Statistical inference for functions of the parameters
of a linear mixed model

by

Jia Liu

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
William Q. Meeker, Co-major Professor
Daniel J. Nordman, Co-major Professor
Peng Liu
Ranjan Maitra
Alyson G. Wilson

Iowa State University
Ames, Iowa
2013

Copyright ©Jia Liu, 2013. All rights reserved.
DEDICATION

To my parents and grandma.
TABLE OF CONTENTS

LIST OF TABLES ............................................................... vi
LIST OF FIGURES ........................................................... vii
ACKNOWLEDGEMENTS ....................................................... viii
ABSTRACT ................................................................. x
CHAPTER 1. GENERAL INTRODUCTION .............................. 1
  1.1 Background ......................................................... 1
  1.2 Motivation .......................................................... 1
    1.2.1 Coverage Probabilities of Confidence Intervals for Functions of the
          Parameters of a Linear Mixed Model ................................. 2
    1.2.2 The Number of MCMC Draws Needed to Compute Bayesian Credible
          Bounds ................................................................. 2
    1.2.3 Implementation of an Algorithm for the Number of Draws Needed
          in Quantile Estimations ................................................. 3
  1.3 Dissertation Organization ......................................... 3

CHAPTER 2. COVERAGE PROBABILITIES OF CONFIDENCE INTERVALS FOR FUNCTIONS OF THE PARAMETERS OF A LINEAR MIXED EFFECTS MODEL ........................................ 4
  2.1 Introduction ........................................................ 5
    2.1.1 Background and Motivation .................................... 5
    2.1.2 Related Literature .............................................. 6
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.3</td>
<td>Overview</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>Development of a Canonical Model</td>
<td>7</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Probability of Flaw Detection in Titanium Forgings</td>
<td>7</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Probability of Failure in Accelerated Repeated Measures Degradation Model</td>
<td>8</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Cracks Detection in Fan Blades</td>
<td>9</td>
</tr>
<tr>
<td>2.2.4</td>
<td>A Canonical Model</td>
<td>10</td>
</tr>
<tr>
<td>2.3</td>
<td>Confidence Interval Methods</td>
<td>12</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Likelihood Confidence Interval Procedures</td>
<td>12</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Wald Confidence Interval Procedures</td>
<td>15</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Bootstrap Confidence Interval Procedures</td>
<td>17</td>
</tr>
<tr>
<td>2.3.4</td>
<td>Bayesian Credible Interval Procedures</td>
<td>19</td>
</tr>
<tr>
<td>2.4</td>
<td>Simulation Experiment</td>
<td>22</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Design of the Simulation Experiment</td>
<td>22</td>
</tr>
<tr>
<td>2.4.2</td>
<td>The Simulation Algorithm</td>
<td>23</td>
</tr>
<tr>
<td>2.5</td>
<td>Simulation Experiment Results</td>
<td>24</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Computational Time Results</td>
<td>25</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Coverage Probability Results</td>
<td>25</td>
</tr>
<tr>
<td>2.6</td>
<td>Concluding Remarks and Areas for Further Research</td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>32</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Background and Motivation</td>
<td>32</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Literature Review</td>
<td>33</td>
</tr>
<tr>
<td>3.1.3</td>
<td>Overview</td>
<td>35</td>
</tr>
<tr>
<td>3.2</td>
<td>Structure of MCMC Output</td>
<td>35</td>
</tr>
<tr>
<td>3.3</td>
<td>Quantile Estimation and the Number of Draws Determination</td>
<td>36</td>
</tr>
</tbody>
</table>
3.3.1 Quantile Estimation for i.i.d. Sequences 37
3.3.2 Quantile Estimation for MCMC Sequences 38
3.4 Application and Evaluation of the Algorithm 41
  3.4.1 Application 1: Linear model with mixed effects 42
  3.4.2 Application 2: Generalized linear model with a Poisson distribution 46
3.5 Concluding Remarks and Areas for Further Research 47

CHAPTER 4. IMPLEMENTATION OF AN ALGORITHM FOR THE
  NUMBER OF DRAWS NEEDED IN QUANTILE ESTIMATIONS 50
  4.1 Introduction 51
  4.2 Estimation Method 51
  4.3 Estimation Algorithm 52
  4.4 Implementations and R Code 54
  4.5 Discussion 56

CHAPTER 5. GENERAL CONCLUSIONS 57

BIBLIOGRAPHY 59
LIST OF TABLES

| Table 2.1 | Factor level combinations for $\sigma_i$ and $\beta_i$ | 22 |
| Table 2.2 | Values of $\mu_D$ and $y_{th}$ giving desired values of POD($x$) and $F(t; x)$ | 23 |
| Table 3.1 | Results of the application of linear mixed-effects model when $g_1 = \beta_1$ is the function of interest | 44 |
| Table 3.2 | Results of the application of linear mixed-effects model when $g_2 = \sigma_1/\sigma_\epsilon$ is the function of interest | 45 |
| Table 3.3 | Results of the application of generalized linear model with Poisson distribution when $g_1 = \beta$ is the function of interest | 47 |
| Table 3.4 | Results of the application of generalized linear model with Poisson distribution when $g_2 = P(Y_3 = 0)$ is the function of interest | 48 |
LIST OF FIGURES

Figure 2.1 Relative computational time for likelihood, Bayesian, and bootstrap methods comparing with Wald method. . . . . . . . . . . . 26

Figure 2.2 Left-tail (displayed on the left) and right-tail (displayed on the right) coverage probabilities for estimating the failure-time CDF versus the number $n$ of units when $\beta = 2$ and $\sigma^2 = 4$. Top to bottom, the rows correspond to results for $m=5, 15, 30$, respectively. The horizontal line indicates the nominal 5% level in each graph. 29

Figure 2.3 Left-tail (displayed on the left) and right-tail (displayed on the right) coverage probabilities for estimating POD function versus the number $n$ of units when $\beta = 2$ and $\sigma^2 = 4$. Top to bottom, the rows correspond to results for $m=5, 15, 30$, respectively. The horizontal line indicates the nominal 5% level in each graph. . . 30
ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere gratitude to everyone who gave me various help through my PhD journey.

I am extremely grateful to my advisors, Dr. William Q. Meeker and Dr. Daniel J. Nordman for their inspiring guidance, continuous support, generous advice, enthusiastic encouragement and immense knowledge during my PhD study and research work. I have been amazingly fortunate to have advisors who cared so much about my work, and who responded to my questions and queries so promptly. Their comments and criticisms at different stages of my research were very insightful and thought-provoking. I learned a lot from them.

My sincere thanks also goes to my committee members Dr. Peng Liu, Dr. Ranjan Maitra, and Dr. Alyson G. Wilson for many insightful discussions and suggestions. I would additionally like to thank Dr. Thomas Mathew as well as other faculty members in UMBC for their guidance throughout the initial stages of my graduate study in UMBC. I will forever be thankful to Department of Statistics at Iowa State University for providing me a lot of invaluable resource and opportunity to broaden my knowledge, skills and insights, and helping me prepared well for my career. Many thanks to all the faculty members, staffs in the department for extending their help and support in various ways.

I am also very grateful to all my friends. Their support and care helped me overcome many difficulties these years and made me happy all the time.

Last but not least, I would like thank my parents and grandma for their endless love, support and encouragement throughout my life. Without them my dream would not come true. Especially, I owe the deepest gratitude towards my beloved mom who is in
heaven. I was greatly encouraged by her to come to USA for the graduate study and I wish she could see my graduation. Thank you mom for giving me the strength and the power. I always feel you in my heart.
ABSTRACT

The linear mixed model is a very popular and powerful tool in many applications, such as engineering, biology and social science. Oftentimes it is of interest to make statistical inference on functions of parameters in a linear mixed model.

In Chapter 2 we constructed a canonical linear mixed-effects model with some non-linear parametric functions of interest based on motivating examples in reliability and nondestructive evaluation. Several competing procedures that can be used to construct confidence intervals for these nonlinear functions of parameters in a linear mixed model, such as likelihood, Wald, Bayesian and bootstrap procedures, are described in this chapter. Then we designed a simulation study to compare the coverage properties and computational cost for different interval estimation procedures.

The Markov chain Monte Carlo (MCMC) procedure introduced in Chapter 2 for Bayesian estimation is an efficient way to produce the credible intervals for statistical models. There is, however, always Monte Carlo error in the estimates because the MCMC procedure involves the use of random numbers. If it is of interest to determine the interval end points, which are quantified as quantile estimates, with certain degrees of repeatability, a large amount of MCMC draws may be required. It is especially true when strong autocorrelation exists in MCMC draws. In Chapter 3 we described a procedure to estimate the number of MCMC draws needed for the quantile estimates with desired precision and confidence level. We also used several examples, where different MCMC procedures are involved in, to illustrate the use of the procedure.

In Chapter 4 we introduced an R function to implement the procedures of estimating the number of draws for either the MCMC sequences or the i.i.d. sequences described in
Chapter 3. The R function takes a vector of pilot draws from either an MCMC sequence or an i.i.d. sequence, the quantile probability, the desired precision and the confidence level as the input, and returns the required number of draws. Details about how to use the function are discussed in this chapter.
CHAPTER 1. GENERAL INTRODUCTION

1.1 Background

A linear model specifies the relationship between a response variable and explanatory variables. When the linear model contains both fixed effects and random effects, it is called a linear mixed-effects model. The linear mixed-effects model is useful in a wide variety of disciplines in the physical, biological and social sciences where data is grouped according to some classification factors, such as repeated measures data, longitudinal data, and spatial data. Many contributions to the statistical literature have provided theory and applications about linear mixed-effects models. Few of these contributions, however, have considered complicated nonlinear functions of the model parameters. In many applications it is also important and interesting to make inference on some functions of parameters in a linear mixed-effects model. Among the most common methods for computing confidence intervals for a linear mixed-effects model are the likelihood method, Wald method, bootstrap method and Bayesian method.

1.2 Motivation

The research is generally motivated by some engineering problems and the detailed motivation for each project is given below.
1.2.1 Coverage Probabilities of Confidence Intervals for Functions of the Parameters of a Linear Mixed Model

Linear mixed-effects models are very useful and commonly used in many areas of application. In many situations it is important to estimate nonlinear functions of the model parameters in a linear mixed model. This project is motivated by applications in reliability and nondestructive evaluation. In this project, we formulated a canonical model with linear mixed-effects according to these applications and introduced some complicated nonlinear parametric functions of interests. We used several procedures, such as likelihood, Wald, Bayesian and parametric bootstrap procedures, to construct the confidence intervals for proposed nonlinear functions. A simulation study was conducted to study the coverage probabilities as well as computational time for these interval estimation approaches. Based on the simulation results, we made recommendations about which approach to use under different scenario.

1.2.2 The Number of MCMC Draws Needed to Compute Bayesian Credible Bounds

When the Bayesian method is used to produce the credible intervals for some functions, there is Monte Carlo error in the results. In order to obtain a reasonable degree of repeatability it may require a very long chain in the Markov chain Monte Carlo (MCMC) procedure, especially when the MCMC draws have a large amount of autocorrelation. In this project we provided an algorithm to compute the number of MCMC draws required to estimate the interval end points with certain degrees of precision. The end points of the credible interval are actually the quantiles of the empirical distribution of the MCMC draws from the marginal posterior distribution. A pilot draw with appropriate burn-in is required and is treated as approximately stationary in this algorithm for quantile estimations. Two examples were used to illustrate the use of the algorithm: one is based on a linear mixed-effects model with a Gibbs sampler procedure, and the other is based
on a generalized linear model with random walk Metropolis procedure. The algorithm introduced here can be applied to any MCMC procedure.

1.2.3 Implementation of an Algorithm for the Number of Draws Needed in Quantile Estimations

The second project introduced an algorithm to predict the MCMC draws that will be required to provide a desired degree of precision in the quantile estimate. In this paper, we developed an R function to implement the procedure of estimating the number of draws for MCMC sequences as well as i.i.d. sequences. The R function takes a vector of pilot draws from either an MCMC sequence or an i.i.d. sequence, the quantile probability, the desired precision and the confidence level as the input, and returns the required number of draws. In this paper we explained how to use the R function by several applications.

1.3 Dissertation Organization

This dissertation contains a general introduction at the beginning, three main chapters in the middle and a general conclusion at the end. Chapter 2 constructs confidence intervals for some nonlinear functions in linear mixed-effects models by using likelihood method, Wald method, Bayesian method and bootstrap method, and compares the performance of these approaches. Chapter 3 describes how to choose the number of MCMC draws needed to obtain the Bayesian credible interval endpoints with some specified amount of precision. Chapter 4 documents an R implementation of the algorithm for computing required number of draws in order to obtain the precise quantile estimates.
CHAPTER 2. COVERAGE PROBABILITIES OF CONFIDENCE INTERVALS FOR FUNCTIONS OF THE PARAMETERS OF A LINEAR MIXED EFFECTS MODEL

A paper to be submitted

Jia Liu
William Q. Meeker
Daniel J. Nordman
Department of Statistics
Iowa State University
Ames, Iowa 50010

Abstract

The linear mixed-effects model is a powerful tool that is used in various areas of application. In many applications, there is a strong interest in estimating nonlinear functions of the model parameters. We are motivated by such applications in reliability, where one needs to estimate the probability of failure, and in nondestructive evaluation, where one needs to estimate the probability of detection. The direct computation of confidence intervals for these nonlinear parametric functions is difficult, and there are several competing procedures for computing such intervals. For example, procedures may be based on likelihood ratio statistics or Wald statistics, involve simulation or bootstrap
methods, or use Bayesian approaches. We use our motivating applications to construct a canonical linear mixed-effects model and associated nonlinear parametric functions of interests. We use the canonical model to examine confidence interval procedures in a carefully designed simulation study. Under this canonical model, we compare the different confidence interval procedures with respect to coverage accuracy and computational speed for probability of detection and failure time probability parameters.

**Keywords:** Bayesian; bootstrap; credible interval; likelihood; Monte Carlo; Wald

## 2.1 Introduction

### 2.1.1 Background and Motivation

Linear mixed-effects models are widely used in many areas of application, including agriculture, biology, economics, and engineering. See, for example, Venables and Ripley (2002) and Pinheiro and Bates (2000) for general theory and application of linear mixed-effects models. In many applications, there is a strong interest in estimating nonlinear functions of the model parameters, which can be complex parametric functions of both mean effects and variance components in such models. The direct computation of confidence intervals for these nonlinear parametric functions can be difficult and computationally intensive. There are several competing methods for computing such intervals. For example, Li et al. (2014) considered estimating probability of detection (POD) in a mixed-effects model for quantifying flaw detection in titanium forging disks. A Bayesian method was used to estimate the model parameters, the mean POD, and a quantile of the POD distribution, and to find corresponding lower credible bounds. Weaver and Meeker (2013) described an accelerated repeated measures degradation model and estimated quantities of the failure-time distribution using maximum likelihood. Gao et al. (2014) used a linear mixed-effects model to describe the capability of a vibrothermography inspection system to detect cracks in fan blades. They showed how to estimate
the POD using the percentile bootstrap method to calculate confidence intervals for the POD function.

The purpose of this paper is to compare the performances of several confidence interval procedures for important nonlinear parametric functions in linear mixed-effects models that occur in reliability and nondestructive evaluation areas. As mentioned above, while various interval estimation procedures have been applied in such problems, little is known about how these methods compare for estimating complicated nonlinear parametric functions, which are commonly of interest. We motivate these functions with several examples, arising in inference about POD functions and failure-time distributions. Such examples allow us to formulate a unifying canonical linear mixed-effects model for purposes of studying estimation procedures over a meaningful class of nonlinear parametric functions. Confidence interval procedures can differ greatly in their computational burdens, so it is also of interest to quantify the coverage accuracy of different procedures against their relative computational costs. In particular, through a large simulation study, we examine interval procedures based on likelihood, normal-theory Wald, Bayesian, and bootstrap methods.

2.1.2 Related Literature

Harville (1976) constructed confidence regions for linear combinations of fixed and random effects. Kackar and Harville (1984) investigated the mean squared errors of estimators of the fixed and random effects in linear mixed models. Rekab and Tahir (2003) introduced a sequential design for estimating a nonlinear function in the simple linear regression model. Staggs (2009) considered inference on fixed effects in the mixed-effect model, based on several parametric bootstrap approaches. Das and Krishen (1999) used bootstrap methods to estimate standard errors of parameters in nonlinear mixed-effect models. A number of similar studies have been conducted for other inference problems in order to provide information on properties of different confidence interval
procedures. Two examples include the following. Jeng and Meeker (2000) compared normal-approximation procedures, likelihood ratio procedures and bootstrap procedures to compute confidence intervals for the quantiles of a Weibull distribution based on Type I censored data. Zuo et al. (2013) conducted a simulation study on confidence interval procedures for the mean cumulative function based on recurrence event data. They compared normal approximation (Wald) procedures and bootstrap-based procedures.

### 2.1.3 Overview

The rest of this chapter is organized as follows. Section 2.2 describes the motivating applications more completely and presents the development of a canonical mixed effects model that captures the key features of these applications. Section 2.3 describes the details of the confidence interval methods used in our study regarding probability of detection and failure time distribution. Section 2.4 describes the design of the simulation experiment. Section 2.5 summarizes the results of the simulation study. Section 2.6 contains concluding remarks and suggests some extensions for further research work.

### 2.2 Development of a Canonical Model

This section outlines some important motivating applications of the mixed-effects model in reliability and nondestructive evaluation, and describes a canonical model and associated nonlinear parametric functions of interest which we will use in our study of confidence interval procedures.

#### 2.2.1 Probability of Flaw Detection in Titanium Forgings

As mentioned in Section 2.1.1, Li et al. (2014) modeled an ultrasonic signal response to describe data from a study about the detection of flaws in titanium forgings that are used in the manufacturing of aircraft engine fan disks. Their physics-based nonlinear
mixed effects model was

\[ Y = \alpha + \tau + \gamma + \epsilon \]

where

\[ \alpha = \log_{10}(\beta) + \log_{10}\left( |R| \frac{\pi w^2}{2} \left( 1 - e^{-2(x/w)^2} \right) \right) \]

with independent \( \tau \sim N(0, \sigma^2_\tau) \), \( \gamma \sim N(0, \sigma^2_\gamma) \), and \( \epsilon \sim N(0, \sigma^2_\epsilon) \). Here \( \tau \), \( \gamma \) and \( \epsilon \) represent an operator random effect, a target random effect, and a measurement error, respectively. Additionally, \( \beta \) is the scaling fitting parameter that accounts for the overall factor of the Kirchhoff approximation, \( x \) is the target radius, \( w \) is the beam size, and \( R \) is a reflectance coefficient that depends on the physical or chemical properties of the target. The mean response function \( \alpha \) is approximately linear in \( x \) when \( x \) is small relative to \( w \). We use this linear approximation in our study. In this application there were seven different targets, six different operators, and a total of 392 observations.

The probability of detection here is given by

\[ \text{POD}(x) = \Pr(Y > y_{th}) = \Phi \left( \frac{\mu_Y - y_{th}}{\sigma_{\text{total}}} \right), \]

where \( \mu_Y = \alpha \), \( \sigma^2_{\text{total}} = \sigma^2_\tau + \sigma^2_\gamma + \sigma^2_\epsilon \), and the detection threshold \( y_{th} \) was chosen such that the probability of a false alarm would be negligible.

### 2.2.2 Probability of Failure in Accelerated Repeated Measures Degradation Model

Weaver and Meeker (2013) used the degradation model

\[ Y_{ijk} = D_{ijk} + \epsilon_{ijk}, \quad i = 1, \ldots, 19; j = 1, \ldots, 5; k = 1, 2, 3 \]

to describe the increase in resistance of carbon film resistors as a function of temperature and exposure time. Here \( Y_{ijk} \) represents the observed degradation for unit \( i \) at time point \( j \) for level \( k \) of the accelerating variable, and the actual degradation path is

\[ D_{ijk} = \gamma x_k \tau_{ij} + b_0 + b_1 \tau_{ij}, \]
based on bivariate normal variables

\[
\begin{pmatrix}
b_0 \\ b_1
\end{pmatrix} \sim \text{BVN} \left[ \begin{pmatrix}
\beta_0 \\ \beta_1
\end{pmatrix}, \Sigma_b = \begin{pmatrix}
\sigma_{b0}^2 & \rho \sigma_{b0} \sigma_{b1} \\
\rho \sigma_{b0} \sigma_{b1} & \sigma_{b1}^2
\end{pmatrix} \right],
\]

which are used to represent the random intercept and slope that vary from unit to unit. Additionally, \( \tau_{ij} \) is the square root of time, \( x_k \) denotes the temperature, the coefficient \( \gamma \) describes how the degradation changes with the temperature, and \( \epsilon_{ijk} \) is an error term for each measurement. Suppose that failure is defined to occur when \( D_{ijk} \geq \mu_D \), where \( \mu_D \) is the prespecified degradation level, and \( T \) denotes the time to failure for a unit with associated degradation path \( D_{ijk} \) (i.e. amount of time until \( D_{ijk} \) reaches \( \mu_D \)). Then the failure-time cumulative distribution function is

\[
F(t; x, \tau_{ij}) = \Pr(T \leq t) = \Pr(D_{ijk} \geq \mu_D) = \Pr(b_0 + b_1 \tau_{ij} \geq \mu_D - \gamma x_k \tau_{ij}) = 1 - \Phi(\kappa)
\]

where

\[
\kappa = \frac{\mu_D - \gamma x_k \tau_{ij} - \beta_0 - \beta_1 \tau_{ij}}{\sqrt{\sigma_{b0}^2 + \tau_{ij}^2 \sigma_{b1}^2 + 2 \tau_{ij} \rho \sigma_{b0} \sigma_{b1}}}
\]

Similarly if a failure is defined by \( D_{ijk} \leq \mu_D \) then \( F(t; x) = \Phi(\kappa) \) is the associated failure-time distribution.

### 2.2.3 Cracks Detection in Fan Blades

Gao et al. (2014) considered a linear mixed-effects model to describe the detection of cracks in fan blades. They used the following model for a given function of the maximum contrast response:

\[
Y = \beta_{0,b} + \beta_{1,b} \log(v) + \beta_{2,b} \log(p) + \beta_{3,b} \log(r) + \beta_{4} \log(v) \log(p) + \beta_{5} [\log(r)]^2 + \epsilon
\]

where \( b \) denotes the number of cracks, \( b = 1, \ldots, 10 \), \( v \), \( p \) and \( r \) are controlled experimental factors in their study and represent vibration amplitude, sonic pulse length, and trigger force, respectively. There were three levels for each of these three factors, and
there were 32 tests for each crack. The model used by Gao et al. (2014) assumed that \( \epsilon \sim N(0, \sigma^2) \) and that

\[
\begin{pmatrix}
\beta_{0,b} \\
\beta_{1,b} \\
\beta_{2,b} \\
\beta_{3,b}
\end{pmatrix}
\sim MVN
\begin{pmatrix}
\beta_0 \\
\beta_1 \\
0 \\
0
\end{pmatrix}, \Sigma_{\beta b},
\]

where the multivariate normal (MVN) distribution for \((\beta_{0,b}, \beta_{1,b}, \beta_{2,b}, \beta_{3,b})'\) describes the crack-to-crack variability. POD is the probability that (conditional on values \(v, p, \) and \(r\)) a response \(Y\) from a thermal image sequence exceeds a threshold \(y_{th}\) for a crack taken at random from the population of cracks. That is,

\[
POD(v, p, r) = \Pr(Y > y_{th}) = 1 - \Phi \left( \frac{y_{th} - \mu(v, p, r)}{\sigma(v, p, r)} \right),
\]

where

\[
\mu(v, p, r) = \beta_0 + \beta_1 \log(v) + \beta_4 \log(v) \log(p) + \beta_5 [\log(r)]^2
\]

is the conditional mean of the response variable \(Y\), and \(\sigma^2(v, p, r) = X_i' \Sigma_{\beta b} X_i + \sigma^2_\epsilon\) is the variance of \(Y\), where \(X_i = (1, \log(v), \log(p), \log(r))'\). Note that, in the above model, \(\beta_0, \beta_1, \beta_4, \) and \(\beta_5\) represent fixed effects.

### 2.2.4 A Canonical Model

In order to study the properties of confidence interval procedures in a more generic setting, we use the following canonical model motivated by the examples in Sections 2.2.1, 2.2.2, and 2.2.3. Namely, let

\[
Y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + b_{0i} + b_{1i} z_i + \epsilon_{ij}, i = 1, \ldots, n, j = 1, \ldots, m, \tag{2.1}
\]

where \(Y_{ij}\) denotes a response for inspection \(j\) of unit \(i\), \(n\) denotes the total number of experimental units, and \(m\) represents the number of inspections within each unit. In the model (2.1), \(\beta_0, \beta_1\) and \(\beta_2\) denote fixed effects (i.e., mean) parameters associated with
fixed regressors $x_{1i}$ and $x_{2ij}$ ($x_{1i}$ only depends on the unit, while $x_{2ij}$ is determined by
the unit and inspection within each unit). Additionally, $b_{0i}$ and $b_{1i}$ represent normally
distributed random intercept and slope effects corresponding to unit $i$, where $z_i$ is a ran-
dom regressor associated with $b_{1i}$ and we assume $z_i = x_{1i}$. In particular, the distribution
of the random intercept and slope effects is given by

$$\left( \begin{array}{c} b_{0i} \\ b_{1i} \end{array} \right) \sim \text{BVN} \left[ \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \Sigma_b = \left( \begin{array}{cc} \sigma_0^2 & \rho \sigma_0 \sigma_1 \\ \rho \sigma_0 \sigma_1 & \sigma_1^2 \end{array} \right) \right],$$

where $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ is a normal error term for unit $i$ at inspection $j$. It is assumed
that $[(b_{0i}, b_{1i})', i = 1, \ldots, n]$ are i.i.d. random vectors, $\{(\epsilon_{ij} : i = 1, \ldots, n; j = 1, \ldots, m)\}$
are i.i.d. random variables and the two collections are independent. The variance com-
ponents in the model are then parameterized as $\sigma_0^2$, $\sigma_1^2$, $\rho$, $\sigma_\epsilon^2$, corresponding to the
variances of $b_{0i}$, $b_{1i}$, the correlation between $b_{0i}$ and $b_{1i}$, and the variance of the error
term $\epsilon_{ij}$. Hence, the data from the canonical model correspond to $[(Y_{ij}, x_{1i}, x_{2ij}, z_i) : i = 1, \ldots, n; j = 1, \ldots, m]$.

An equivalent expression for the $m$-dimensional response vector $Y_i = (Y_{i1}, \ldots, Y_{im})'$
for the unit $i$ is

$$Y_i = X_i \beta + Z_i b_i + \epsilon_i, \quad i = 1, \ldots, n, \quad b_i \sim N(0, \Sigma_b), \quad \epsilon_i \sim N(0, \sigma_\epsilon^2 I), \quad (2.2)$$

where $b_i = (b_{0i}, b_{1i})'$ is the vector of random effects, $\beta = (\beta_0, \beta_1, \beta_2)'$ is the vector of fixed
effects, $X_i = (1 \ x_{1i} \ x_{2ij})_{m \times 3}$ is the design matrix for fixed effects, $Z_i = (1 \ z_i)_{m \times 2}$
is the design matrix for random intercept/slope effects, and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{im})'$. The
parameters of interest are $\theta = (\beta_0, \beta_1, \beta_2, \sigma_0^2, \sigma_1^2, \rho, \sigma_\epsilon)$, and a generic parametric function
of interest is denoted by $g(\theta)$. In terms of the canonical model (2.2), a failure-time
CDF, which is analogous to the parametric function of interest in the degradation model
reliability application (c.f. Section 2.2.2) can be expressed as
\[ g_1(\theta; x_1, x_2) = \Pr(T \leq t) = 1 - \Pr(b_0 + b_1 x_1 \leq \mu_D - \beta_0 - \beta_1 x_1 - \beta_2 x_2) \] (2.3)
where
\[ \kappa_1 = \frac{\mu_D - \beta_0 - \beta_1 x_1 - \beta_2 x_2}{\sqrt{\sigma^2 b_0 + x_1^2 \sigma^2 b_1 + 2x_1 \rho \sigma b_0 \sigma b_1}}. \]
\(\mu_D\) denotes a given degradation level, and \(x_1\) and \(x_2\) are some given values for regressors \(x_{1i}\) and \(x_{2ij}\). To mimic the nondestructive evaluation examples (c.f. Sections 2.2.1 and 2.2.3), a POD function in terms of the canonical model can be written as
\[ g_2(\theta; x_1, x_2) = \Pr(Y > y_{th}) = 1 - \Phi(\kappa_2), \] (2.4)
where
\[ \kappa_2 = \frac{y_{th} - \beta_0 - \beta_1 x_1 - \beta_2 x_2}{\sqrt{\sigma^2 b_0 + x_1^2 \sigma^2 b_1 + 2x_1 \rho \sigma b_0 \sigma b_1 + \sigma^2 \epsilon}}. \]
\(y_{th}\) denotes a prespecified detection threshold, and \(x_1\) and \(x_2\) are some given values for regressors \(x_{1i}\) and \(x_{2ij}\).

We next describe confidence interval methods for estimating the parametric functions \(g_1(\theta)\) and \(g_2(\theta)\), which will be evaluated by simulations in Sections 2.4 and 2.5 under the canonical model (2.2).

### 2.3 Confidence Interval Methods

This section describes several competing procedures that can be used to construct confidence intervals for functions of the model parameters, as described in Section 2.2.4.

#### 2.3.1 Likelihood Confidence Interval Procedures

Likelihood-based confidence interval procedures generally perform better than the more commonly used intervals based on the Wald statistics. The log-likelihood function
for the canonical model (2.2) is

\[ L = C - \frac{1}{2} \sum_{i=1}^{n} \log |\Sigma_i| - \frac{1}{2} \sum_{i=1}^{n} (Y_i - X_i\beta)'\Sigma_i^{-1}(Y_i - X_i\beta) \]

where

\[ \Sigma_i = \text{Var}(X_i\beta + Z_i b_i + \epsilon_i) = Z_i \Sigma b' Z_i + \sigma^2 \epsilon I_i, \]

and \( C \) is a constant that does not depend on any of the model parameters. The random effects covariance matrix \( \Sigma_b \) can be written as \( \sigma^2 \epsilon D_q \), where \( D_q^{-1} = \Delta_q' \Delta_q \), \( q = 2 \), and we take \( \Delta_q \) to be the transpose of the inverse of the Cholesky decomposition of \( D_q \). Following this approach, Bates and Pinheiro (1998) showed that the likelihood function can be written as

\[ L(\beta, D_q, \sigma^2 \epsilon | y) = \prod_{i=1}^{n} p(Y_i | \beta, D_q, \sigma^2 \epsilon) \]

\[ = \prod_{i=1}^{n} \int p(Y_i | b_i, \beta, \sigma^2 \epsilon)p(b_i | D_q, \sigma^2 \epsilon)db_i \]

\[ = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2 \epsilon} |D_q|} \int \exp\left[ -\frac{1}{2\sigma^2 \epsilon} \left( \|Y_i - X_i\beta - Z_i b_i\|^2 + b_i'D_q^{-1}b_i \right) \right] db_i. \]

The expression \( \|Y_i - X_i\beta - Z_i b_i\|^2 + b_i'D_q^{-1}b_i \) can be written as

\[ \|Y_i - X_i\beta - Z_i b_i\|^2 + \|0 - 0\beta - \Delta_q b_i\|^2 = \|\tilde{Y}_i - \tilde{X}_i\beta - \tilde{Z}_i b_i\|^2, \]

where

\[ \tilde{Y}_i = \begin{pmatrix} Y_i \\ 0 \end{pmatrix}, \tilde{X}_i = \begin{pmatrix} X_i \\ 0 \end{pmatrix}, \tilde{Z}_i = \begin{pmatrix} Z_i \\ \Delta_q \end{pmatrix}. \]

Again, as in Bates and Pinheiro (1998), the QR decomposition on \( \tilde{Z}_i \) gives \( \tilde{Z}_i = Q(i) \begin{pmatrix} R_{11(i)} \\ 0 \end{pmatrix} \), where \( Q(i) \) is a \((n + q) \times (n + q)\) orthogonal matrix and \( R_{11(i)} \) is an upper-triangular \( q \times q \) matrix. Based on the properties of orthogonal matrices

\[ \|\tilde{Y}_i - \tilde{X}_i\beta - \tilde{Z}_i b_i\|^2 = \|Q(i)'(\tilde{Y}_i - \tilde{X}_i\beta - \tilde{Z}_i b_i)\|^2 \]

\[ = \|c_{1(i)} - R_{10(i)}\beta - R_{11(i)}b_i\|^2 + \|c_{0(i)} - R_{00(i)}\beta\|^2 \]
where \[
\begin{pmatrix}
R_{10(i)} \\
R_{00(i)}
\end{pmatrix}
= Q'_{(i)} \tilde{X}_i 
\text{ and } 
\begin{pmatrix}
c_{1(i)} \\
c_{0(i)}
\end{pmatrix}
= Q'_{(i)} \tilde{Y}_i.
\]
Again according to the orthogonal-triangular decomposition
\[
\begin{bmatrix}
R_{00(1)} & c_{0(1)} \\
\vdots & \vdots \\
R_{00(m)} & c_{0(m)}
\end{bmatrix}
= Q_0
\begin{bmatrix}
R_{00} & c_0 \\
0 & c_{-1}
\end{bmatrix}
\]
and along with \(1/\sqrt{|D_q|} = \text{abs} |\Delta_q|\), the likelihood function becomes
\[
L(\beta, D_q, \sigma^2_\epsilon | y) = (2\pi \sigma^2_\epsilon)^{-N/2} \exp \left( \frac{\|c_{-1}\|^2 + \|c_0 - R_{00} \beta\|^2}{-2\sigma^2_\epsilon} \right) \prod_{i=1}^{n} \text{abs} \left( \frac{|\Delta_q|}{R_{11(i)}} \right).
\] (2.5)

To obtain a likelihood-based confidence interval for \(g(\theta)\) we reparameterize, replacing \(\beta_0\) with \(g(\theta)\) giving the parameter vector \(\theta^\dagger = (g(\theta), \beta_1, \beta_2, \sigma^2_0, \sigma^2_1, \rho, \sigma_\epsilon)\). That is,
\[
\beta_0 = \mu_D - \beta_1 x_1 - \beta_2 x_2 - \Phi^{-1}(1 - g_1(\theta)) \sqrt{\sigma^2_{b_0} + x_1^2 \sigma^2_{b_1} + 2x_1 \rho \sigma_{b_0} \sigma_{b_1}},
\]
where \(g_1(\theta)\) refers to the equation (2.3), or
\[
\beta_0 = y_{th} - \beta_1 x_1 - \beta_2 x_2 - \Phi^{-1}(1 - g_2(\theta)) \sqrt{\sigma^2_{b_0} + x_1^2 \sigma^2_{b_1} + 2x_1 \rho \sigma_{b_0} \sigma_{b_1} + \sigma^2_\epsilon},
\]
where \(g_2(\theta)\) refers to the equation (2.4), which is substituted into the function (2.5) to get the likelihood function in terms of \(\theta^\dagger\).

Let \(\hat{\theta}^\dagger\) denote ML estimate of \(\theta^\dagger\) and let \(\theta^\dagger_{-1} = (\beta_1, \beta_2, \sigma^2_0, \sigma^2_1, \rho, \sigma_\epsilon)\). Then the profile likelihood for \(g(\theta)\) is
\[
R(g(\theta)) = \max_{\theta^\dagger_{-1}} \left[ \frac{L(g(\theta), \theta^\dagger_{-1})}{L(\hat{\theta}^\dagger)} \right].
\]
Under the null hypothesis, the profile likelihood ratio statistic \(-2 \log[R(g(\theta))]\) approximately follows a chi-square distribution with 1 degree of freedom. Thus an approximate \(100(1-\alpha)\)% likelihood-based confidence interval for \(g(\theta)\) is
\[
\{ g(\theta) : R(g(\theta)) > \exp(-\chi^2_{1,1-\alpha}/2) \}. 
\]
2.3.2 Wald Confidence Interval Procedures

Confidence interval procedures based on a Wald statistic are widely used because they are computationally simple. In general a Wald procedure can be viewed as being based on a quadratic approximation to the log likelihood. Meeker and Escobar (1995) described the relationship between confidence regions/intervals based on likelihood ratio and Wald statistics and the advantages of likelihood-based methods. The Wald statistic for a scalar function of the parameters is

\[
\frac{g(\hat{\theta}) - g(\theta)}{se_g(\hat{\theta})},
\]

where \( \hat{\theta} \) is the MLE of \( \theta \) and \( se_g(\hat{\theta}) \) denotes the standard error of \( g(\hat{\theta}) \), which approximately follows a normal distribution in large samples. The Fisher information, which is the negative expectation of the Hessian matrix, can be used to obtain the standard error of the MLE \( g(\hat{\theta}) \) of \( g(\theta) \), or an estimate of this standard error.

The Hessian Matrix, \( H \), is given by

\[
H = \begin{pmatrix}
H_{\beta\beta} & H_{\beta\theta} \\
H_{\theta\beta} & H_{\theta\theta}
\end{pmatrix} = \begin{pmatrix}
\frac{\partial^2 l}{\partial \beta \partial \beta} & \frac{\partial^2 l}{\partial \beta \partial \theta} \\
\frac{\partial^2 l}{\partial \theta \partial \beta} & \frac{\partial^2 l}{\partial \theta \partial \theta}
\end{pmatrix}
\]

where \( l \) is the log-likelihood, i.e the logarithm of (2.5), \( \beta = (\beta_0, \beta_1, \beta_2)' \) is the vector of the fixed effect parameters, and \( \vartheta=(\sigma_0, \sigma_1, \rho, \sigma_\epsilon) \) is the vector of variance component parameters. Jennrich and Schluchter (1986) gave the following expressions for the elements of \( H \) for a linear mixed effects model:

\[
H_{\beta\beta} = -\sum_{i=1}^n X_i' \Sigma_i^{-1} X_i;
\]

\[
[H_{\beta\theta}]_{kj} = [H_{\theta\beta}]_{jk} = -\sum_{i=1}^n x_{ij}' \Sigma_i^{-1} \hat{\Sigma}_{ik} \Sigma_i^{-1} e_i, \quad j = 1, \ldots, 3; k = 1, \ldots, 4;
\]

\[
[H_{\theta\theta}]_{kr} = - \frac{1}{2} \sum_{i=1}^n \text{tr}(\Sigma_i^{-1} \hat{\Sigma}_{ik} \Sigma_i^{-1} (2e_i e_i' - \Sigma_i) \Sigma_i^{-1} \hat{\Sigma}_{ir})
\]

\[
+ \frac{1}{2} \sum_{i=1}^n \text{tr}(\Sigma_i^{-1} (e_i e_i' - \Sigma_i) \Sigma_i^{-1} \hat{\Sigma}_{i,k} \hat{\Sigma}_{r}), \quad k, r = 1, \ldots, 4.
\]
Here \( e_i = Y_i - X_i\beta \), \( x_{ij} \) is column \( j \) of \( X_i \), \( \hat{\Sigma}_{ik} = \partial \Sigma_i / \partial \theta_k \), and \( \hat{\Sigma}_{i,kr} = \partial^2 \Sigma_i / \partial \theta_k \partial \theta_r \).

The Fisher information matrix is the expected value of the negative Hessian matrix

\[
I(\theta) = E(-H) = \begin{pmatrix}
\sum_{i=1}^n X_i'\Sigma^{-1}_i X_i & 0 \\
0 & \sum_{i=1}^n T_i
\end{pmatrix},
\]

where \( T_i \) is a \( 4 \times 4 \) symmetric matrix with elements

\[
T_{jk}^i = \frac{1}{2} \text{tr}(\Sigma_i^{-1}\hat{\Sigma}_{ij}\Sigma_i^{-1}\hat{\Sigma}_{ik}), \quad j = 1, \ldots, 4; \quad k = 1, \ldots, 4.
\]

The observed information matrix, \( \hat{I}(\theta) \), is given by

\[
\hat{I}(\theta) = -\hat{H},
\]

where \( \hat{H} \) is \( H \) evaluated at the ML estimator \( \hat{\theta} \) of \( \theta \).

To obtain Wald confidence intervals for \( g(\theta) \), we first obtain a Wald interval for \( \kappa \) representing \( \kappa_1 \) in (2.3) for which \( g_1(\theta) = 1 - \Phi(\kappa_1) \) or \( \kappa_2 \) in (2.4) for which \( g_2(\theta) = 1 - \Phi(\kappa_2) \), in order to avoid the confidence interval endpoints falling outside the parameter settings. Then by the delta method, estimates of the standard error of \( \hat{\kappa} \), the MLE of \( \kappa \), are given by

\[
\hat{se}_{1,\hat{\kappa}} = \left[ \left( \frac{\partial \kappa}{\partial \theta} \right)^T E(-H)^{-1} \left( \frac{\partial \kappa}{\partial \theta} \right) \right]^{\frac{1}{2}} \quad (2.6)
\]

based on the expected information matrix and

\[
\hat{se}_{2,\hat{\kappa}} = \left[ \left( \frac{\partial \kappa}{\partial \theta} \right)^T (-H)^{-1} \left( \frac{\partial \kappa}{\partial \theta} \right) \right]^{\frac{1}{2}} \quad (2.7)
\]

based on the observed information matrix. Then a \( 100(1-\alpha)\% \) Wald confidence interval for \( \kappa \) based on the expected information matrix is given by

\[
[\bar{\kappa}, \tilde{\kappa}] = \hat{\kappa} \pm z_{1-\alpha/2} \hat{se}_{1,\hat{\kappa}},
\]

and \( 100(1-\alpha)\% \) Wald confidence interval for \( \kappa \) based on the observed information matrix is given by

\[
[\bar{\kappa}, \tilde{\kappa}] = \hat{\kappa} \pm z_{1-\alpha/2} \hat{se}_{2,\hat{\kappa}},
\]
where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. Then the Wald confidence interval for $g(\theta)$ is $[1-\Phi(\tilde{\kappa}), 1-\Phi(\kappa)]$, found by applying the decreasing monotone transformation $1-\Phi(\cdot)$ to the endpoints of $[\tilde{\kappa}, \kappa]$, which is the Wald confidence interval for $\kappa$.

### 2.3.3 Bootstrap Confidence Interval Procedures

The likelihood and Wald procedures described in Sections 2.3.1 and 2.3.2, respectively, may be adequate when the sample size is relatively large. The bootstrap method, however, can be expected to provide more accurate approximate confidence intervals. Of possible bootstrap methods, we will evaluate the simple percentile method, the bias-corrected (BC) percentile method, the bias-corrected and accelerated (BCa) method, and the bootstrap-$t$ method to construct a confidence interval for $g(\theta)$. The simple percentile bootstrap procedure in Efron (1981, 1982) is transformation invariant and easy to implement. This simple method, however, is only first-order accurate and may result in a less satisfactory coverage probability. The validity of this simple method depends on the implicit existence of a transformation that symmetrizes the distribution of the estimator, and such a transformation may not exist. The BC percentile procedure by Efron (1981, 1982) and the BCa percentile procedure by Efron (1987) are expected to have better coverage probability than the simple percentile procedure. The parametric bootstrap-$t$ confidence interval procedure by Efron (1981, 1982) also can be expected to have properties that are better than the corresponding Wald procedures. Both the bootstrap-$t$ and BCa procedures are second-order accurate under some regularity assumptions. The bootstrap-$t$ method, like the Wald method, requires specification of a transformation.

The idea of the parametric bootstrap is to simulate samples

$$
\begin{pmatrix}
  b_{0i}^* \\
  b_{1i}^*
\end{pmatrix}
; \quad \{\epsilon_{ij}^*\}_{i=1,...,n; j=1,...,m}
$$
from the fitted parametric model using the ML estimate $\hat{\theta}$ in place of the true parameter $\theta$ and to calculate the bootstrap estimates $\theta^*$ and $\kappa^*$ from the simulated data. This procedure is repeated $B$ times, giving $\theta^*_1, \ldots, \theta^*_B; \kappa^*_1, \ldots, \kappa^*_B$. The 100(1-$\alpha$)% simple percentile bootstrap confidence interval for $\kappa$ is thus

$$[\kappa, \tilde{\kappa}] = [\kappa^*_{\alpha/2}, \kappa^*_1 - \alpha/2],$$

where $\kappa^*_q$ generically denotes the $q$ percentile of the bootstrap realizations $\kappa^*_1, \ldots, \kappa^*_B$. If the bootstrap distribution is symmetric and centered on the ML estimate, the simple percentile method will perform well. Otherwise, if the distribution is non-symmetric or bias exits, the simple percentile method will be less appropriate. The BC method adjusts the bias in the bootstrap distribution by the constant $\hat{z}_0 = \Phi^{-1} \left( \#(\hat{\kappa}^* < \hat{\kappa})/B \right)$, where $\#(\hat{\kappa}^* < \hat{\kappa})/B$ denotes the proportion of the $B$ bootstrap samples for which the bootstrap estimate $\hat{\kappa}^*$ is less than the MLE $\hat{\kappa}$ from the original data. Then the 100(1-$\alpha$)% BC confidence interval for $\kappa$ is

$$[\kappa, \tilde{\kappa}] = [\kappa^*_\alpha_1, \kappa^*_\alpha_2],$$

where the adjusted quantiles are

$$\alpha_1 = \Phi(2\hat{z}_0 - z_{1-\alpha/2}); \quad \alpha_2 = \Phi(2\hat{z}_0 + z_{1-\alpha/2}).$$

The BCa method additionally adjusts skewness by using an acceleration constant $a$. There are various ways to approximate the constant $a$ and one of the easiest to implement is the delete-one-observation at a time jackknife estimate of skewness

$$\hat{a} = \frac{\sum_{i=1}^{n} (\hat{\kappa}(i) - \hat{\kappa}(\cdot))^3}{6(\sum_{i=1}^{n} (\hat{\kappa}(i) - \hat{\kappa}(\cdot))^2)^{3/2}},$$

where $\hat{\kappa}(i)$ is estimated from the original sample with the $i$th point $X_i$ deleted and $\hat{\kappa}(\cdot) = \sum_{i=1}^{n} \hat{\kappa}(i)/n$. Thus, the 100(1-$\alpha$)% BCa interval for $\kappa$ is

$$[\kappa, \tilde{\kappa}] = [\kappa^*_\alpha_1, \kappa^*_\alpha_2].$$
where

\[ \tilde{\alpha}_1 = \Phi \left( \hat{z}_0 + \frac{\hat{z}_0 - z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 - z_{1-\alpha/2})} \right), \]

\[ \tilde{\alpha}_2 = \Phi \left( \hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 - z_{1-\alpha/2})} \right). \]

The 100(1-\alpha)\% bootstrap-\(t\) confidence interval for \(\kappa\) based on the expected information matrix is

\[ [\hat{\kappa}, \tilde{\kappa}] = [\hat{\kappa} - \hat{s}\hat{e}_{1,\hat{\kappa}}t^*_{1-\alpha/2}, \hat{\kappa} + \hat{s}\hat{e}_{1,\hat{\kappa}}t^*_{\alpha/2}]. \]

Here \(t^*_{\alpha/2}\) and \(t^*_{1-\alpha/2}\) are \(\alpha/2\) and \((1-\alpha/2)\) quantiles of \(t^*_b = (\kappa^*_b - \hat{\kappa})/se_{1,\kappa^*_b}, b = 1, \ldots, B,\) and \(\hat{s}\hat{e}_{1,\hat{\kappa}}\) and \(se_{1,\kappa^*_b}\) are derived from the equation (2.6), the first being based on ML estimates and the second being based on bootstrap estimates. The bootstrap-\(t\) confidence interval based on the observed information matrix is derived in a similar way by using (2.7).

For any bootstrap procedure, a confidence interval \([\hat{\kappa}, \tilde{\kappa}]\) for \(\kappa\) is then translated into an interval for \(g(\theta)\) as \([1 - \Phi(\tilde{\kappa}), 1 - \Phi(\hat{\kappa})]\), similar to the normal-theory Wald interval of Section 2.3.2.

### 2.3.4 Bayesian Credible Interval Procedures

In many other areas of application, Bayesian credible interval procedures, with an appropriate choice of diffuse prior distributions, have been shown generally to have excellent frequentist coverage properties [c.f. Kass and Wasserman (1996)]. Agresti and Min (2005) suggested that the use of diffuse priors to provide good frequentist performance when Bayesian methods are used. Browne and Draper (2006) used both Bayesian and likelihood-based methods to fit multilevel models. They found that both methods lead to approximately unbiased point estimates when fitting two-level variance-components models, and Bayesian methods with diffuse prior distributions provide well-calibrated point and interval estimates in the three-level random-effect logistic regression models. It is of interest to see if this result holds also for nonlinear functions of the parameters.
$g(\theta)$ in linear mixed-effects models. In the canonical model, we use diffuse conjugate priors for $\beta$, $\Sigma_b$, and $\sigma_\epsilon$ as follows:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \sim \text{MVN}(\mu_0, \Lambda_0),$$

$$\Sigma_b^{-1} \sim \text{Wishart}(S_0, \eta_0),$$

$$\sigma_\epsilon^2 \sim \text{inverse-gamma}(\alpha_0, \nu_0),$$

where $(\mu_0, \Lambda_0, \eta_0, S_0, \nu_0, \alpha_0)$ is a set of initial values. To make the prior diffuse, we set

$$\Lambda_0 = \begin{pmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 100 \end{pmatrix}, S_0 = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix},$$

and $\alpha_0 = \nu_0 = 0.001$. For the Wishart distribution, $\eta_0$ is the parameter that denotes the degrees of freedom and must be larger $q - 1$, which is 1 here, to ensure the prior distribution is proper. And $\eta_0$ can be thought of as a ”prior sample size”, so we chose a small value for $\eta_0$, which is 2, in order to obtain a less informative prior. Additionally, we set the prior of $\beta$ at the ML estimates $\mu_0 = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)'$.

To perform Bayesian inference, we use Gibbs sampling to obtain samples from the posterior distribution based on full conditional posterior distributions as follows. The full conditional distribution for $\beta$ is:

$$p(\beta|.) \sim \text{MVN}(\mu_m, \Lambda_m),$$

where

$$\Lambda_m = \left( \sum_{i=1}^n X_i'(\sigma^2 I)^{-1} X_i + \Lambda_0^{-1} \right)^{-1},$$

$$\mu_m = \Lambda_m \left( \sum_{i=1}^n X_i'(\sigma^2 I)^{-1} (Y_i - Z_i b_i) + \Lambda_0^{-1} \mu_0 \right).$$
The full conditional distribution for \( b \) is:

\[
p(b|.) \sim \text{MVN}(\mu_b, \Sigma_b), \tag{2.9}
\]

where

\[
\Sigma_{b_i} = (Z'_i(\sigma^2 \epsilon I)^{-1}Z_i + \Sigma^{-1}_b)^{-1},
\]
\[
\mu_{b_i} = \Sigma_{b_i}[Z'_i(\sigma^2 \epsilon I)^{-1}(Y_i - X_i)].
\]

The full conditional distribution for \( \Sigma_b \) is:

\[
p(\Sigma_b|.) \sim \text{inverse-Wishart}(\eta_0 + m, (S_0 + S_b)^{-1}), \tag{2.10}
\]

where

\[
S_b = \sum_{i=1}^{n} b'_i b_i.
\]

And the full conditional distribution for \( \sigma_\epsilon \) is:

\[
p(\sigma^2_\epsilon|.) \sim \text{inverse-gamma}\left(\alpha_0 + \frac{n}{2}, \nu_0 + \frac{SSR}{2}\right), \tag{2.11}
\]

where

\[
SSR = \sum_{i=1}^{n} \sum_{j=1}^{m} (Y_{ij} - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2ij} - b_{0i} - b_{1i} z_i)^2.
\]

A Monte Carlo Markov chain (MCMC) sampler of posterior values of \( \tilde{\theta} = (\tilde{\beta}, \tilde{\Sigma}_b, \tilde{\sigma}_\epsilon) \) can be implemented using, for example, OpenBUGS. We initiated the chain at ML estimators for each variable in \((\beta, \Sigma_b, \sigma_\epsilon)\). In our simulation, each MCMC sample \( \tilde{\theta}_j \) was sequentially drawn from the full conditional distributions (2.8), (2.9), (2.10), and (2.11) based on iteratively updated values. After removing samples from a sufficiently long burn-in period, we obtained a sequence of MCMC samples, \( \tilde{\theta}_1, \ldots, \tilde{\theta}_S \), of length \( S \). Then posterior evaluations of the functions \( g(\tilde{\theta}_1), \ldots, g(\tilde{\theta}_S) \) were computed. A 100(1-\(\alpha\))% Bayesian credible interval for \( g(\theta) \) was obtained from the \( \alpha/2 \) and \((1 - \alpha/2) \) sample quantiles of the posterior evaluations of \( g(\tilde{\theta}) \), denoted by

\[
[g(\tilde{\theta})_{\alpha/2}, g(\tilde{\theta})_{1-\alpha/2}].
\]
2.4 Simulation Experiment

2.4.1 Design of the Simulation Experiment

There are several factors that can be expected to have effect on the coverage properties and computational time of an interval estimation procedure for the nonlinear functions $g(\theta)$ of interest. In this section, we describe the design of a simulation experiment to compare the four types of interval procedures described in Section 2.3. Important simulation considerations are outlined in the following.

1. The number of experimental units is denoted by $n$ and the number of inspections within a unit is denoted by $m$. In order to examine the effects of these two factors, we considered all combinations of levels $n = (15, 30, 60)$ and $m = (5, 15, 30)$.

2. The values in the parameter vector $\theta = (\beta_0, \beta_1, \beta_2, \sigma_0^2, \sigma_1^2, \sigma_2^2, \rho, \sigma_\epsilon)$ are chosen based on different levels of the signal-to-noise ratios $\beta_i/\sigma_i, i = 0, 1$ and the variance ratios $\sigma_i^2/\sigma_\epsilon^2, i = 0, 1$. We set $\sigma_\epsilon = 1$, $\rho = 0$, $\sigma_0 = \sigma_1$, and $\beta_0 = \beta_1 = \beta_2$. Other values of $\rho$, given by 0.3 and 0.8, were also tried and similar results were obtained. Levels of the four previous ratios were taken to be 1 or 4 to indicate low or high noise/variance in the data generation. The resulting values of $\sigma_i$ and $\beta_i, i = 0, 1$, are listed in Table 2.1.

<table>
<thead>
<tr>
<th>$\sigma_i^2/\sigma_\epsilon^2$</th>
<th>$\beta_i/\sigma_i$</th>
<th>$\sigma_i^2/\sigma_\epsilon^2$</th>
<th>$\beta_i/\sigma_i$</th>
<th>$\sigma_i^2/\sigma_\epsilon^2$</th>
<th>$\beta_i/\sigma_i$</th>
<th>$\sigma_i^2/\sigma_\epsilon^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\beta_i/\sigma_i = 1$</td>
<td>1</td>
<td>$\beta_i/\sigma_i = 4$</td>
<td>4</td>
<td>$\beta_i/\sigma_i = 1$</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>16</td>
</tr>
</tbody>
</table>

In the canonical model (2.1) the values of $x_{1i} = z_i$ were chosen to be $n$ equally spaced values between $-1$ and 1, and the values of $x_{2ij}$ were chosen to be $m$ equally spaced values between $-1$ and 1. Because larger values of probability of detection (POD) are
of primary interest while small values of the failure time CDF \( F(t; x) \) are usually of primary interest, we chose values of \( x_1, x_2, \mu_D \) and \( y_{th} \) for evaluations such that \( POD = 0.9 \) and \( F(t; x) = 0.1 \). Then after setting \( x_1 = 0.05 \) and \( x_2 = 0.1 \), the needed values of \( \mu_D \) and \( y_{th} \) were determined as listed in Table 2.2 according to the different values of \( \theta \).

<table>
<thead>
<tr>
<th>POD(( x ))</th>
<th>( \beta_i = 1 )</th>
<th>( \beta_i = 2 )</th>
<th>( \beta_i = 4 )</th>
<th>( \beta_i = 16 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_i^2 )</td>
<td>( 1 )</td>
<td>( 4 )</td>
<td>( 1 )</td>
<td>( 4 )</td>
</tr>
<tr>
<td>( y_{th} )</td>
<td>-0.67</td>
<td>-0.57</td>
<td>2.78</td>
<td>15.52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( F(t; x) )</th>
<th>( \beta_i = 1 )</th>
<th>( \beta_i = 2 )</th>
<th>( \beta_i = 4 )</th>
<th>( \beta_i = 16 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_i^2 )</td>
<td>( 1 )</td>
<td>( 4 )</td>
<td>( 1 )</td>
<td>( 4 )</td>
</tr>
<tr>
<td>( \mu_D )</td>
<td>2.43</td>
<td>4.86</td>
<td>5.88</td>
<td>21.08</td>
</tr>
</tbody>
</table>

2.4.2 The Simulation Algorithm

We evaluated the two-sided 90% confidence interval procedures described in Section 2.3. For each procedure, combination of \((n, m)\), and configuration of noise ratios, the coverage probabilities were approximated following the steps below. The total computational time was also recorded.

1. Simulate a data set for each combination of factors from the simulation design.

2. Construct the likelihood-based confidence interval and Wald-based confidence interval described in Section 2.3.1 and Section 2.3.2.

3. Compute bootstrap samples and use these to construct the bootstrap-\( t \), simple bootstrap percentile, BC and BCa bootstrap confidence intervals described in Section 2.3.3. As Efron and Tibshirani (1994) recommended, the number of bootstrap samples used here is \( B=2,000 \).
4. Generate MCMC samples using a Gibbs sampling algorithm and construct the Bayesian credible interval described in Section 2.3.4. The length of the Markov chain was 10,000 after the first 5,000 burn-in samples were dropped. The burn-in period was determined by inspecting parameter trace plots for a number of simulated data sets.

5. If the confidence intervals obtained in Steps 2 to 4 cover the true value of $g(\theta)$ then assign an indicator a value of 1; if the lower bound of the confidence interval is larger than the true value then assign the indicator a value of 0; otherwise assign the indicator a value of 2.

6. Repeat Steps 1 to 5 for a total of 2,000 times and then compute the proportion of simulation runs for which the lower endpoint of the interval is less than the true value of $g(\theta)$ and the proportion of runs for which the upper endpoint of the interval is greater than the true value of $g(\theta)$.

2.5 Simulation Experiment Results

In our simulation study, we first simulated data using $n=15, 60$ and $m=5, 30$ and examined coverage probabilities for different noise/variance ratio settings. We found that the coverage probabilities were nearly identical for the same $\sigma_i$ values; that is, across all the procedures, coverage probabilities were the same when $\beta_i = 1$, $\sigma_i^2 = 1$ and $\beta_i = 4$, $\sigma_i^2 = 1$, and coverage probabilities were the same for $\beta_i = 2$, $\sigma_i^2 = 4$ and $\beta_i = 16$, $\sigma_i^2 = 4$. Hence, in order to provide a more efficient summarization of our results, we only present results for $\beta_i=2$, $\sigma_i^2=4$ and $\beta_i=4$, $\sigma_i^2=1$, but on all combinations of $n$ and $m$. Our goal is to compare different confidence interval procedures and to assess the effect of $n$ and $m$ on coverage probabilities and computational time.
2.5.1 Computational Time Results

With respect to computational time, the Wald procedure is the fastest among the four procedures. The likelihood and Bayesian procedures require somewhat more computation time. The bootstrap procedure requires the most computing time. The computational time for each procedure relative to Wald method under different values of $n$ and $m$ when $\beta_i = 2$ and $\sigma_i^2 = 4$ are shown in Figure 2.1. The results are similar for other variance/noise ratio settings. Of the bootstrap methods, here we only show the computing time for the bootstrap-$t$ procedure as the other bootstrap procedures require similar amount of time. When $n$ increases, the ratio of computational time for both likelihood and Bayesian procedures relative to the Wald method also increases and rates of change are positively related with values of $m$. When $n$ is fixed and $m$ increases, the ratio of computational time for the likelihood procedure does not change much while that for the Bayesian procedure does increase some. Interestingly, the sample size $n$ has little effect on the ratio of computational time for the bootstrap procedure, which is relatively large in all cases in Figure 2.1.

2.5.2 Coverage Probability Results

In evaluating the coverage probability of 90% confidence intervals, we examined both left-tail and right-tail (error) probabilities and compared them with the 5% nominal level for each tail. The Monte Carlo margin of error is approximately $2[0.05(1 - 0.05)/2000]^{1/2} = 0.0097$, or approximate 1%. We say that a procedure is accurate as long as the one-sided error probability is within ±1% error of the nominal level. Because the results based on two different parameter settings gave us similar conclusions, we will only show the most interesting and useful results from the simulation experiment when $\beta_i = 2$ and $\sigma_i^2 = 4$. Figure 2.2 shows the effect of sample size on tail error probabilities for the failure-time CDF $g_1(\theta)$, and Figure 2.3 shows the effect of sample size on error probabilities for POD $g_2(\theta)$. There are six plots within each figure, arranged in three
Figure 2.1 Relative computational time for likelihood, Bayesian, and bootstrap methods comparing with Wald method.

rows and two columns. In row 1, $m$ is 5, in row 2, $m$ is 15, and in row 3, $m$ is 30. The left-tail probabilities are presented in the plots on the left side of figures, while right tail probabilities are given on the right side of the figures. Because the coverage probabilities for the Wald and bootstrap-$t$ procedures based on the expected Hessian matrix and the observed Hessian matrix are almost exactly the same, we only present the results based on the observed Hessian matrix.

From Figures 2.2 and 2.3, we found that $n$ has the substantial effect on the performance of the four procedures, while $m$ reaches a point of diminishing returns. When $n$
increases, the tail probabilities become closer to the nominal 5% level. When \( m \) changes from 5 to 15, these probabilities become closer to the nominal level though the tail probabilities do not change much when \( m \) increases from 15 to 30. The bootstrap-t procedure and BCa bootstrap procedure have considerably better performance in coverage accuracy than other procedures and tend to be accurate even when the sample size is relatively small. The BC bootstrap, likelihood and Wald procedures have similar performances and they are not accurate for small sample sizes. In general the performance of Bayesian procedure is a little worse than the bootstrap-t and BCa bootstrap procedures, while somewhat better than BC bootstrap, likelihood and Wald procedures. The simple percentile bootstrap procedure is not an accurate procedure and, even for the large sample sizes, its tail probabilities are still quite far away from the nominal level.

In the case when \( m \) is 30, however, the right-tail probabilities with the Bayesian procedure for both the failure-time CDF and POD quantities are seen to deviate a bit from the nominal level when \( n \) increases. To avoid a potential case of improper use of a prior distribution, different set of priors were also considered. Gelman (2006) suggested noninformative uniform prior distributions on a wide range for hierarchical variance parameters when the number of experimental units is larger than 5, and half-Cauchy family when the number of experimental units is below 5. For our study since the number of units is always larger than 5 the noninformative uniform distributions, i.e. \( \sigma_\epsilon \sim \text{Uniform}(0,100) \), \( \sigma_0 \sim \text{Uniform}(0,100) \), \( \sigma_1 \sim \text{Uniform}(0,100) \), and \( \rho \sim \text{Uniform}(-1,1) \), were tried. The right-tail probabilities still deviated some from the nominal level but in the opposite direction in this case, which was consistent with the posterior distributions. For larger sample sizes, tail probabilities are less affected by different sets of diffuse prior distributions and converge to the nominal level.

As the true value for \( F(t; \mathbf{x}) \) is 0.1 and the true value for POD is 0.9 in the simulations, there is less variability in left-tail probabilities for the failure-time CDF and right-tail probabilities for the POD function, but more variability in results for right-tail
probabilities for the failure-time CDF and left-tail probabilities for the POD function, as shown in Figures 2.2 and 2.3.

2.6 Concluding Remarks and Areas for Further Research

Overall different noise/variance ratio settings have little effect on the coverage probability, while sample sizes have great impact on the coverage accuracy. Compared with other procedures, the bootstrap-\(t\) and BCa bootstrap procedures exhibit very good coverage accuracy properties. This can be attributed to the fact that, theoretically, they are second-order accurate. The disadvantage is their computational cost. The Wald procedure is the fastest to compute. When the sample size is relatively small, however, the Wald procedure is not accurate. If computational time is an issue, the Wald procedure can be a good choice. If some intensive computation is acceptable, then bootstrap-\(t\) and BCa procedures will provide the most accurate interval estimates. We do not recommend the simple percentile bootstrap procedure.

There are several potential areas for further investigation. Our study focused on probability functions in reliability and nondestructive evaluation. We would expect the similar results for other nonlinear functions of parameters in a linear mixed-effects model, but this general topic requires further attention. Additionally, the likelihood, Wald, Bayesian, and bootstrap procedures could potentially be studied for inference about complex parametric functions in nonlinear mixed effects models.
Figure 2.2  Left-tail (displayed on the left) and right-tail (displayed on the right) coverage probabilities for estimating the failure-time CDF versus the number \( n \) of units when \( \beta_i = 2 \) and \( \sigma^2_i = 4 \). Top to bottom, the rows correspond to results for \( m=5, 15, 30 \), respectively. The horizontal line indicates the nominal 5% level in each graph.
Figure 2.3 Left-tail (displayed on the left) and right-tail (displayed on the right) coverage probabilities for estimating POD function versus the number $n$ of units when $\beta_i=2$ and $\sigma^2_i=4$. Top to bottom, the rows correspond to results for $m=5$, 15, 30, respectively. The horizontal line indicates the nominal 5% level in each graph.
CHAPTER 3. THE NUMBER OF MCMC DRAWS
NEEDED TO COMPUTE BAYESIAN CREDIBLE BOUNDS

A paper to be submitted

Jia Liu
Daniel J. Nordman
William Q. Meeker
Department of Statistics
Iowa State University
Ames, Iowa 50010

Abstract

Bayesian credible bounds produced from Markov chain Monte Carlo (MCMC) procedures contain Monte Carlo error and thus may require a long chain in order to have a reasonable degree of repeatability. This is especially true when there is a substantial amount of autocorrelation in the chain realization. Repeatability would be important in some applications where it would be undesirable to report numerical values containing substantial Monte Carlo error. The endpoints of a credible interval correspond to quantiles of the empirical distribution of the MCMC draws from the marginal posterior distribution of the quantity of interest. Our goal is to provide an algorithm to choose the number of MCMC draws that will provide, with high probability, a specified amount
of precision (i.e., the number of correct significant digits) in the interval endpoints.

**Keywords:** Markov chain Monte Carlo (MCMC); nonparametric quantile estimation; precision

3.1 Introduction

3.1.1 Background and Motivation

Markov chain Monte Carlo (MCMC) methods are commonly used for Bayesian inference computations, for numerically obtaining draws (or approximating draws) from a posterior distribution of interest. Such computations are often efficient and easy to implement, even for complicated data and model combinations. Additionally, MCMC procedures are particularly helpful for sampling posterior distributions that are very complicated and sometimes of high dimension. Metropolis et al. (1953) first introduced the idea of MCMC methods for evaluating complex integrals arising from physical problems. The integrals were restated as expectations of random variables having a distribution function $f(\cdot)$ and then samples were generated from $f(\cdot)$ to estimate the expectations. Hastings (1970) generalized the method to solve statistical problems. Given a distribution function $f(\cdot)$ that needs to be evaluated, the MCMC procedure constructs a Markov chain having a stationary distribution $f(\cdot)$. After a large number of draws the chain is then used to estimate particular functions of the parameters of $f(\cdot)$. The sample draws can be used to make inferences for unknown quantities of interest, based on the data and prior distribution specification. The quantity of interest can be a parameter in a model or some function of parameters.

There are several different MCMC algorithms that can be used to obtain the sample draws. These include the Metropolis algorithm by Metropolis et al. (1953), the Metropolis-Hastings algorithm by Hastings (1970) and the Gibbs sampler algorithm described by Geman and Geman (1984) and Gelfand and Smith (1990).
A credible interval can be used to quantify the statistical uncertainty of the unknown quantity of interest. The end points of the credible interval are defined by the quantiles of the empirical distribution of the MCMC draws from the marginal posterior distribution. Because MCMC methods involve random sampling, we would not expect to obtain the same set of MCMC draws each time that we run the chain. However, one may be interested in having a certain degree of precision that would provide a specified amount of repeatability for the quantile estimates in terms of Monte Carlo error. Usually there is more variability in the tail of the marginal posterior distribution, thus a large number of MCMC draws may be needed in order to achieve a desired degree of precision for the end points of Bayesian credible intervals. This suggests the need for a method to choose the number of draws required to estimate the quantile with certain degrees of precision. The main idea of the method is to obtain a pilot stretch of MCMC draws which can be treated as an approximately stationary realization, and then apply state-of-the-art techniques for quantile estimation of a stationary, weakly dependent time process. This leads to a procedure for estimating MCMC sample sizes for Bayesian credible intervals of desired precision. The method provided in this paper can be applied to any MCMC algorithm.

3.1.2 Literature Review

A number of books and papers have described different MCMC algorithms for Bayesian computations. Gelfand and Smith (1990) reviewed and compared three sampling approaches, which are stochastic substitution, the Gibbs sampler, and the sampling-importance-resampling algorithm, for different model structures in applications. Geyer (1992) suggested one long run of the Markov chain and estimated variances based on MCMC output by using window estimators, batch means, and specialized Markov chain estimators. Smith and Roberts (1993) reviewed implementations of the Gibbs sampler with some examples and also briefly described other MCMC methods, such as the Metropolis-

There is also previous work that introduced various ways to obtain nonparametric confidence intervals for quantiles. Woodruff (1952) proposed a method of obtaining confidence intervals for medians and other position measures by inverting the end points of confidence intervals for the corresponding distribution function under any sampling scheme. Sitter and Wu (2001) assessed the performance of confidence intervals for quantiles by varying tail probabilities. Gilat and Hill (1996) derived distribution-free confidence intervals for quantiles from any distribution based on order statistics and i.i.d assumptions. Chen and Hall (1993) used smoothed empirical likelihood confidence intervals for quantiles and showed the procedure had good coverage properties.

Most directly relevant to our work, Raftery and Lewis (1992) proposed a method for computing the total number of MCMC draws as well as the length of burn-in period when the tail probability of the posterior distribution of a function is to be estimated within certain degree of precision. They focus on the precision for estimating a tail probability, and not the quantile itself, which is a different quantity and the scales of the two quantities (probabilities and quantiles) are difficult to relate in terms of precision. In fact, the approach of Raftery and Lewis (1992) translates an MCMC draw into binary (0-1) time series, which may be useful for estimating proportions but presents a loss of information for other quantities, such as quantiles. Our focus is on estimating a particular quantile with a specified degree of precision, which has practical interpretation for the final interval estimator itself that is often of direct interest.
3.1.3 Overview

The rest of this chapter is organized as follows. Section 3.2 summarizes the structure of MCMC algorithm output. Section 3.3 describes the details of the quantile estimation for both i.i.d. and MCMC sequences. Section 3.4 presents some examples to illustrate the use of the method. Section 3.5 contains concluding remarks and suggests some extensions for further research work.

3.2 Structure of MCMC Output

An MCMC procedure is first used to randomly generate a sequence of draws from the joint posterior distribution of the unknown parameters in a model. After removal of initial 'burn-in' draws, a sequence of draws can be stored in a matrix

\[
[X] = \begin{pmatrix}
x_{11} & x_{12} & \cdots & x_{1p} \\
x_{21} & x_{22} & \cdots & x_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
x_{S1} & x_{S2} & \cdots & x_{Sp}
\end{pmatrix},
\]

where \( S \) represents the number of draws, \( p \) represents number of unknown parameters \( \theta \) in the model, and \((x_{i1}, \ldots, x_{ip})\) denotes the \( i \)th MCMC draw for \( \theta \). If additional quantities are of interest, like some function of parameters \( g(\theta) \), extra columns will be added to the output matrix with the same number of rows. Then the new output matrix would be

\[
[\tilde{X}] = \begin{pmatrix}
x_{11} & \cdots & x_{1p} & g(x)_1 \\
x_{21} & \cdots & x_{2p} & g(x)_2 \\
\vdots & \vdots & \vdots & \vdots \\
x_{S1} & \cdots & x_{Sp} & g(x)_S
\end{pmatrix}.
\]

After generating MCMC draws point estimates of \( g(\theta) \) can be obtained. One possibility is the median of the draws from the marginal posterior distribution. All \( S \) draws for each parameter are placed in ascending order such that \( g(x)_{i(1)} < g(x)_{i(2)} < \cdots <
\( g(x)_{i(S)}, \quad i = 1, \ldots, p, \) and then the posterior median of the parameter \( g(\theta) \), which is denoted by \( M_i \), is defined by

\[
M_i = \begin{cases} 
  g(x)_{i,((1+S)/2)} & \text{if } S \text{ is odd}, \\
  (g(x)_{i,(S/2)} + g(x)_{i,(S/2+1)})/2 & \text{if } S \text{ is even}.
\end{cases}
\]

In addition to the point estimate, a credible interval quantifies statistical uncertainty. A \( 100(1-\alpha)\% \) Bayesian credible interval for the parameter \( g(\theta) \) may be obtained from the \( \alpha/2 \) and \( (1-\alpha/2) \) sample quantiles of the empirical posterior distribution of MCMC draws, denoted by

\[
[g(x)_{i,(S(\alpha/2))}, \quad g(x)_{i,(S(1-\alpha/2))}].
\]

It is the precision resulting from the use of MCMC draws to obtain such intervals that interest us here and, in particular, how many MCMC draws \( S \) are needed to obtain a desired degree of precision.

After a 'burn-in' period, it is common to assume that the Markov Chain has been initialized for obtaining draws that approximately follow the stationary distribution (i.e., the posterior distribution) of the Markov chain; this is supported by the theory of Harris recurrent Markov chains; see Athreya and Lahiri (2006) Ch. 14. Then, by the transition probability structure of the MCMC algorithm, a sample \( g(x)_{i(1)}, \ldots, g(x)_{i(S)} \) may be then treated as a realization of a stationary, time dependent sequence from the marginal posterior distribution. From this, techniques of quantile estimation for such time series may then be applied for determining the number of draws \( S \) needed to achieve a desired precision in the endpoints of a Bayesian credible interval.

### 3.3 Quantile Estimation and the Number of Draws

Determination

Here we provide background for quantile estimation and a method for the number of draws determination with stationary time series. We begin with reviewing this process
for i.i.d data in Section 3.3.1 and describe the methodology of interest for time series
(i.e., for MCMC draws) in Section 3.3.2.

### 3.3.1 Quantile Estimation for i.i.d. Sequences

Let $F$ be the increasing cumulative distribution function (cdf) of the i.i.d. continuous
random variables $X_1, \ldots, X_S$ so that $F(x) = \text{Pr}(X_i \leq x), \quad i = 1, \ldots, S$. Also let $f$ be
the corresponding probability density function given by $f(x) = F'(x)$. For a particular
$0 < p < 1$, the $p$ quantile of the distribution function $F$ is denoted by $\xi_p$ and is defined
such that $F(\xi_p) = p$. The quantile $\xi_p$ can be estimated by the sample quantile $\hat{\xi}_p$ of the
distribution, which is defined as

$$
\hat{\xi}_p = \begin{cases} 
x_{(Sp)} & \text{if $Sp$ is an integer}, \\
x_{(\lfloor Sp \rfloor + 1)} & \text{otherwise},
\end{cases}
$$

(3.1)

where $\lfloor . \rfloor$ denotes the floor function and $x_{(i)}$ represents the $i$th order statistic among
$x_{(1)} < x_{(2)} < \cdots < x_{(S)}$.

We can also refer to the sample quantile $\hat{\xi}_p$ as the $p$ quantile of the empirical distri-
bution function $F_S$, which is defined as

$$
F_S = \frac{1}{S} \sum_{i=1}^{S} I(X_i \leq x),
$$

where $I(\cdot)$ is the indicator function. From Stuart and Ord (1994), it holds that

$$
\frac{\hat{\xi}_p - \xi_p}{\sqrt{p(1-p)/[Sf^2(\xi_p)]}} \xrightarrow{d} N(0, 1), \quad \text{as } S \to \infty.
$$

Hence, for large sample sizes $\hat{\xi}_p$ is approximately normally distributed with mean equal
to the population quantile $\xi_p$ and variance

$$
p(1-p)/[Sf^2(\xi_p)].
$$

(3.2)

Given a desired relative precision $d$ and the confidence level $1 - \alpha$, one may wish to
determine the number of draws $S$ so that $\Pr(|\hat{\xi}_p - \xi_p| \leq d\xi_p) = 1 - \alpha$ holds approximately.
Note here $d$ represents precision relative to the quantile $\xi_p$ of interest in estimation and $d\xi_p$ represents absolute precision. For example, if $d = 0.01$, then we are seeking the number of draws $S$ at which the estimate $\hat{\xi}_p$ matches the target quantile $\xi_p$ to two decimal place accuracy (in the case that $1 - d \leq \hat{\xi}_p/\xi_p \leq 1 + d$ with confidence $1 - \alpha$). We first run the procedure based on an initial or pilot sample of draws $\tilde{X}_1, \ldots, \tilde{X}_S$ ($\tilde{X}_1, \ldots, \tilde{X}_S$ may correspond to the first $\tilde{S}$ draws in $X_1, \ldots, X_S$), and get the estimates $\tilde{\xi}_p$ and $\tilde{f}(\tilde{\xi}_p)$ for $\xi_p$ and $f(\xi_p)$. Substituting $\tilde{\xi}_p$ and $\tilde{f}(\tilde{\xi}_p)$ in (3.2) and using normal theory, one may equate $d\tilde{\xi}_p = Z_{1-\alpha/2} \sqrt{p(1-p)/[S\tilde{f}(\tilde{\xi}_p)^2]}$, where $Z_{1-\alpha/2}$ denotes the upper $\alpha/2$ quantile of the standard normal distribution. Then the approximate total number of draws that will be needed can be estimated as

$$S = \left\lfloor \frac{Z_{1-\alpha/2} [p(1-p)]}{[d\tilde{\xi}_p \tilde{f}(\tilde{\xi}_p)]^2} \right\rfloor + 1.$$

### 3.3.2 Quantile Estimation for MCMC Sequences

MCMC draws generally have some dependence structure and may be treated as a stationary time series realization after a burn-in period (cf. Section 3.2). For stationary time processes, the degree of dependency can be quantified by the autocovariance function. Suppose $X_1, \ldots, X_S$ are random variables from a real-valued stationary time process $\{X_t\}$, having a (increasing) marginal cdf and pdf given by $F(\cdot)$ and $f(\cdot)$, and $p$ quantile $\xi_p$ satisfying $F(\xi_p) = p \in (0, 1)$. Let $\hat{\xi}_p$ denote the $p$ sample quantile from $X_1, \ldots, X_S$. 

Under mild regularity conditions on the time dependence, $\hat{\xi}_p$ is approximately normal in large samples with mean $\xi_p$ and large-sample variance given by

$$\text{Avar}(\hat{\xi}_p) = \frac{\sigma^2(p; S)}{Sf^2(\xi_p)},$$

(3.3)

where

$$\sigma^2(p; S) = S \times \text{Var} \left( \frac{1}{S} \sum_{i=1}^{S} Y_i \right) = p(1-p) + 2 \sum_{j=1}^{S-1} \left( 1 - \frac{j}{S} \right) \gamma(j),$$

for $Y_i \equiv I(X_i < \xi_p), \quad i = 1, \ldots, S$, and $\gamma(j) \equiv \text{cov}(Y_1, Y_{j+1}), \quad j = 1, \ldots, S - 1$. See Chen and Tang (2005) for distributional details about sample quantiles in dependent
sequences. Note that if random variables $X_1, \ldots, X_S$, are i.i.d., then $\{Y_i\}_{i=1}^S$ are i.i.d. Bernoulli($p$) random variables, so that (3.3) reduces to the large-sample variance of the sample quantile $\xi_p$ in the i.i.d. setting (3.2) by using $\gamma(0) = p(1 - p)$ and $\gamma(j) = 0$ for $j = 1, \ldots, S - 1$.

In order to formulate a sample size determination based on the large sample normality of $\hat{\xi}_p$, we need to estimate both the variance $\sigma^2(p; S)$ and the pdf $f(\cdot)$ in (3.3). To do so, we use kernel estimation based on the flat-top lag window of Politis and Romano (1995), which is denoted as $\lambda_T(t)$. There are various choices for the family of flat-top kernels. The simplest form shown in Politis (2003) is defined as

$$
\lambda_T(t) = \begin{cases} 
1 & 0 \leq |t| \leq 0.5 \\
2(1 - |t|) & 0.5 \leq |t| \leq 1 \\
0 & \text{otherwise.}
\end{cases}
$$

It has been shown that the flat-top kernel has good asymptotic properties and leads to fast convergence rates in variance and pdf estimations (cf. Politis 2003).

An estimated spectral density from the series $\{\hat{Y}_i \equiv I(X_i < \hat{\xi}_p)\}_{i=1}^S$ can be used to estimate $\sigma^2(p; S)$. In particular, if $\hat{\phi}(0)$ denotes the estimated spectral density of $\{Y_i\}$ at the origin, based on $\{\hat{Y}_i\}_{i=1}^S$, then $2\pi \hat{\phi}(0)$ estimates $\sigma^2(p; S)$ using that $\lim_{S \to \infty} \sigma^2(p; S) = 2\pi \phi(0)$ when $\sum_{j=0}^{\infty} |r(j)| < \infty$, where

$$
\phi(\omega) = \frac{1}{(2\pi)} \sum_{k=-\infty}^{\infty} e^{i k \omega} \hat{r}(k), \quad -\pi \leq \omega \leq \pi,
$$

denotes the spectral density function of $\{\hat{Y}_i \equiv I(X_i < \xi_p)\}_{i=1}^S$. The corresponding sample version of $\phi(\omega)$ is

$$
\hat{\phi}(\omega) = \frac{1}{(2\pi)} \sum_{k=-H}^{H} \lambda_T(k/H)e^{i k \omega} \hat{r}(k),
$$

based on the flat-top kernel $\lambda_T(\cdot)$ with a bandwidth parameter $H$ and the lag-$k$ sample autocovariance

$$
\hat{r}(k) = \frac{1}{S} \sum_{i=1}^{S-|k|} (\hat{Y}_i - \bar{Y})(\hat{Y}_{i+k} - \bar{Y}), \quad |k| < S,
$$

(3.4)
where $\bar{Y} = \sum_{i=1}^{S} \tilde{Y}_i / S$. There is a simple empirical rule that can be used to choose the kernel bandwidth. The optimal value of $H$, which is denoted by $H_{\text{opt}}$, minimizes the MSE of $\hat{\phi}(\omega)$. From Politis (2003), an estimate of $H_{\text{opt}}$ is given by $\hat{H} = 2h$, where $h$ is the smallest positive integer such that

$$\max_{k=1,\ldots,K} |\hat{\varrho}(h + k)| < c \sqrt{\log S / S},$$

for $\hat{\varrho}(k) = \hat{\varrho}(k)/\hat{\varrho}(0)$. Here $c$ is a positive constant and was recommended to be 2 and the value of $K$ was recommended to be 5.

We obtain an estimate $\hat{f}(x)$ of the marginal pdf of $\{X_t\}$ as follows, also based on the flat-top kernel. Based on an inversion of the Fourier transform, we have

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixt} Q(t) dt, \quad x \in \mathbb{R},$$

where

$$Q(t) = \int_{-\infty}^{\infty} e^{-itx} f(x) dx, \quad t \in \mathbb{R}$$

is the characteristic function of $f$. We estimate $f(x)$ by

$$\hat{f}(x) = \frac{1}{2\pi} \int_{-M}^{M} \lambda^T(t/M) e^{itx} \hat{Q}(t) dt = \frac{1}{\pi S} \sum_{j=1}^{S} g(x,x_j,M),$$

$$g(x,x_j,M) = \begin{cases} \frac{2}{M(x-x_j)^2} [\cos(M/2(x-x_j)) - \cos(M(x-x_j))] & \text{if } x \neq x_j \\ \frac{5}{8} M & \text{if } x = x_j, \end{cases}$$

where $M$ is a bandwidth and

$$\hat{Q}(t) = \frac{1}{S} \sum_{j=1}^{S} e^{-itx_j}$$

is the sample characteristic function of $\{X_t\}_{t=1}^{S}$. In Politis (2003), a bandwidth estimate is given by $\hat{M} = 2m$, where $m$ is the smallest positive real number such that

$$\max_{t \in (0,K)} |\hat{Q}(m + t)| < c \sqrt{\log S / S}.$$
Now suppose that we wish to choose the number of draws $S$ such that $\Pr(\hat{\xi}_p - \xi_p \leq d\xi_p) = 1 - \alpha$ holds for dependent draws $X_1, \ldots, X_S$. Based on a pilot set of draws $\tilde{X}_1, \ldots, \tilde{X}_S$ of length $\tilde{S}$ ($\tilde{X}_1, \ldots, \tilde{X}_S$ may correspond to the first $\tilde{S}$ draws in $X_1, \ldots, X_S$), we apply the kernel estimation above to obtain estimates $\tilde{\xi}_p$, $2\pi \tilde{\phi}(0)$ and $\hat{f}(\tilde{\xi}_p)$ of $\xi_p$, $\sigma^2(p; S)$ and $f(\xi_p)$ for substitution in the large sample variance formula (3.3). Then we equate

$$d\tilde{\xi}_p = Z_{1-\alpha/2} \sqrt{2\pi \tilde{\phi}(0) / [S \hat{f}(\tilde{\xi}_p)^2]},$$

and solve for

$$S = \left\lceil \frac{Z_{1-\alpha/2}^2 2\pi \tilde{\phi}(0)}{(d\tilde{\xi}_p \hat{f}(\tilde{\xi}_p))^2} \right\rceil + 1. \quad (3.6)$$

3.4 Application and Evaluation of the Algorithm

In this section we illustrate the use of the algorithm by choosing the number of MCMC draws in the Bayesian analysis of two data examples, involving the fit of a linear model with mixed effects and a generalized linear model with a Poisson distribution and a log link. For each example we will also evaluate the algorithm by repeating the procedure 1,000 times and checking the average relative precision and coverage.

First we need to determine a gold standard to serve as the true posterior quantile of interest in these examples. We run the MCMC procedure long enough, i.e. 50,000,000 MCMC draws, and use the empirical quantile of some posterior quantity of interest based on these 50,000,000 MCMC draws as the gold standard, or the true quantile $\xi_p$ of the marginal posterior distribution (which depends on the sample and the parametric inference problem). In order to obtain the spectral density estimates and probability density estimates, certain length of initial MCMC draws are needed. So we run an initial 10,000 iterations and discard the first 2,000 draws as burn-in. After the burn-in we assume these 8,000 initial draws are an approximately stationary realization from the marginal posterior distribution of interest. Based on these initial draws, we can estimate
the sample autocovariance function as in (3.4) and sample characteristic function in (3.5) in order to determine the number of draws $S$ needed to estimate the marginal posterior quantile $\xi_p$ of interest with a desired level of precision. That is, based on (3.6), the total number of draws needed after burn-in in the MCMC procedure can be calculated. For estimating the marginal posterior quantile of some parametric functions, the initial run with 8,000 MCMC draws is more than enough to obtain the desired precision and we then simply use the available 8,000 MCMC draws to estimate the marginal posterior quantile. However, for credible intervals with a desired precision regarding other parametric functions, more draws are required and we need to continue running the MCMC procedure to obtain the calculated number of draws. We apply the algorithm 1,000 times and the marginal posterior quantile of a corresponding parametric function is estimated each time based on actual MCMC draws used. We record the average number of MCMC draws needed as well as the average number of MCMC draws actually used over 1,000 runs. The mean absolute relative precision $|\hat{\xi}_p - \xi_p|/\xi_p$ based on actual MCMC draws used and the proportion of 1,000 simulation runs for which $|\hat{\xi}_p - \xi_p| \leq d\xi_p$ holds are also recorded.

There are several factors that can be expected to have an effect on the required number of MCMC draws.

1. The desired relative precision, denoted by $d$, is set as 0.01 and 0.005.

2. The confidence level, denoted by $1 - \alpha$, is set as 0.9 and 0.95.

3. The tail probability $p$, which is set at 0.8, 0.9, 0.95 to vary from moderate to extreme cases.

3.4.1 Application 1: Linear model with mixed effects

We consider an analysis example based on the 2002 Education Longitudinal Study (ELS) data from Hoff (2009), which involves 10th grade students from 100 different large
urban public high schools. The linear mixed-effects model used here is

$$Y_{ij} = b_{0i} + b_{1i}x_{ij} + \epsilon_{ij}, \quad j = 1, \ldots, n; \ i = 1, \ldots, n;$$

(3.7)

where $Y_{ij}$ denotes the normalized math scores for student $j$ in school $i$, $n$ denotes the total number of schools, and $n_i$ denotes the total number of students in school $i$. In (3.7), $b_{0i}$ and $b_{1i}$ represent normally distributed random intercept and slope effects corresponding to school $i$. The variable $x_{ij}$ is a regressor associated with $b_{1i}$, representing the centered socioeconomic status (SES) scores of $i$th student’s family in school $j$. In particular, the distribution of random intercept and slope effects is given by

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim \text{MVN} \left( \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \Sigma_b = \begin{pmatrix} \sigma_0^2 & \rho \sigma_0 \sigma_1 \\ \rho \sigma_0 \sigma_1 & \sigma_1^2 \end{pmatrix} \right).$$

Here $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ is a normal error term for student $j$ in school $i$. It is assumed that $\{(b_{0i}, b_{1i}), i = 1, \ldots, n\}$ are i.i.d. random vectors, $\{\epsilon_{ij}: i = 1, \ldots, n; j = 1, \ldots, m\}$ are i.i.d. random variables and the two collections are independent.

A Gibbs sampling algorithm with conjugate priors is used in this example. The priors for $\beta$, $\Sigma_b$, and $\sigma_\epsilon$ are as follows:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim \text{MVN}(\mu_0, \Lambda_0),$$

$$\Sigma_b^{-1} \sim \text{Wishart}(S_0, \eta_0),$$

$$\sigma_\epsilon^2 \sim \text{inverse-gamma}(\alpha_0, \nu_0),$$

The details about the model and the initial values ($\mu_0, \Lambda_0, \eta_0, S_0, \nu_0, \alpha_0$) can be found in Hoff (2009, p. 200).

We consider estimating the marginal posterior quantile of two different parametric functions. The first is $g_1 = \beta_1$ which can be interpreted as the mean of the slopes that vary from school to school. The second, more complicated function is the ratio of standard deviations for the slope and error terms $g_2 = \sigma_1/\sigma_\epsilon$. 
Table 3.1 shows results of the application of linear mixed-effects model when \( g_1 = \beta_1 \) is the function of interest. \( S \) denotes the average number of MCMC draws calculated over 1,000 different runs for estimating \( p \) quantile based on the desired relative precision \( d \) and the confidence level \( 1 - \alpha \) after burn-in. \( S' \) denotes the actual number of MCMC draws used averaging over 1,000 runs. For some cases, the initial 8,000 draws are sufficient to achieve the required precision in quantile estimation. In this case, the actual number of MCMC draws used for the simulation is 8,000. For other cases, the initial 8,000 draws are not enough to achieve the required precision so that more draws are used. \( 1 - \hat{\alpha} \) denotes the proportion of 1,000 runs for which \( |\hat{\xi}_p - \xi_p| \leq d \xi_p \) holds based on the actual number of MCMC draws used. \( \bar{d} \) denotes the mean of the estimated absolute relative precisions over 1,000 runs based on actual number of MCMC draws used. Table 3.2 shows results of the application of linear mixed-effects model when \( g_2 = \sigma_1/\sigma_e \) is the parametric function of interest.

Table 3.1  Results of the application of linear mixed-effects model when \( g_1 = \beta_1 \) is the function of interest.

<table>
<thead>
<tr>
<th>Relative Precision ( d )</th>
<th>Confidence Level 1-( \alpha )</th>
<th>Quantile ( p )</th>
<th>Number Needed ( S )</th>
<th>Number Used ( S' )</th>
<th>Estimated Coverage ( 1 - \hat{\alpha} )</th>
<th>Mean Precision ( \bar{d} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.010</td>
<td>0.90</td>
<td>0.80</td>
<td>2,236</td>
<td>8,000</td>
<td>0.999</td>
<td>0.0027</td>
</tr>
<tr>
<td>0.010</td>
<td>0.90</td>
<td>0.90</td>
<td>2,501</td>
<td>8,000</td>
<td>0.995</td>
<td>0.0029</td>
</tr>
<tr>
<td>0.010</td>
<td>0.90</td>
<td>0.95</td>
<td>3,076</td>
<td>8,000</td>
<td>0.982</td>
<td>0.0032</td>
</tr>
<tr>
<td>0.010</td>
<td>0.95</td>
<td>0.80</td>
<td>3,158</td>
<td>8,000</td>
<td>0.997</td>
<td>0.0027</td>
</tr>
<tr>
<td>0.010</td>
<td>0.95</td>
<td>0.90</td>
<td>3,544</td>
<td>8,000</td>
<td>0.999</td>
<td>0.0027</td>
</tr>
<tr>
<td>0.010</td>
<td>0.95</td>
<td>0.95</td>
<td>4,366</td>
<td>8,000</td>
<td>0.995</td>
<td>0.0029</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.80</td>
<td>8,952</td>
<td>8,952</td>
<td>0.982</td>
<td>0.0032</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.90</td>
<td>10,021</td>
<td>10,021</td>
<td>0.907</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.95</td>
<td>12,494</td>
<td>12,494</td>
<td>0.900</td>
<td>0.0024</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.80</td>
<td>12,673</td>
<td>12,673</td>
<td>0.944</td>
<td>0.0021</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.90</td>
<td>14,190</td>
<td>14,190</td>
<td>0.946</td>
<td>0.0020</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.95</td>
<td>17,364</td>
<td>17,364</td>
<td>0.932</td>
<td>0.0021</td>
</tr>
</tbody>
</table>

From Table 3.1 and Table 3.2 we find when the desired amount of precision in quantile estimations increases then needed number of MCMC draws also increases. Also, the
Table 3.2  Results of the application of linear mixed-effects model when \( g_2 = \sigma_1/\sigma_\epsilon \) is the function of interest.

<table>
<thead>
<tr>
<th>Relative Precision ( d )</th>
<th>Confidence Level 1-( \alpha )</th>
<th>Quantile ( p )</th>
<th>Number Needed ( S )</th>
<th>Estimated Coverage 1 - ( \hat{\alpha} )</th>
<th>Mean Precision ( \bar{d} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.010</td>
<td>0.90</td>
<td>0.80</td>
<td>12,065</td>
<td>0.904</td>
<td>0.0048</td>
</tr>
<tr>
<td>0.010</td>
<td>0.90</td>
<td>0.90</td>
<td>12,267</td>
<td>0.878</td>
<td>0.0051</td>
</tr>
<tr>
<td>0.010</td>
<td>0.90</td>
<td>0.95</td>
<td>13,940</td>
<td>0.886</td>
<td>0.0050</td>
</tr>
<tr>
<td>0.010</td>
<td>0.95</td>
<td>0.80</td>
<td>17,124</td>
<td>0.936</td>
<td>0.0042</td>
</tr>
<tr>
<td>0.010</td>
<td>0.95</td>
<td>0.90</td>
<td>17,490</td>
<td>0.935</td>
<td>0.0044</td>
</tr>
<tr>
<td>0.010</td>
<td>0.95</td>
<td>0.95</td>
<td>19,702</td>
<td>0.939</td>
<td>0.0043</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.80</td>
<td>48,268</td>
<td>0.880</td>
<td>0.0026</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.90</td>
<td>49,661</td>
<td>0.893</td>
<td>0.0026</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.95</td>
<td>55,197</td>
<td>0.885</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.80</td>
<td>68,806</td>
<td>0.942</td>
<td>0.0020</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.90</td>
<td>69,878</td>
<td>0.934</td>
<td>0.0022</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.95</td>
<td>77,807</td>
<td>0.942</td>
<td>0.0021</td>
</tr>
</tbody>
</table>

Number of MCMC draws is highly correlated with the quantile probability and the confidence level. More draws are needed if larger quantile probability and smaller confidence level are of interests. Overall the estimated coverage probability 1 - \( \hat{\alpha} \) is very close to the nominal level 1 - \( \alpha \), and the mean of the estimated absolute relative precision \( \bar{d} \) is always about half of the proposed relative precision \( d \). This latter behavior is expected to be seen when the estimation procedure (3.6) is accurate. To see this, note that equation (3.3) entails the large-sample normality of the sample quantile under data dependence

\[
(\hat{\xi}_p - \xi_p) \sim N \left( 0, \frac{2\pi \phi(0)}{Sf^2(\xi_p)} \right)
\]

so that, if the estimated number of draws follows equation (3.6)

\[
S \approx \frac{Z_{1-\alpha/2}^2 2\pi \phi(0)}{(\xi_p d f(\xi_p))^2},
\]

then

\[
(\hat{\xi}_p - \xi_p) \sim N \left( 0, \frac{d^2 \xi_p^2}{Z_{1-\alpha/2}^2} \right)
\]

holds, implying

\[
\hat{d} = \frac{(\hat{\xi}_p - \xi_p)}{\xi_p} \sim N \left( 0, \frac{d^2}{Z_{1-\alpha/2}^2} \right).
\]
and

\[ \text{E}[\hat{d}] = \text{E}\left[ \frac{\hat{\xi}_p - \xi_p}{\xi_p} \right] \approx \frac{d}{|Z_{1-\alpha/2}|} \sqrt{\frac{2}{\pi}} \]

by normal theory; for \(1-\alpha = 0.9\) or 0.95, we then have \(\text{E}[\hat{d}] \approx 0.49d\) or \(0.41d\). So that we expect observed relative precision \(\bar{d}\) in Tables 3.1-3.2 to be about half of \(d\).

### 3.4.2 Application 2: Generalized linear model with a Poisson distribution

The example is from Whyte et al. (1987) and was presented in Dobson (1990). The number of deaths due to AIDS in Australia per 3 month period from January 1983 to June 1986 was recorded. When the response is a count, an appropriate model for linking the response and other explanatory variables is often a Poisson regression model. The Poisson model with a logarithm link function used here is

\[ Y_i \sim \text{Poisson}(\lambda_i), \]

\[ \log(\lambda_i) = \alpha + \beta \log(x_i). \]

Here \(Y_i\) denotes the number of deaths in Australia due to AIDS and \(x_i = i\) denotes time point (measured in multiple of 3 month after January 1983). A random walk Metropolis algorithm is used here to generate draws from the posterior distribution. A multivariate normal prior is assumed on \((\alpha, \beta)\). The first function we consider in this example is \(g_1 = \beta\) which can be interpreted as the marginal effect of time period on the logarithm of the expected number of deaths dues to AIDS. The second, more complicated function is \(g_2 = P(Y_3 = 0)\) which represents the probability of no deaths accrued in the third quarter in 1983. The results are shown in Table 3.3 and Table 3.4 when \(g_1 = \beta\) and \(g_2 = P(Y_3 = 0)\) are the functions of interests, respectively. The definitions of \(S, p, d, \alpha, S', \hat{\alpha}, \) and \(\bar{d}\) refer to those in Tables 3.3 and 3.4. And the results are consistent with the ones shown in Application 1.
Table 3.3  Results of the application of generalized linear model with Poisson distribution when $g_1 = \beta$ is the function of interest.

<table>
<thead>
<tr>
<th>Relative Precision $d$</th>
<th>Confidence Level $1-\alpha$</th>
<th>Quantile $p$</th>
<th>Number Needed $S$</th>
<th>Number Used $S'$</th>
<th>Estimated Coverage $1 - \hat{\alpha}$</th>
<th>Mean Precision $\hat{d}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.010</td>
<td>0.90</td>
<td>0.80</td>
<td>3,224</td>
<td>8,000</td>
<td>0.989</td>
<td>0.0031</td>
</tr>
<tr>
<td>0.100</td>
<td>0.90</td>
<td>0.90</td>
<td>4,422</td>
<td>8,000</td>
<td>0.956</td>
<td>0.0037</td>
</tr>
<tr>
<td>0.010</td>
<td>0.90</td>
<td>0.95</td>
<td>6,420</td>
<td>8,000</td>
<td>0.930</td>
<td>0.0044</td>
</tr>
<tr>
<td>0.010</td>
<td>0.95</td>
<td>0.80</td>
<td>4,578</td>
<td>8,000</td>
<td>0.989</td>
<td>0.0031</td>
</tr>
<tr>
<td>0.010</td>
<td>0.95</td>
<td>0.90</td>
<td>6,278</td>
<td>8,000</td>
<td>0.956</td>
<td>0.0037</td>
</tr>
<tr>
<td>0.100</td>
<td>0.95</td>
<td>0.95</td>
<td>9,115</td>
<td>9,115</td>
<td>0.941</td>
<td>0.0041</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.80</td>
<td>12,895</td>
<td>12,895</td>
<td>0.889</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.90</td>
<td>17,685</td>
<td>17,685</td>
<td>0.886</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.95</td>
<td>25,680</td>
<td>25,680</td>
<td>0.877</td>
<td>0.0026</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.80</td>
<td>18,309</td>
<td>18,309</td>
<td>0.950</td>
<td>0.0021</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.90</td>
<td>25,110</td>
<td>25,110</td>
<td>0.929</td>
<td>0.0022</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.95</td>
<td>36,462</td>
<td>36,462</td>
<td>0.931</td>
<td>0.0022</td>
</tr>
</tbody>
</table>

3.5  Concluding Remarks and Areas for Further Research

This chapter proposed a method to estimate the number of MCMC draws needed in order to obtain the desired precision for quantile estimates used in constructing Bayesian credible intervals. This procedure is especially useful when it is important to eliminate Monte Carlo error in reported results. The number of draws required will depend on the function of interest. Some functions may only require a moderate number of MCMC draws, while others may need a surprisingly large number. The number of draws is also positively related with the quantile probability, the desired precision, and the confidence level. If we want to have more confidence to get more precise estimates for a larger quantile probability, then more MCMC draws are needed.

In this paper we initialize the Markov Chain by choosing appropriate ‘burn-in’ period and then assume the rest of draws approximately follow the stationary distribution. If initialization issues are intended to be avoided the idea based on regenerative simulation or the batch means approach could be used. A regeneration is when the chain restarts
Table 3.4  Results of the application of generalized linear model with Poisson distribution when $g_2 = P(Y_3 = 0)$ is the function of interest.

<table>
<thead>
<tr>
<th>Relative Precision $d$</th>
<th>Confidence Level 1-$\alpha$</th>
<th>Quantile $p$</th>
<th>Number Needed $S$</th>
<th>Estimated Coverage $1 - \hat{\alpha}$</th>
<th>Mean Precision $\hat{d}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.010</td>
<td>0.90</td>
<td>0.80</td>
<td>46,431</td>
<td>0.880</td>
<td>0.0051</td>
</tr>
<tr>
<td>0.010</td>
<td>0.90</td>
<td>0.90</td>
<td>53,309</td>
<td>0.878</td>
<td>0.0051</td>
</tr>
<tr>
<td>0.010</td>
<td>0.90</td>
<td>0.95</td>
<td>75,690</td>
<td>0.94</td>
<td>0.0042</td>
</tr>
<tr>
<td>0.010</td>
<td>0.95</td>
<td>0.80</td>
<td>65,924</td>
<td>0.943</td>
<td>0.0042</td>
</tr>
<tr>
<td>0.010</td>
<td>0.95</td>
<td>0.90</td>
<td>75,690</td>
<td>0.94</td>
<td>0.0042</td>
</tr>
<tr>
<td>0.010</td>
<td>0.95</td>
<td>0.95</td>
<td>93,855</td>
<td>0.930</td>
<td>0.0042</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.80</td>
<td>18,572</td>
<td>0.884</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.90</td>
<td>213,233</td>
<td>0.904</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.005</td>
<td>0.90</td>
<td>0.95</td>
<td>264,400</td>
<td>0.899</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.80</td>
<td>263,696</td>
<td>0.949</td>
<td>0.0021</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.90</td>
<td>302,758</td>
<td>0.944</td>
<td>0.0021</td>
</tr>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.95</td>
<td>375,407</td>
<td>0.934</td>
<td>0.0021</td>
</tr>
</tbody>
</table>

itself and then the tours between regenerations are i.i.d.. The batch means approach breaks MCMC chains into batches of equal sized and then these batches are assumed to be approximately i.i.d.. So both approaches allows the variance of Monte Carlo estimates to be computed based on i.i.d. assumptions. Mykland et al. (1995) applied the splitting technique, which was first introduced by Athreya and Ney (1978) and Nummelin (1978), in regenerative simulation and explicitly stated how to incorporate it into some Markov chain samplers. Hobert et al. (2002) discussed the use of regenerative simulation under appropriate assumptions. Doss and Tan (2013) developed a way to calculate standard errors for estimates of ratios of normalizing constants based on MCMC draws by using regenerative simulation. Jones et al. (2006) considered a stopping rule for MCMC procedure using the width of a confidence interval and used regenerative simulation and modified batch means methods to estimate the variance. The regenerative simulation has theoretically superior properties and the batch means approach is easy to implement. Both regenerative simulation and the batch means approach could be extended to estimate the variance for quantile estimates and then further determine the number of
MCMC draws needed for accurate quantile estimates. However, the main issue related to techniques based on regenerations is that it is computationally difficult to automate in a way that would be immediately applicable to any MCMC sampler and the estimate of the transition density is hard to obtain. This issue has been discussed by Bertail and Clémençon (2006) and Harari-Kermadec (2011) and they proposed the regenerative block bootstrap method which is based on small blocks of random lengths.

In addition, the approach proposed in this paper can be further extended to other applications. If a certain degree of repeatability in bootstrap interval estimates is desirable, then a similar approach can be used to decide the number of bootstrap samples needed. Details on how to do this require further research.
CHAPTER 4. IMPLEMENTATION OF AN ALGORITHM FOR THE NUMBER OF DRAWS NEEDED IN QUANTILE ESTIMATIONS

A paper to be submitted

Jia Liu
Daniel J. Nordman
William Q. Meeker
Department of Statistics
Iowa State University
Ames, Iowa 50010

Abstract

This paper introduces an R implementation of an algorithm for obtaining the number of draws of both i.i.d. sequences and Markov chain Monte Carlo (MCMC) sequences when a specified amount of precision is needed for quantile estimates. A pilot sample of draws is needed for the estimation procedure. The R function receives a vector of a pilot sample of draws from either MCMC sequence or i.i.d. sequence, the quantile probability, the desired precision and the confidence level as the input. And the function returns the number of draws needed.
Keywords: Markov chain Monte Carlo (MCMC); nonparametric quantile estimation; R; precision

4.1 Introduction

Liu et al. (2013) provides an algorithm for computing the number of MCMC draws needed to achieve a desired degree of precision when Bayesian credible interval endpoints are estimated. The Bayesian credible interval endpoints are given by the quantiles of the empirical distribution of the MCMC draws. Liu et al. (2013) also provided an algorithm to compute the number of draws needed for specified precision on quantile estimates under i.i.d. assumptions. In this paper, we will describe how to use a computational tool, presented as an R package, to implement the algorithm suitable for both dependent and i.i.d. sequences. We will explain the R functions and illustrate the use of the functions with several applications.

4.2 Estimation Method

Let $\xi_p$ denote the $p$ quantile of the distribution function $F$ of the random variables $X_1, \ldots, X_n$. We have $F(\xi_p) = p$. The $i$th order statistics is denoted by $x_{(i)}$ such that $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$. Then the sample $p$ quantile of the distribution, which is denoted by $\hat{\xi}_p$, can be expressed as

$$
\hat{\xi}_p = \begin{cases} 
x_{(np)} & \text{if } np \text{ is an integer,} \\
x_{(\lfloor np \rfloor + 1)} & \text{otherwise,}
\end{cases}
$$

(4.1)

where $n$ is the total number of draws and $\lfloor . \rfloor$ denotes the floor function.

When $n$ is large enough, $\hat{\xi}_p$ is approximately normally distributed and centers at the population quantile $\xi_p$. Then we have

$$
\frac{\hat{\xi}_p - \xi_p}{\text{Var}(\hat{\xi}_p)} \to N(0, 1).
$$
In Stuart and Ord (1994) it is shown that for independent sequences $\text{Var}(\hat{\xi}_p) = p(1 - p)/[nf^2(\xi_p)]$, where $f(\xi_p)$ is the probability density function estimated at the population quantile $\xi_p$. And in Chen and Tang (2005) it is shown for dependent sequences from a stationary process $\text{Var}(\hat{\xi}_p) = 2\pi \hat{\phi}(0)/[nf^2(\xi_p)]$, where $\hat{\phi}(0)$ represents the estimated spectral density of $\{Y_i \equiv I(X_i < \xi_p)\}_{i=1}^n$ at the origin. Our goal is to estimate the number of draws needed in order to provide some specified precision with certain confidence level such that $\Pr(|\hat{\xi}_p - \xi_p| \leq d\xi_p) = 1 - \alpha$ holds, where $d$ represents the relative precision for estimating $\xi_p$ and $1 - \alpha$ denotes the confidence level. A pilot set of draws $\tilde{X}_1, \ldots, \tilde{X}_S$ is needed to obtain estimates of $\xi_p, f(\xi_p)$ and $\phi(0)$, which correspond to $\hat{\xi}_p, \hat{f}(\hat{\xi}_p)$ and $\hat{\phi}(0)$.

For an MCMC sequence it is required to select an appropriate length of burn-in and the draws after burn-in are assumed approximately to be stationary. Then the approximate number of draws for independent sequences can be showed to be

$$n = \left\lfloor \frac{Z_{1-\alpha/2}^2[p(1 - p)]}{[d\xi_p f(\xi_p)]^2} \right\rfloor + 1. \tag{4.2}$$

The approximate number of draws for dependent sequences is

$$n = \left\lfloor \frac{Z_{\alpha/2}^2 2\pi \hat{\phi}(0)}{(d\xi_p \hat{f}(\xi_p))^2} \right\rfloor + 1. \tag{4.3}$$

### 4.3 Estimation Algorithm

The most challenging part in this algorithm is to find an accurate way to estimate the probability density function and the spectral density function. We used the flat-top kernel from Politis and Romano (1993, 1995, 1996, 1999) here because it has good asymptotic properties and converges rapidly. Among different forms of flat-top kernels, the simplest one, which is denoted as $\lambda^T(\cdot)$, is given by Politis (2003).

$$\lambda^T(t) = \begin{cases} 1 & 0 \leq |t| \leq 0.5 \\ 2(1 - |t|) & 0.5 \leq |t| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$
Following Liu et al. (2013) the probability density function can be estimated based on the flat-top kernel. In particular,

\[
\hat{f}(x) = \frac{1}{2\pi} \int_{-M}^{M} \lambda^T(t/M)e^{itx} \hat{Q}(t) dt = \frac{1}{\pi S} \sum_{j=1}^{S} g(x, x_j, M),
\]

(4.4)

\[
g(x, x_j, M) = \begin{cases} 
\frac{2}{M(x-x_j)^2}[\cos(\frac{M}{2}(x - x_j)) - \cos(M(x - x_j))] & \text{if } x \neq x_j \\
\frac{5}{8}M & \text{if } x = x_j,
\end{cases}
\]

where \(\hat{Q}(t)\) is the sample characteristic function of \(\{X_i\}_{i=1}^{S}\), and \(M\) is a bandwidth that is defined as the smallest positive real number such that

\[
\max_{t \in (0, K)} |\hat{Q}(M/2 + t)| < c \sqrt{\log S/S}.
\]

Here \(c\) and \(K\) are some positive constants, and Politis (2003) recommended \(c\) to be 2 and \(K\) to be 5. Combining (4.1) and (4.4), the number of draws needed for estimating \(p\) quantile of certain independent sequence based on desired relative precision \(d\) and confidence level \(1 - \alpha\) can be obtained by (4.2).

The sample spectral density function can also be easily obtained based on the flat-top kernel

\[
\hat{\phi}(\omega) = \frac{1}{(2\pi)} \sum_{k=-H}^{H} \lambda^T\left(\frac{k}{H}\right)e^{ik\omega} \hat{R}(k),
\]

(4.5)

where \(\hat{R}(k)\) is the lag-\(k\) sample autocovariance, and the bandwidth parameter \(H\) is the smallest positive integer such that

\[
\max_{k=1,...,K} \frac{\hat{R}(k)}{\hat{R}(0)} < c \sqrt{\log S/S}.
\]

Here \(c\) and \(K\) are still set as 2 and 5, respectively. Then combining (4.1), (4.4) and (4.5), the number of draws needed for estimating \(p\) quantile of some dependent sequence with stationary property based on desired relative precision \(d\) and confidence level \(1 - \alpha\) can be obtained by (4.3).
4.4 Implementations and R Code

In this section we look at how to obtain the appropriate number of draws under both independent and dependent scenarios. The main R function is `numSample`, which returns the number of draws needed. The input of function `numSample` is described below:

- `samp`: A vector of pilot draws. In many situations the number of these pilot draws is far smaller than the actual number needed. For a MCMC sequence these are MCMC draws after an appropriate length of burn-in.
- `d`: The relative precision.
- `p`: The quantile probability. Then the corresponding quantile is $\xi_p$.
- `prob`: The confidence level.
- `ind`: The indicator to denote whether it is an independent sequence or a dependent sequence. When it is an independent sequence then `ind=TRUE`; otherwise `ind=FALSE`.

One example related to an independent sequence is shown below. Suppose we have some initial 10,000 i.i.d. draws which are randomly sampled from normal distribution with mean 2 and standard deviation 4. We first set the seed in order to make results reproducible then generate draws using R function `rnorm`.

```
R> set.seed(100)
R> samp<- rnorm(10000,2,4)
```

If we want to have 95% confidence in estimating 0.8 quantile with 0.01 relative precision then the required number of draws can be obtained as below.

```
R> res<- numSample(samp=samp,d=0.01,p=0.8,prob=0.95,ind=TRUE)
```
Next we will describe how to obtain the number of draws for a MCMC sequence. We use the data about the number of deaths due to AIDS in Australia from the first quarter in 1983 to the second quarter in 1986 inclusively, which was reported in Whyte et al. (1987), as an example. The data set AIDS shown below contains two columns. The first column is the number of deaths in Australia due to AIDS and the second column is the quarterly time point.

```
R> AIDS
   AIDs time
[1,]  0  1
[2,]  1  2
[3,]  2  3
[4,]  3  4
[5,]  1  5
[6,]  4  6
[7,]  9  7
[8,] 18  8
[9,] 23  9
[10,] 31 10
[11,] 20 11
[12,] 25 12
[13,] 37 13
[14,] 45 14
```

A Bayesian Poisson regression model is fit in this data set with AIDs as the response \(Y\) and logarithm of time as the explanatory variable \(X\). The R function MCMC-
Cpoisson in the MCMCpack package is used here to generate a MCMC sample of draws from the posterior distribution of a Poisson regression model. The details about the use of package MCMCpack can be found in Martin et al. (2011). After a graphical check, we throw away the first 2,000 draws and use the following 8,000 draws as our pilot sample of draws. The function of interest, which is denoted as samp in the following code, is the probability of no deaths accrued in the third quarter of year 1983, which is \( P(Y_3 = 0) \).

```R
R> library(MCMCpack)
R> res_fit<- MCMCpoisson(AIDs~log(time),burnin=2000,mcmc=8000)
R> samp<- exp(-exp(res_fit[,1]+res_fit[,2]*log(3)))
```

Then if we are interested in estimating the 0.95 quantile of function samp with 0.005 relative precision and 90% confidence level, the number of MCMC draws required is computed as following.

```R
R> res<- numSample(samp=samp,d=0.005,p=0.95,prob=0.9,ind=FALSE)
R> res
277856
```

### 4.5 Discussion

This paper described how to implement the procedure of estimating the number of draws for both independent and dependent sequences. The algorithm is useful when a certain precision is needed for quantile estimates. R functions are used in implementing the procedure, receiving the inputs and producing the number of draws needed. According to users’ interests, the functions allow users to specify desired precision, the confidence level, the quantile probability, and a sequence of initial draws.
CHAPTER 5. GENERAL CONCLUSIONS

In this dissertation we developed statistical methods on the inference of functions of parameters in some commonly used statistical models. The three main chapters of this dissertation are papers that have been prepared for publication.

In Chapter 2 we conducted a simulation study to explore the performance of different approaches on constructing confidence intervals for nonlinear functions of parameters in a linear mixed-effects model. The simulation results show that the bootstrap-t and BCa bootstrap procedures provide the most accurate coverage properties among all the procedures. These results match the theoretical fact that both procedures are second-order accurate. Both bootstrap procedures, however, cost a lot of computational time. The Wald procedure with observed Hessian matrix and expected Hessian matrix provides almost the same interval estimates. And the Wald procedure and likelihood procedure are quite similar regarding to coverage properties in most situations. Even though the Wald procedure is not accurate for relative small sample size, it is the fastest one to compute. So we would recommend the Wald procedure when computational time is an issue. If, on the other hand, some intensive computation is acceptable, then we would recommend bootstrap-t and BCa procedures. And the simple percentile bootstrap procedure should be avoided.

In Chapter 3 we derived an algorithm to compute the number of MCMC draws needed in order to obtain desired precision for computing credible intervals. In this chapter we used two data applications to illustrate the use of the algorithm. One involves the fit of a linear model with mixed effects and the other involves a generalized linear model with
a Poisson distribution and a log link. Results from these two applications show that the number of MCMC draws depends on the function of interest. Some functions may only require a moderate number of MCMC draws, while others may need a surprisingly large number of draws. The quantile estimates with larger quantile probability, smaller precision, and higher confidence level may require more MCMC draws.

In Chapter 4 we developed an R function that implements the algorithm described in Chapter 3. The R function takes a vector of a pilot sample of draws from either an MCMC sequence or an i.i.d. sequence, the quantile probability, the desired precision and the confidence level as the input, and returns the required number of draws.
BIBLIOGRAPHY


