \sim \frac{1}{N+1}. \quad \text{(A.131)}$$

For $J_{343}(N)$, we have

$$\begin{aligned} J_{343}(N) &\le \frac{B^{-2}(a,1)}{N+1}\sum_{j=1}^{N}\frac{(\log j)^{1/2}}{aj} \\ &\approx \frac{B^{-2}(a,1)}{N+1}\int_1^N \frac{(\log x)^{1/2}}{ax}dx \le \frac{B^{-2}(a,1)(\log N)^{3/2}}{a(N+1)}. \end{aligned} \tag{A.132}$$

Combining (A.122), (A.126), (A.131), and (A.132), we get (A.107).

Now we prove (A.108). Define

$$g_2(x) = \frac{1}{(x+a)^{1-a}(j-x+a)}, \quad 0 \le x \le j. \tag{A.133}$$

Take the derivative of $\log g_2(x)$ with respect to x and we have

$$\frac{d}{dx}\log g_2(x) = \frac{(2-a)x - [(1-a)(j+a)-a]}{(x+a)(j-x+a)}. \tag{A.134}$$

Let $j_0 = [\frac{(2-a)x-[(1-a)(j+a)-a]}{(x+a)(j-x+a)}]$. Notice that $g_2(x)$ decreases when $0 \le x \le j_0$ and increases when $j_0+1 \le x \le j$. Therefore we have

$$0 \le \sum_{i=0}^{j_0} g_2(i) - \int_0^{j_0} g_2(x)dx \le g_2(0) = \frac{1}{a^{1-a}(j+a)}, \tag{A.135}$$

$$0 \le \sum_{i=j_0+1}^{j} g_2(i) - \int_{j_0}^{j} g_2(x)dx \le g_2(j) = \frac{1}{a(j+a)^{1-a}}. \tag{A.136}$$

Adding up (A.135) and (A.136) will yield that

$$0 \le \sum_{i=0}^{j} g_2(i) - \int_0^j g_2(x)dx \le \frac{1}{a^{1-a}(j+a)} + \frac{1}{a(j+a)^{1-a}}. \tag{A.137}$$

Then we have

$$J_{35j} \le \frac{1}{a^{1-a}(j+a)} + \frac{1}{a(j+a)^{1-a}} \approx \frac{1}{a(j+a)^{1-a}}, \tag{A.138}$$

and

$$J_{35}(N) \leq \frac{B^{-2}(a,1)}{N+1} \sum_{j=0}^{N} \frac{\Gamma(j+1)}{\Gamma(j+a+1)} \frac{1}{(j+a)^{1-a}} \sim \frac{B^{-2}(a,1)\log N}{a(N+1)}. \tag{A.139}$$

To study $J_{36}(N)$, we observe the facts that

$$\frac{\Gamma(i+a)}{\Gamma(i+1)} = \frac{1}{(i+a)^{1-a}} \left[1 + \frac{a(1-a)}{2(i+a)} + T_1(i)\right], \tag{A.140}$$

$$\frac{\Gamma(j-i+a)}{\Gamma(j-i+1)} = \frac{1}{(j-i+a)^{1-a}} \left[1 + \frac{a(1-a)}{2(j-i+a)} + T_2(i)\right], \tag{A.141}$$

$$\frac{\Gamma(j-i+a)}{\Gamma(j-i+a+1)} = \frac{1}{(j-i+a)^{a}} \left[1 - \frac{a(1-a)}{2(j-i+a)} + T_3(i)\right], \tag{A.142}$$

where

$$T_1(i) = O\left(\frac{1}{(i+a)^2}\right), \quad i \to \infty, \tag{A.143}$$

$$T_2(i) = O\left(\frac{1}{(j-i+a)^2}\right), \quad j-i \to \infty, \tag{A.144}$$

$$T_3(i) = O\left(\frac{1}{(j-i+a)^2}\right), \quad j-i \to \infty. \tag{A.145}$$

We can rewrite J_{36j} as

$$\begin{aligned} J_{36j} &= \sum_{i=0}^{j} \frac{1}{(i+a)^{1-a}(j-i+a)} \Bigg[\left(1 + \frac{a(1-a)}{2(i+a)} + T_1(i)\right) \\ &\quad \left(1 + \frac{a(1-a)}{2(j-i+a)} + T_2(i)\right)\left(1 - \frac{a(1-a)}{2(j-i+a)} + T_3(i)\right) - 1\Bigg]. \\ &= J^*_{361j} + J^*_{362j} + o\left(\min\left(J^*_{361j}, J^*_{362j}\right)\right), \end{aligned} \tag{A.146}$$

where J^*_{361j} and J^*_{362j} are the leading terms and

$$J^*_{361j} = \sum_{i=0}^{j} \frac{a(1-a)}{2(i+a)^{2-a}(j-i+a)}, \tag{A.147}$$

$$J^*_{362j} = -\sum_{i=0}^{j} \frac{2a^2}{2(i+a)^{1-a}(j-i+a)^2}. \tag{A.148}$$

It is easy to show that both J^*_{361j} and J^*_{362j} are of the order of $1/j^{1-a}$. Therefore we have

$$J_{36}(N) \le \frac{B^{-2}(a,1)}{N+1} \sum_{j=0}^{N} \frac{\Gamma(j+1)}{\Gamma(j+a+1)} \frac{1}{j^{1-a}} \sim \frac{B^{-2}(a,1)\log N}{N+1}. \tag{A.149}$$

This completes the proof of the Fact. $\square$

Fact A.0.2 and (A.106) yield (A.103). To prove (A.104), similar as (A.37), we have

$$\mathrm{E}\left(\frac{1}{(N-n+1)^2}\right) = \sum_{l=1}^{3} \frac{B^{-2}(a,1)\Gamma(N+1)}{\Gamma(N+a+1)} \sum_{n=0}^{N} \frac{J^*_{3l}(n)}{(N-n+1)^2}. \tag{A.150}$$

Proposition A.0.2 implies that for $l=2$,

$$\frac{B^{-2}(a,1)\Gamma(N+1)}{\Gamma(N+a+1)} \sum_{n=0}^{N} \frac{J^*_{32}(n)}{(N-n+1)^2} \le J_{32}(N) \le \frac{1}{N}; \tag{A.151}$$

and for $l=3$,

$$\frac{B^{-2}(a,1)\Gamma(N+1)}{\Gamma(N+a+1)} \sum_{n=0}^{N} \frac{J^*_{33}(n)}{(N-n+1)^2} \le J_{33}(N) \le \frac{\log N}{N}. \tag{A.152}$$

For $l = 1$, we use integral to approximate the partial sum. Notice that

$$
\begin{aligned}
&\frac{J_{31}^*(n)}{(N-n+1)^2} \\
= \;& \frac{1}{(N-n+1)^2}\int_0^n \frac{1}{(x+1)^{1-a}(n-x+1)^{1-a}(N-x+1)^a}dx \\
\leq \;& \frac{1}{(N-n+1)^2}\int_0^n \frac{1}{(x+1)^{1-a}(n-x+1)}dx \\
\leq \;& \frac{1}{(N-n+1)^2}\frac{\log(n+1)+2^{1-a}(1-\log 2)/a}{(n+1)^{1-a}} \\
\leq \;& 2\log(N+1)\frac{1}{(n+1)^{1-a}(N-n+1)^2}.
\end{aligned} \tag{A.153}
$$

Therefore we get

$$
\begin{aligned}
&\frac{B^{-2}(a,1)\Gamma(N+1)}{\Gamma(N+a+1)}\sum_{n=0}^{N}\frac{J_{31}^*(n)}{(N-n+1)^2} \\
\leq \;& \frac{2B^{-2}(a,1)\Gamma(N+1)\log(N+1)}{\Gamma(N+a+1)}\int_0^N \frac{1}{(x+1)^{1-a}(N-x+1)^2}dx \\
\approx \;& \frac{2B^{-2}(a,1)\Gamma(N+1)\log(N+1)}{\Gamma(N+a+1)}\int_0^N \frac{1}{(x+1)^{1-a}(N-x+1)^2}dx \\
= \;& \frac{2B^{-2}(a,1)\Gamma(N+1)\log(N+1)}{N^{2-a}\Gamma(N+a+1)}\int_0^1 \frac{1}{(y+1/N)^{1-a}(1-y+1/N)^2}dy \\
\sim \;& \frac{2B^{-2}(a,1)N\Gamma(N+1)\log(N+1)}{N^{2-a}\Gamma(N+a+1)} \approx \frac{2B^{-2}(a,1)\log(N+1)}{N}.
\end{aligned} \tag{A.154}
$$

Combining (A.150), (A.151), (A.152), and (A.154), we get (A.104). □

Combining Propositions A.0.2 and A.0.3, we have the following corollary.

Corollary A.0.1. *For $k = 2$ and $0 < a \leq 1$,*

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

$$
\frac{C_2B^{-2}(a,b)\log(N+1)}{N^b} \leq E_2(J_3(N)) \leq \frac{C_2B^{-2}(a,b)\log(N+1)}{N^b} + \frac{\log^b N}{N^b}; \tag{A.155}
$$

(2) *when* $b = 1$,

$$\frac{B^{-2}(a,1)\log^2 N}{2(N+1)} \le E_2(J_3(N)) \le \frac{B^{-2}(a,1)\log^2 N}{2(N+1)} + \frac{B^{-2}(a,1)\log^{3/2} N}{a(N+1)}. \tag{A.156}$$

In general, we assume there are $k+1$ sampling occasions. Then we have

$$n \mid N, \boldsymbol{\theta} \sim \text{Bin}\left(N, 1 - \prod_{u=1}^{k+1}(1-\theta_i)\right). \tag{A.157}$$

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 $J_3(N)$ can be expressed as

$$\begin{aligned}
\mathrm{E}_{k+1}(J_3(N)) &= \sum_{n=0}^{N}\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\frac{N!}{n!(N-n)!}\int_0^1\cdots\int_0^1 \\
&\left(1-\prod_{u=1}^{k+1}\delta_u\right)^n\left(\prod_{u=1}^{k+1}x_u\right)^{N-n} B^{-(k+1)}(a,b)\prod_{u=1}^{k+1}\delta_u^{b-1}\prod_{u=1}^{k+1}(1-\delta_u)^{a-1}d\delta_1\cdots d\delta_{k+1} \\
&= B^{-(k+1)}(a,b)\sum_{n=0}^{N}\sum_{i=0}^{\infty}\frac{1}{(N-n+1+i)^2}\frac{N!}{n!(N-n)!}\int_0^1\cdots\int_0^1 \\
&(1-\delta_{k+1}+\delta_{k+1}(1-\prod_{u=1}^{k}\delta_u))^n\prod_{u=1}^{k+1}\delta_u^{N-n+b-1}\prod_{u=1}^{k+1}(1-\delta_u)^{a-1}d\delta_1\cdots d\delta_{k+1} \\
&= \sum_{n=0}^{N}\sum_{i=0}^{\infty}\frac{B^{-(k+1)}(a,b)\Gamma(N+1)}{(N-n+1+i)^2\Gamma(n+1)\Gamma(N-n+1)}\sum_{j=0}^{n}\frac{\Gamma(n+1)}{\Gamma(j+1)\Gamma(n-j+1)} \\
&\int_0^1\cdots\int_0^1(1-\delta_{k+1})^j\delta_{k+1}^{n-j}\left(1-\prod_{u=1}^{k}\delta_u\right)^{n-j}\delta_u^{N-n+b-1}\prod_{u=1}^{k+1}(1-\delta_u)^{a-1}d\delta_1\cdots d\delta_{k+1}
\end{aligned} \tag{A.158}$$

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

$$
\begin{aligned}
& \text{The RHS} \\
= \quad & \frac{B^{-(k+1)}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{n=0}^{N} \left(\sum_{i=0}^{\infty} \frac{1}{(N-n+1+i)^2} \right) \frac{1}{\Gamma(N-n+1)} \sum_{j=0}^{n} \frac{\Gamma(j+a)\Gamma(N-j+b)}{\Gamma(j+1)\Gamma(n-j+1)} \\
& \int_0^1 \cdots \int_0^1 \left(1 - \prod_{u=1}^{k} \delta_u \right)^{n-j} \prod_{u=1}^{k} \delta_u^{N-n+b-1} \prod_{u=1}^{k} (1-\delta_u)^{a-1} d\delta_1 \cdots d\delta_k \\
= \quad & \frac{B^{-(k+1)}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{j=0}^{N} \sum_{r=0}^{N-j} \left(\sum_{i=0}^{\infty} \frac{1}{(N-j-r+1+i)^2} \right) \frac{\Gamma(j+a)\Gamma(N-j+b)}{\Gamma(j+1)\Gamma(N-j+1)} \\
& \int_0^1 \cdots \int_0^1 \frac{\Gamma(N-j+1)}{\Gamma(r+1)\Gamma(N-j-r+1)} \left(1 - \prod_{u=1}^{k} \delta_u \right)^{r} \prod_{u=1}^{k} \delta_u^{N-j-r+b-1} \prod_{u=1}^{k} (1-\delta_u)^{a-1} d\delta_1 \cdots d\delta_k \\
= \quad & \frac{B(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{m=0}^{N} \frac{\Gamma(N-m+a)\Gamma(m+b)}{\Gamma(N-m+1)\Gamma(m+1)} \sum_{r=0}^{m} \left(\sum_{i=0}^{\infty} \frac{1}{(m-r+1+i)^2} \right) \\
& \int_0^1 \cdots \int_0^1 \frac{B^{-k}(a,b)\Gamma(m+1)}{\Gamma(r+1)\Gamma(m-r+1)} \left(1 - \prod_{u=1}^{k} \delta_u \right)^{r} \prod_{u=1}^{k} \delta_u^{m-r+b-1} \prod_{u=1}^{k} (1-\delta_u)^{a-1} d\delta_1 \cdots d\delta_k \\
= \quad & \frac{B(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{m=0}^{N} \frac{\Gamma(m+b)\Gamma(N-m+a)}{\Gamma(m+1)\Gamma(N-m+1)} \mathrm{E}_k(J_3(m)),
\end{aligned} \tag{A.159}
$$

where $\mathrm{E}_k(J_3(N))$ denotes the expectation of $J_3(N)$ 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 $m \to \infty$,

$$|\mathrm{E}_k(J_3(m)) - \frac{C_2 B^{-k}(a,b) \log^{k-1}(m)}{m^b}| \leq \frac{C_5(k) \log^{k-2+b}(m)}{m^b}. \tag{A.160}$$

For $b = 1$, we assume that

$$|\mathrm{E}_k(J_3(m)) - \frac{B^{-k}(a,1) \log^{k}(m)}{m}| \leq \frac{C_6(k) \log^{k-1/2}(m)}{m}. \tag{A.161}$$

We have shown in Corollary A.0.1 that assumptions (A.160) and (A.161) hold for $k = 2$.

Proposition A.0.4. *Assuming $k+1$ sampling occasions and assumptions (A.160)* *and (A.161), we have for* $0 < b < 1$,

$$|E_{k+1}(J_3(N)) - \frac{C_2 B^{-(k+1)}(a,b)\log^k(N)}{N^b}| \le \frac{C_5(k+1)\log^{k-1+b}(N)}{N^b}. \tag{A.162}$$

For $b = 1$,

$$|E_{k+1}(J_3(N)) - \frac{B^{-(k+1)}(a,1)\log^{k+1}(N)}{N}| \le \frac{C_6(k+1)\log^{k+1/2}(N)}{N}. \tag{A.163}$$

Proof. We focus on the case when $0 < b < 1$ and prove (A.162). We decompose $\mathrm{E}_{k+1}(J_3(N))$ after removing the first $M+1$ terms (M is a fixed integer).

$$\begin{aligned}\mathrm{E}_{k+1}(J_3(N)) &= \frac{C_2 B^{-(k+1)}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{m=M}^{N} \frac{\Gamma(m+b)\Gamma(N-m+a)\log^{k-1} m}{m^b \Gamma(m+1)\Gamma(N-m+1)} \\ &+ \frac{B(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{m=M}^{N} \frac{\Gamma(m+b)\Gamma(N-m+a)}{\Gamma(m+1)\Gamma(N-m+1)} \left(\mathrm{E}_k(J_3(m)) - \frac{C_2 B^{-k}(a,b)\log^{k-1} m}{m^b}\right) \\ &\equiv J_{3k1}(N) + J_{3k2}(N).\end{aligned} \tag{A.164}$$

By the similar approach in the proof of Proposition A.0.2, we can approximate $J_{3k1}(N)$ by

$$J_{3k1}(N) \approx \frac{C_2 B^{-(k+1)}(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} I_k(N), \tag{A.165}$$

where

$$I_k(N) = \int_M^N \frac{\log^{k-1} x}{(x+1)(N-x+1)^{1-a}} dx. \tag{A.166}$$

Fact A.0.3. *For any* $0 < b < 1$ *and* $k \ge 2$, *there exist positive constants* C_7 *such that*

$$\frac{\log^k N}{N^{1-a}} - \frac{C_7 \log^{k-1} N}{N^{1-a}} \le I_k(N) \le \frac{\log^k N}{N^{1-a}}. \tag{A.167}$$

Proof. First we find the order of $I_k(N)$ for $k = 2$. Let $y = N - x$,

$$\begin{aligned} I_2(N) = & \int_M^N \frac{\log x}{(x+1)(N-x+1)^{1-a}} dx \\ = & \frac{1}{N^{1-a}} \int_{M/N}^1 \frac{\log y + \log N}{(y+1/N)(1-y+1/N)^{1-a}} dy \\ \sim & \frac{\log^2 N}{N^{1-a}} + \frac{1}{N^{1-a}} \int_{M/N}^1 \frac{\log y}{(y+1/N)(1-y+1/N)^{1-a}} dy \\ \leq & \frac{\log^2 N}{N^{1-a}}. \end{aligned} \tag{A.168}$$

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

$$\begin{aligned} & \frac{1}{N^{1-a}} \int_{M/N}^{1/2} \frac{\log y}{(y+1/N)(1-y+1/N)^{1-a}} dy \\ \geq & \frac{1}{(1/2+1/N)N^{1-a}} \int_{M/N}^{1/2} \frac{\log y}{(1-y+1/N)^{1-a}} dy \\ \geq & \frac{\log M - \log N}{(1/2+1/N)N^{1-a}} \int_{M/N}^{1/2} \frac{1}{(1-y+1/N)^{1-a}} dy \\ = & \frac{\log M - \log N}{a(1/2+1/N)N^{1-a}} \left[(1-M/N+1/N)^a - (1/2+1/N)^a\right] \\ \approx & \frac{\log M - \log N}{aN^{1-a}}. \end{aligned} \tag{A.169}$$

We also have

$$\begin{aligned} & \frac{1}{N^{1-a}} \int_{1/2}^1 \frac{\log y}{(y+1/N)(1-y+1/N)^{1-a}} dy \geq -\frac{\log 2}{N^{1-a}} \int_{1/2}^1 \frac{1}{(1-y+1/N)^{1-a}} dy \\ = & -\frac{\log 2}{a(1/2+1/N)N^{1-a}} \left[(1/2+1/N)^a - (1/N)^a\right] \approx -\frac{\log 2}{aN^{1-a}}. \end{aligned} \tag{A.170}$$

It follows from (A.168), (A.169), and (A.170) that

$$\frac{\log^2 N}{N^{1-a}} + \frac{\log M - \log N}{aN^{1-a}} - \frac{\log 2}{aN^{1-a}} \leq I_2(N) \leq \frac{\log^2 N}{N^{1-a}}. \tag{A.171}$$

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

$$\begin{aligned}
I_{k+1}(N) &= \int_M^N \frac{\log^k x}{(x+1)(N-x+1)^{1-a}}dx \\
&= \frac{1}{N^{1-a}} \int_{M/N}^1 \frac{(\log y + \log N)^k}{(y+1/N)(1-y+1/N)^{1-a}}dy \\
&= \frac{1}{N^{1-a}} \int_{M/N}^1 \frac{(\log y + \log N)^{k-1}(\log y + \log N)}{(y+1/N)(1-y+1/N)^{1-a}}dy \\
&= I_k(N)\log N + \frac{1}{N^{1-a}} \int_{M/N}^1 \frac{(\log y + \log N)^{k-1}\log y}{(y+1/N)(1-y+1/N)^{1-a}}dy \\
&\leq I_k(N)\log N. \qquad (A.172)
\end{aligned}$$

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

$$\begin{aligned}
&\frac{1}{N^{1-a}} \int_{M/N}^{1/2} \frac{(\log y + \log N)^{k-1}\log y}{(y+1/N)(1-y+1/N)^{1-a}}dy \\
\geq\ & \frac{(\log M - \log N)\log^{k-1} M}{N^{1-a}} \int_{M/N}^{1/2} \frac{1}{(y+1/N)(1-y+1/N)^{1-a}}dy \\
=\ & \frac{(\log M - \log N)\log^{k-1} M}{(1/2+1/N)N^{1-a}} \int_{M/N}^{1/2} \frac{1}{(1-y+1/N)^{1-a}}dy \\
=\ & \frac{(\log M - \log N)\log^{k-1} M}{a(1/2+1/N)N^{1-a}} \left[(1-M/N+1/N)^a - (1/2+1/N)^a\right] \\
\approx\ & \frac{2(1-2^{-a})(\log M - \log N)\log^{k-1} M}{aN^{1-a}}. \qquad (A.173)
\end{aligned}$$

Secondly,

$$\begin{aligned}
&\frac{1}{N^{1-a}} \int_{1/2}^{1} \frac{(\log y + \log N)^{k-1}\log y}{(y+1/N)(1-y+1/N)^{1-a}}dy \\
\geq\ & -\frac{(\log N - \log 2)^{k-1}\log 2}{N^{1-a}} \int_{1/2}^{1} \frac{1}{(y+1/N)(1-y+1/N)^{1-a}}dy \\
\geq\ & -\frac{(\log N - \log 2)^{k-1}\log 2}{(1/2+1/N)N^{1-a}} \int_{1/2}^{1} \frac{1}{(1-y+1/N)^{1-a}}dy \\
=\ & -\frac{(\log N - \log 2)^{k-1}\log 2}{a(1/2+1/N)N^{1-a}} \left[(1/2+1/N)^a - (1/N)^a\right] \\
\approx\ & -\frac{2^{1-a}\log 2(\log N - \log 2)^{k-1}}{aN^{1-a}}. \qquad (A.174)
\end{aligned}$$

It follows from (A.172), (A.173), and (A.174) that

$$\begin{aligned} 0 &\le I_k(N)\log N - I_{k+1}(N) \\ &\le \frac{2(1-2^{-a})(\log N - \log M)\log^{k-1} M}{aN^{1-a}} + \frac{2^{1-a}\log 2(\log N - \log 2)^{k-1}}{aN^{1-a}} \end{aligned} \tag{A.175}$$

Combining (A.171) and (A.175), we have proved Fact A.0.3. □

It follows from (A.165) and Fact A.0.3 that

$$|J_{3k1}(N) - \frac{C_2 B^{-(k+1)}(a,b)\log^k N}{N^b}| \le \frac{C_2 C_7 B^{-(k+1)}(a,b)\log^{k-1} N}{N^b}. \tag{A.176}$$

Similarly we can show that

$$\begin{aligned} |J_{3k2}(N)| &\le \frac{B(a,b)\Gamma(N+1)}{\Gamma(N+a+b)} \sum_{m=M}^{N} \frac{\Gamma(m+b)\Gamma(N-m+a)\log^{k-2+b} N}{m^b\Gamma(m+1)\Gamma(N-m+1)} \\ &\le \frac{C_8 \log^{k-1+b} N}{N^b}, \end{aligned} \tag{A.177}$$

where C_8 is a positive constant. Combining (A.164), (A.176), (A.177), we get (A.162). In the case of $b = 1$, (A.163) can be proved similarly. □

Proposition A.0.1 leads to (2.22) and (2.24), and Proposition A.0.4 yield (2.23) and (2.25) in Lemma 2.3.1.

Bibliography

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

Agresti, A. (1994), 'Simple capture-recapture models permitting unequal catchability and variable sampling effort', *Biometrics* **50**, 494–450.

Barger, K. & Bunge, J. (2008), 'Bayesian estimation of the number of species using noninformative priors', *Biometrical Journal* **50**, 1064–1076.

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. O. & Bernardo, J. M. (1992), On the development of reference priors, *in* 'Bayesian Statistics 4. (J. M. Bernardo, J. O. Berger, D. V. Lindley and A. F. M. Smith, eds)', Oxford University Press, London, pp. 35–60 (with discussion).

Berger, J. O., Bernardo, J. M. & Sun, D. (2009b), 'Natural induction: an objective Bayesian approach', *Rev. R. Acad. Cienc. Exactas Fis. Nat., Ser. A Mat.* **103**, in press.

Berger, J. O., Bernardo, J. M. & Sun, D. (2011), 'Objective priors for discrete parameter space', *Journal of American Statistical Association* .

Bernardo, J. M. (1979), 'Reference posterior distributions for Bayesian inference', *Journal of the Royal Statistical Society, Series B* **41**, 113–147 (with discussion). Reprinted in *Bayesian Inference* **1** (G. C. Tiao and N. G. Polson, eds.), Oxford: Edward Elgar, 1995, 229–263.

Bernardo, J. M. & Smith, A. F. M. (1994), *Bayesian Theory*, John Wiley & Sons.

Blumenthal, S. & Marcus, R. (1975), 'Estimating population-size with exponential failure', *Journal Of The American Statistical Association* **70**, 913–922.

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

Burnham, K. & Overton, K. (1978), 'Estimation of the size of a closed population when capture probabilities vary among animals', *Biometrika* **65**, 625–633.

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

Dorazio, R. & Royle, J. (2003), 'Mixture models for estimating the size of a closed population when capture rates vary among individuals', *Biometrics* **59**, 351–364.

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

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

Gilks, W., Best, N. & Tan, K. (1995), 'Adaptive rejection metropolis sampling within gibbs sampling', *Applied Statistics* **44**, 455–472.

Goudie, I. B. J. & Goldie, C. M. (1981), 'Initial size estimation for the pure death process', *Biometrika* **68**, 543–550.

Huggins, R. (1989), 'On the statistical analysis of capture experiments', *Biometrika* **76**, 133–40.

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

Joe, H. & Reid, N. (1985), 'Estimating the number of faults in a system', *Journal Of The American Statistical Association* **80**, 222–6.

Kinderman, A. & Monahan, J. (1977), 'Computer generation of random variables using the ratio of uniform deviates', *ACM Trans. Math. Software* **3**, 257–260.

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

Mukerjee, R. & Ghosh, M. (1997), 'Second-order probability matching priors', *Biometrika* **84**, 970–975.

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

Pledger, S. (2000), 'Unified maximum likelihood estimates for closed capture-recapture models using mixtures', *Biometrics* **56**, 434–442.

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

Raftery, A. E. (1988a), 'Inference for the binomial n parameter: a hierarchical Bayes approach', *Biometrika* **75**, 223–228.

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

Royle, J. (2009), 'Analysis of capture-recapture models with individual covariates using data augmentation', *Biometrics* **65**, 267–274.

Scholz, F.-W. (1986), 'Software reliability modeling and analysis', *IEEE Trans. Software Eng.* **12**(1), 25–31.

Silvapulle, M. (1981), 'On the existence of maximum likelihood estimators for the binomial response models', *Journal of the Royal Statistical Society, Series B* **43**, 310–313.

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

Speckman, P., Lee, J. & Sun, D. (2009), 'Existence of the mle and propriety of the posteriors for a general multinomial choice model', *Statistica Sinica* **19**, 731–748.

Stoklosa, J., Hwang, W.-h., Wu, S.-h. & Huggins, R. (2011), 'Heterogeneous capture-recapture models with covariates: A partial likelihood approach for closed populations', *Biometrics* **67**, 1659–1665.

Wang, X., Chong, H. & Sun, D. (2007), 'Bayesian population estimation for small sample capture-recapture data using noninformative priors', *Journal Of Statistical Planning And Inference* **137**, 1099–1118.

Xu, C., Sun, D. & He, C. (2012), 'Objective bayesian analysis for a capture-recapture model', *Technical Report* .

VITA

Chang Xu was born on November 11, 1984, in Jiujiang, Jiangxi Province, People's Republic of China. He received his B.S. in Mathematics from Peking University in 2006. He joined the graduate program in the Department of Statistics at the University of Missouri in January 2008. He will receive his Ph.D. in Statistics in July 2012. As of August 2012, He will be a visiting assistant professor in the Department of Mathematical Sciences, University of Cincinnati.