Generalized fiducial inference for mixed linear models

Jessi Cisewski

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2012

Approved by:

Jan Hannig

Chuanshu Ji

Shu Lu

Haipeng Shen

Richard L. Smith

ABSTRACT
JESSI CISEWSKI: Generalized fiducial inference for mixed linear models.
(Under the direction of Jan Hannig.)

Fiducial inference was proposed by R.A. Fisher in 1930 to overcome what he perceived as a deficiency in Bayesian methodology – assuming a prior distribution without prior knowledge. Due to some controversy, fiducial inference quickly fell into disfavor by the statistical community and was left undeveloped by Fisher. There were several attempts over the subsequent decades to revive fiducial inference. Eventually a connection was drawn between fiducial inference and generalized inference, called generalized fiducial inference (GFI). Under the GFI paradigm, inference is performed by considering the generalized fiducial distribution on the parameter space with a flexibility similar to a posterior distribution in the Bayesian framework.

GFI can be thought of as a transference of probability from the model space to the parameter space, and a generalized fiducial distribution is defined for the unknown parameters of the model. In this dissertation, we apply the generalized fiducial framework to the normal linear mixed model setting and to logistic models with mixed effects. GFI is a computationally-based mode of inference, and we develop sequential Monte Carlo algorithms to obtain samples from the generalized fiducial distribution on the parameter space. In the normal linear mixed model setting, the proposed method is found to be competitive or better when evaluated based on frequentist criteria of empirical coverage and average length of confidence intervals. In the logistic setting with mixed effects setting, the simulation study reveals that the generalized fiducial approach tends to have correct empirical coverage, along with providing finite confidence intervals in cases where competing methods have disjoint, infinite, or non-calculable intervals.

For the final part of this dissertation, we developed a methodology for classifying an unknown powder as a particular harmful substance (*Bacillus anthracis* spores) or not. A wavelet transformation was incorporated to allow for possible thresholding or standardization, and

then a linear model technique using the known elemental structure of the harmful substance was used for dimension reduction, and finally a support vector machine approach was employed for the final classification of the substance. The method was applied to real-data produced from a laser-induced breakdown spectroscopy device.

Keywords: generalized fiducial inference; normal linear mixed model; random effects; variance components; logistic regression; sequential Monte Carlo; variance component; classification; support vector machine; wavelet; dimension reduction; laser induced breakdown spectroscopy

ACKNOWLEDGMENTS

I am extremely thankful for my advisor, Professor Jan Hannig. His encouragement, wisdom, and patience throughout my studies emphatically contributed to my positive experience at the University of North Carolina at Chapel Hill. His generosity with his time and knowledge has greatly shaped and improved my abilities as a statistician. Furthermore, his exuberance for statistics made it all really fun!

I would like to thank Professors Chuanshu Ji, Shu Lu, Haipeng Shen, and Richard L. Smith for their support and participation on my dissertation committee. A special thanks to Professor Smith for his helpful suggestions and encouragement both in statistics and running.

I would also like to thank the Department of Statistics and Operations Research for giving me the opportunity to be a part of this outstanding department for the past five years. It has truly been a privilege of which I will always be grateful.

The other students in the department, especially my cohort, has made my experience here quite an adventure; I thank them all for making life interesting. Additionally, I am very thankful for Elizabeth and Brian for their advice, encouragement, and friendship.

I sometimes wonder if my main purpose of coming here was to become a part of the community at the Chapel Hill Bible Church. I am eternally thankful for the great friends and mentors that I met there, especially within my small group. I would not be the woman I am today without Amanda, Cara, and Rachel.

My friends and relatives who live all around the United States are always an inspiration for me. I thank them for their encouragement of me in all my endeavors and for their love. Finally, I want to thank my amazing family - Mom, Dad, Wally and Jodi. I love you more than I could ever sufficiently express.

"Praise the LORD.

Praise God in his sanctuary; praise him in his mighty heavens.

Praise him for his acts of power; praise him for his surpassing greatness.

Praise him with the sounding of the trumpet, praise him with the harp and lyre,

praise him with timbrel and dancing, praise him with the strings and pipe,

praise him with the clash of cymbals, praise him with resounding cymbals.

Let everything that has breath praise the LORD.

Praise the LORD." Psalm 150 (NIV)

Table of Contents

List of Figures

Chapter 1

Background

1.1 Introduction

This dissertation includes three main contributions discussed in Chapters 2, 3, and 4. In Chapters 2 and 3, the focus is on using the framework of generalized fiducial inference in the normal linear mixed model and logistic mixed model settings. Because generalized fiducial inference is a computationally-based methodology, the development of computational algorithms was required. For both model forms, sequential Monte Carlo algorithms were designed, implemented and tested. Generalized fiducial inference and sequential Monte Carlo methodology are introduced in Chapter 1.

Chapter 4 is not related to generalized fiducial inference and can be read independently of the other chapters. The methodology developed in Chapter 4 was motivated and derived using real data provided by researchers at the US Environmental Protection Agency.

Supplementary material is included for completeness in appendices.

Except where noted or obvious, the notational convention that will be used for matrices, vectors, and single values will be, respectively, bold and capital letters for matrices, capital letters for vectors, and lowercase letters for single values (e.g. $\mathbf{A}$, A, a). The acronym 'i.i.d.' will be used for independent and identically distributed random variables.

1.2 Generalized fiducial inference

Fiducial inference was introduced by R.A. Fisher (Fisher, 1930) to rectify what he saw as a weakness in the Bayesian philosophy when a prior distribution is assumed without sufficient prior knowledge. In Fisher (1930), Fisher writes that inverse probability (i.e. posterior probability) is "fundamentally false and devoid of foundation." Fisher expounds his concerns with the idea of noninformative priors when he wrote the following paragraph in the same manuscript:

> The underlying mental cause is not to be confused with the various secondary errors into which one is naturally led in deriving a formal justification of a false position, such as for example Laplace's introduction into his definition of probability of the unelucidated phrase "equally *possible* cases" which, since we must be taken to know what cases are equally possible before we know that they are equally probable, can only lead to the doctrine, known as the "doctrine of insufficient reason," that cases are equally probably (to us) unless we have reason to think the contrary, and so reduces all probability to a subjective judgement. [Emphasis Fisher's]

While Fisher made several attempts at justifying his definition of fiducial inference (Fisher, 1933, 1935), it was not developed to the point of acceptance by the statistical community. In particular, fiducial inference fell into disrepute when it was discovered that some of the properties Fisher claimed did not hold (Lindley, 1958; Zabell, 1992). The main concerns with Fisher's fiducial inference were related to its nonexactness (in the frequentist sense) and nonuniqueness. Over the subsequent decades, a number of other papers appeared with attempts to revitalize fiducial inference by incorporating more structure (Dempster, 1968; Dawid and Stone, 1982; Fraser, 1961b,a, 1966, 1968; Barnard, 1995), but still the ideas did not seem to fully catch the attention of the statistical community.

In Tsui and Weerahandi (1989), generalized p-values were introduced generalizing the notion of classical p-values leading to the development of generalized inference. Generalized inference generalizes the notions of classical statistics providing a framework that allows

classical approaches in problems that were otherwise not manageable in that framework (e.g. in the presence of nuisance parameters). More recently, Hannig et al. (2006) connects Fisher's fiducial inference to generalized inference introduced in Tsui and Weerahandi (1989) and Weerahandi (1993). Hannig et al. (2006) defines a subclass of generalized pivotal quantities, which the call *fiducial generalized pivotal quantities* (FGPQ), outline a recipe for constructing confidence intervals using FGPQ's, and prove the asymptotic exactness of these intervals.

A thorough introduction to generalized fiducial inference can be found in Hannig (2009b). The main idea of fiducial inference is a transference of randomness from the model space to the parameter space. For a simple illustration of the fiducial argument, let y be a realization of a random variable $Y \sim N(\mu, 1)$ (where $N(\mu, 1)$ represents a normal distribution with unknown mean μ and standard deviation 1). The random variable Y can be represented as $Y = \mu + Z$ where $Z \sim N(0, 1)$. Given the observed value y, the fiducial argument solves this equation for the unknown parameter μ to get $\mu = y - Z$ (e.g. suppose $y = 4.8$, then $\mu = 4.8 - Z$ would suggest $\mu \sim N(4.8, 1)$). While the actual value of the realization of Z is unknown, the distribution of Z is fully known and can be used to frame a distribution on the unknown parameter μ. This distribution on μ is called the fiducial distribution.

The generalized fiducial recipe starts with a data-generating equation, also referred to as the *structural equation*, which defines the relationship between the data and the parameters. Let Y be a random vector indexed by parameter(s) $\xi \in \Xi$, then assume Y can be represented as

$$Y = G(\xi, U), \tag{1.1}$$

where G is a jointly measurable function indicating the structural equation, and U is a random element with a fully known distribution (void of unknown parameters). In Chapter 2, the function G will take the form of a normal linear mixed model, and the random components U will be standard normal random variables (see Equation (2.2)). Following the fiducial argument, we define a set-valued function, the inverse image of G, as $Q(\mathbf{y}, \mathbf{u}) = \{\xi : \mathbf{y} = G(\xi, \mathbf{u})\}$, where $\mathbf{y}$ is the observed data vector and $\mathbf{u}$ is an arbitrary realization of U. The set-function $Q(\mathbf{y}, \mathbf{u})$ is then used to define the fiducial distribution on the parameter space. Since

the distribution of U is completely known, independent copies of U, U^*, can be generated to produce a random sample of $Q(\mathbf{y}, \mathbf{u}^*)$ for the given data $\mathbf{y}$ (where $\mathbf{u}^*$ is a realization of U^*). There are several sources of nonuniqueness in this framework. In particular, nonuniqueness could occur if Q has more than one element, if Q is empty, or due to the definition of the structural equation. Nonuniqueness due to the definition of the structural equation will not be addressed here as we assume the form of the model is known (i.e. normal linear mixed model and logistic regression with mixed effects). To resolve the case when there is more than one element in Q, we can define a rule, call it V, for selecting an element of Q. Furthermore, since the parameters ξ are fixed but unknown, there must be some realization of the random variable U such that $\mathbf{y} = G(\xi, \mathbf{u})$ has occurred (i.e. $\{Q(\mathbf{y}, \mathbf{u}) \neq \emptyset\}$). The generalized fiducial distribution of ξ is defined as

$$V(Q(\mathbf{y}, U^*)) \mid \{Q(\mathbf{y}, U^*) \neq \emptyset\}. \tag{1.2}$$

A random element having the distribution of (1.2) is denoted as $\mathcal{R}_\xi$ and called the generalized fiducial quantity (GFQ). The generalized fiducial density of $\mathcal{R}_\xi$ is defined in Hannig (2009b) assuming the structural equation of (1.1) can be factorized as $Y_i = g_i(\xi, \mathbf{U})$ for $i = 1, \ldots, n$, and letting $\mathbf{G} = (g_1, \ldots, g_n)$. If p is the dimension of the unknown parameter(s) vector ξ, define $\mathbf{Y_i} = (Y_{i_1}, \ldots, Y_{i_p})$ and $\mathbf{U_i} = (U_{i_1}, \ldots, U_{i_p})$ so that $\mathbf{Y_i} = \mathbf{G_i}(\xi, \mathbf{U_i})$ (where $\mathbf{i} = (i_1, \ldots, i_p)$ is an arbitrary selection of p of n equations, with $\mathbf{Y_{i^c}}$, $\mathbf{U_{i^c}}$ and $\mathbf{G_{i^c}}$ representing the components of $\mathbf{Y}$, $\mathbf{U}$ and $\mathbf{G}$, respectively, not selected). Assuming the $\mathbf{G_i}$ are invertible and differentiable for each $\mathbf{U_i}$, then the generalized fiducial density is defined as

$$f_{\mathcal{R}_\xi}(\xi) = \frac{f_{\mathbf{Y}}(\mathbf{y}|\xi) J(\mathbf{y}, \xi)}{\int_\Xi f_{\mathbf{Y}}(\mathbf{y}|\xi') J(\mathbf{y}, \xi') \partial \xi'}, \tag{1.3}$$

where

$$J(\mathbf{y}, \xi) = \binom{n}{p}^{-1} \sum_{i=(i_1,\ldots,i_p)} \left| \frac{\det\left(\frac{\partial}{\partial \xi} \mathbf{G_i^{-1}}(\mathbf{y}_i, \xi)\right)}{\det\left(\frac{\partial}{\partial y_i} \mathbf{G_i^{-1}}(\mathbf{y}_i, \xi)\right)} \right|.$$

Note that $p \leq n$ equations of $\mathbf{G}$ are selected to solve the structural equation for ξ, and the remaining $n - p$ equations of $\mathbf{G}$ are then conditioned on the first p equations being correct.

Since the p equations selected are arbitrary, $J(\mathbf{y}, \xi)$ is an average over all possible selections of p equations. The theoretical validation of equation (1.3) is presented in Hannig (2009b). Though the form of the generalized fiducial distribution displayed in Equation (1.3) is visibly similar to a Bayesian posterior distribution with the $J(\mathbf{y}, \xi)$ term acting as a prior, $J(\mathbf{y}, \xi)$ would need to be decomposable into two measurable functions: one a function *only* of the data (i.e. independent of the parameters), and the other a function of the parameters. The decomposition does not hold in general. In fact, Grundy (1956) provides a one-parameter example where no prior distribution can be found that would produce the fiducial distribution.

Defining the generalized fiducial distribution as (1.2) leads to a potential source of nonuniqueness due to conditioning on events with zero probability (i.e. if $P(\{Q(\mathbf{y}, \mathbf{u}) \neq \emptyset\}) = 0$). This is known as the Borel Paradox (e.g. Casella and Berger, 2002). Fortunately this can be resolved by noting that most data has some degree of known uncertainty due, for example, to the resolution of the instrument collecting the data or computer storage. Because of this, instead of considering the value of a datum, an interval around the value can be used (Hannig, 2009a; Hannig et al., 2007). For example, suppose the datum value is $y = 1.632$ meters measuring the height of a woman. If the resolution of the instrument used to measure the woman is 0.001 m (i.e. 1 mm), then her actual height is between 1.631 m and 1.632 m (or between 1.632 m and 1.633 m depending on the practice of the measurer).

By considering interval data, the issue of nonuniqueness due to the Borel Paradox is resolved since the probability of observing our data will never be zero since $P(Q((a, b], U^*) \neq \emptyset) \geq P(Y \in (a, b]) > 0$ where $a < b$ are the endpoints of the interval.

To illustrate how discretized data will be used in the context of generalized fiducial inference, consider the following structural equation

$$Y = \mu + \sigma Z$$

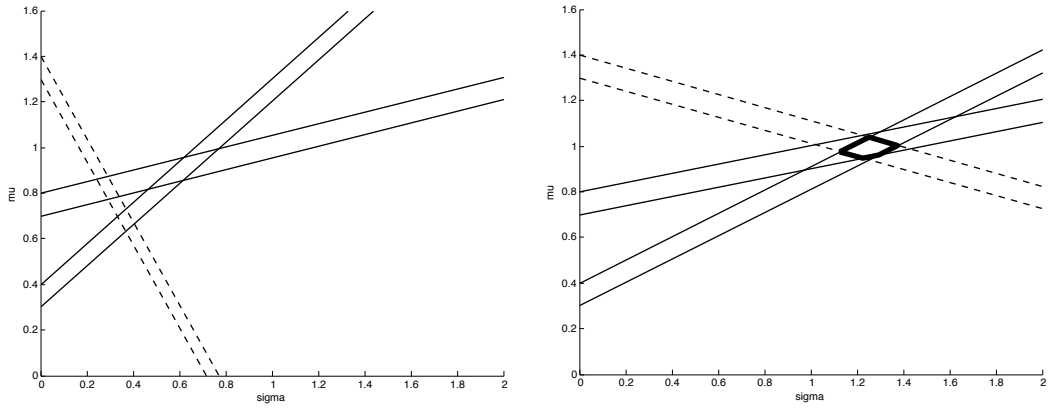
where $Z \sim N(0, 1)$, and (μ, σ) are unknown parameters. Suppose we observe $Y = (y_1, y_2, y_3)' = (0.32, 0.75, 1.37)'$, but due to the precision of the measuring device, we actually only know the following:

$$0.3 < y_1 \leq 0.4, \quad 0.7 < y_2 \leq 0.8, \quad 1.3 < y_3 \leq 1.4.$$

To find the set Q, we simulate $Z = (z_1, z_2, z_3)'$ such that

$$0.3 < \mu + \sigma z_1 \leq 0.4, \quad 0.7 < \mu + \sigma z_2 \leq 0.8, \quad 1.3 < \mu + \sigma z_3 \leq 1.4.$$

Since this is a system of three equations with two unknowns ($p = 2$), we can use the first two inequalities to simulate any values for z_1 and z_2 (with the only requirement that $z_1 \neq z_2$, which has probability zero), and then generate z_3 so that Q is non-empty. Figure 1.1(a) and Figure 1.1(b) illustrate this idea geometrically. Figure 1.1(a) does not produce a viable sample of Z as (z_1, z_2, z_3) does *not* satisfy the system of inequalities. The generated of Z's in Figure 1.1(b) *does* satisfy the system of inequalities.



(a) (b)

Figure 1.1: (a) The solid lines reflect the discretized data values for $x_1 = 0.32$ and $x_2 = 0.75$ with sampled standard normal values $z_1 = -0.9040$ and $z_2 = -0.2544$ and $z_3 = 1.8080$, and the set Q is the bold parallelogram formed by the intersection of these two sets of parallel lines. (b) The setting is the same as (a) except with a new sample of $(z_1, z_2, z_3) = (-0.5115, -0.2022, 0.2880)$, which *does* solve the system of inequalities. The bold interior polygon is the new set Q.

Asymptotic results for interval data using generalized fiducial inference can be found in Hannig (2009a). Asymptotic theory was addressed for the case when the sample size is fixed, but the fatness (i.e. the width of the interval) decreases to zero, and when the fatness is fixed, but the sample size increases. An assumption in both settings is that the data are independent.

Interval data are not explicitly required for generalized fiducial inference, but is useful in the proposed setting of normal linear mixed models. Generalized fiducial inference has a number of other applications in various settings such as wavelet regression (Hannig and Lee, 2009), confidence intervals for extremes (Wandler and Hannig, In press), metrology (Hannig et al., 2007), and variance component models (E et al., 2008), which applies the generalized fiducial framework to unbalanced normal mixed linear models with two variance components.

1.3 Sequential Monte Carlo methods

As previously noted, generalized fiducial inference is a computationally-based mode of inference, and we employ sequential Monte Carlo (SMC) methods in Chapters 2 and 3. The framework of SMC is introduced in this section. When integrals of interest are very complex or unsolvable by analytical methods, simulation-based methods can be used. SMC, or particle filters, is a collection of simulation methods used to sample from an evolving target distribution (i.e. the distribution of interest) accomplished by propagating a system of weighted particles through time or some other index. A solid introduction and applications of SMC methods can be found in Doucet et al. (2001). There are many dimensions to the theory and uses of SMC algorithms (e.g. Chopin (2002, 2004); Del Moral et al. (2006); Douc and Moulines (2008); Kong et al. (1994); Liu and Chen (1998); Liu and West (2001)), but a basic introduction to the methodology is presented below.

SMC has its roots in importance sampling (IS). Given a (possibly vector) random variable Z with probability density function π supported on the set $\mathcal{Z}$, it is often of interest to find the expected value of some measurable function ϕ, which is

$$\mathbb{E}_\pi(\phi(Z)) = \int_{z\in\mathcal{Z}} \phi(z)\pi(z)dz. \tag{1.4}$$

If this integral cannot be solved analytically, but a random sample from π is obtainable, one can estimate the expected value in (1.4) under some assumptions by invoking the law of large

numbers using the estimator

$$\widehat{\mathbb{E}}_{\pi}(\phi(Z)) = \frac{1}{N}\sum_{J=1}^{N}\phi(z^{(J)})$$

where $z^{(J)}$ is sampled from π. Then as $N \longrightarrow \infty$, $\widehat{\mathbb{E}}_{\pi}(\phi(Z)) \longrightarrow \mathbb{E}_{\pi}(\phi(Z))$. If it was not easy, or possible, to directly sample from π, IS can be used to generate the sample.

The idea of IS is to find a *proposal density* $\widetilde{\pi}$ such that it is possible to obtain a random sample from it, and then re-weight the values back to the *target density* π using the following relation

$$\mathbb{E}_{\pi}\phi(\mathbf{Z}) = \int_{z\in\mathcal{Z}}\phi(z)\pi(z)dz = \int_{z\in\mathcal{Z}}\phi(z)\cdot\frac{\pi(z)}{\widetilde{\pi}(z)}\cdot\widetilde{\pi}(z)dz = \mathbb{E}_{\widetilde{\pi}}\Big(\phi(\mathbf{Z})\cdot\frac{\pi(\mathbf{Z})}{\widetilde{\pi}(\mathbf{Z})}\Big),$$

which can then be approximated by

$$\hat{\mathbb{E}}_{\widetilde{\pi}}\phi(z) = \frac{\sum_{J=1}^{N}\phi(z^{(J)})W^{(J)}}{\sum_{K=1}^{N}W^{(K)}},$$

where $W^{(J)}$ are the unnormalized importance weight $\frac{\pi(z^{J)})}{\widetilde{\pi}(z^{(J)})}$, $J = 1, \ldots, N$.

An acceptable proposal distribution is one such that its support contains the support of the target distribution, it should have the same shape as the target distribution, and, of course, it should be possible to obtain a random sample from it. It can also desirable for the proposal distribution to have heavier tails than the target distribution (Geweke, 1989).

If the data arrive sequentially, IS would not be an efficient method since all the importance weights would need to be recomputed upon the arrival of new data. Sequential importance sampling (SIS) is an iterative alternative to IS, and allows the importance weights to be updated sequentially. The framework is slightly more complicated than presented above for IS. Consider the case where we have data arriving sequentially as $Y_{1:t} = (y_1, \ldots, y_t)^T$ with an underlying, unobservable signal $Z_{0:t} = (z_0, \ldots, z_t)^T$.[1] (For simplicity, think of $y_i = \mu + \sigma z_i$ where $z_i \sim N(0,1)$, then the vector $Y_{1:t}$ is the observed data, but the $Z_{1:t}$ is unobservable.) The goal, then is to obtain a sample of the $Z_{0:t}$ given the observable $Y_{1:t}$, and hence target the

[1]The hidden process $Z_{0:t}$ starts with 0 rather than 1 in order to initiate the process. This is often used in the Bayesian context, but will not be necessary later for the proposed generalized fiducial method.

distribution $\pi_{1:t}(Z_{0:t}|Y_{1:t})$. If it is not possible to sample from $\pi_{1:t}$, then a proposal distribution $\widetilde{\pi}_{1:t}$ can be selected. The proposal distribution is defined in a manner so that the importance weights $\pi_{1:t}/\widetilde{\pi}_{1:t}$ can be updated recursively with the arrival of a new data point y_{t+1}. In the IS setting, the *unnormalized* importance weight at time t for particle $J = 1, \ldots, N$ would be written

$$W_{1:t}(Z_{0:t}^{(J)}) = \frac{\pi_{1:t}(Z_{0:t}^{(J)}|Y_{1:t})}{\widetilde{\pi}_{1:t}(Z_{0:t}^{(J)}|Y_{1:t})}.$$

Rather than throw out the previously generated values at each new time step, the weights can be determined recursively by noting the following relationships. First, the target distribution at time $t+1$, $\pi_{1:t+1}$ can be written in terms of a marginal and conditional distribution as

$$\pi_{1:t+1}(Z_{0:t+1}|Y_{1:t+1}) = \pi_{1:t}(Z_{0:t}|Y_{1:t}) \cdot \pi_{t+1|1:t}(Z_{t+1}|Z_{0:t}, Y_{1:t}).$$

A similar formulation can be applied to the proposal distribution, resulting in importance weights at time $t+1$ as

$$W_{1:t+1} = W_{1:t} \cdot \frac{\pi_{t+1|1:t}(Z_{t+1}|Z_{0:t}, Y_{1:t})}{\widetilde{\pi}_{t+1|1:t}(Z_{t+1}|Z_{0:t}, Y_{1:t})}. \tag{1.5}$$

The importance weights of (1.5) can be simplified even further under additional assumptions such as if the hidden process, $Z_{0:t}$, is Markovian. See the introductory chapter of Doucet et al. (2001) for more details about this formulation. Liu and Chen (1998) also discusses the ideas of SIS.

The generated values of $Z_{0:t}^{(J)}$, for $J = 1, \ldots, N$ are typically referred to as *particles*. With their associated *normalized* importances weights, $\{Z_{0:t}^{(J)}, \hat{W}_{1:t}^{(J)}\}_{J=1}^{N}$ is called the *particle system*, where $\hat{W}_{1:t}^{(J)} = \frac{W_{1:t}(Z_{0:t}^{(J)})}{\sum_{K=1}^{N} W_{1:t}(Z_{0:t}^{(K)})}$.

Unfortunately, this method will fail as t increases because the importance weights degenerate to a single particle (i.e. one particle has a normalized importance weight equal to one, the rest zero). The degeneracy of the particle system is often measured by the effective sample size (ESS), which is a measure of the distribution of the weights of the particles. Kong et al. (1994) presents the ESS as having an inverse relation with the coefficient of variation of

the particle weights, and proved that this coefficient of variation increases as the time index increases (i.e. as more data becomes available) in the SIS setting. The ESS at time t is often estimated

$$\text{ESS}_t = \Big(\sum_{J=1}^{N} W_{1:t}^{(J)} \Big)^2 \times \Big(\sum_{K=1}^{N} (W_{1:t}^{(K)})^2 \Big)^{-1}, \tag{1.6}$$

with unnormalized importance weights, $W_{1:t}$ as defined in (1.5). The desire is for the ESS to be as close to N as possible, but as noted above, it can drop fairly quickly to one. An intuitive explanation of ESS, and a rejuvenation procedure can be found in Liu and Chen (1995), which lead to sequential importance sampling resampling procedures, and ultimately SMC methodology.

SMC builds on ideas of sequential importance sampling (SIS) by incorporating a resampling step to resolve issues with the degeneracy of the particle system. Once the ESS for the particle system has dropped below some designated threshold or at some pre-specified time, the particle system is resampled removing inefficient particles with low weights and replicating the particles with higher weights (Liu and Chen, 1995). There are various methods for resampling with the most basic being multinomial resampling, which resamples particles based on the normalized importance weights (see Douc et al. (2005) or Liu and Chen (1998) for a comparison of several resampling methods). The resampling method employed throughout the work presented in this dissertation comes from the appendix of Kitagawa (1996) due to its computational efficiency requiring the generation of only a single random number. It is a form of deterministic, residual resampling where a completely randomized resampling scheme would select each resampled particle according to its importance weight. After resampling, all the particles are assigned an importance weight of $W_{1:t}^{(J)} = N^{-1}$ for $J = 1, \ldots, N$.

Examples of general SMC algorithms can be found in Del Moral et al. (2006) or Jasra et al. (2007). The main idea of SMC methods is to iteratively target a sequence of distributions $\{\pi_{t|1:t-1}\}_{t \in \mathbf{Z}^+}$, where $\pi_{t|1:t-1}$ is often some distribution based on the data available up to time t. The algorithm comprises three main sections after the initialization step: sampling, correction, and resampling. The sampling step arises at a new time step t when particles are sampled from some evolving conditional proposal distribution $\widetilde{\pi}_{t|1:t-1}$. The correction step

is concerned with the calculation of the weights and the idea of reweighting the particles to target the desired conditional distribution at time t, $\pi_{t|1:t-1}$. The resampling step is performed when the ESS of the particle system falls below some desired threshold T (e.g. $T = N/2$). The asymptotic correctness for SMC algorithms can be found in Chopin (2004) and Douc and Moulines (2008).

Chapter 2

Normal linear mixed models

2.1 Summary

While normal linear mixed modeling methods are foundational concepts introduced in any statistical education, adequate general methods for inference involving models with more than a few variance components are lacking, especially in the unbalanced setting. Generalized fiducial inference provides a possible framework that accommodates this dearth of methodology. Under the fabric of generalized fiducial inference along with sequential Monte Carlo methods, we present an approach for inference for both balanced and unbalanced Gaussian linear mixed models. We compare the proposed method to classical and Bayesian results in the literature in a simulation study of two-fold nested models and two-factor crossed designs with an interaction term. The proposed method is found to be competitive or better when evaluated based on frequentist criteria of empirical coverage and average length of confidence intervals. A MATLAB implementation of the proposed algorithm is available.

2.2 Introduction

Inference on parameters of normal mixed linear models has an extensive history (see Khuri and Sahai (1985) for a survey of variance component methodology, or Chapter 2 of Searle et al. (1992) for a summary). There are many inference methods for variance components such as ANOVA-based methods Satterthwaite (1941, 1946); Burdick and Graybill (1992);

Hernandez et al. (1992); Hernandez and Burdick (1993); Jeyaratnam and Graybill (1980), maximum likelihood estimation (MLE) and restricted maximum likelihood (REML) Hartley and Rao (1967); Searle et al. (1992) along with Bayesian methods Gelman (2006); Gelman et al. (2004); Wolfinger and Kass (2000). Many of the ANOVA-based methods become quite complex with complicated models (e.g. due to nesting or crossing data structures), and are not guaranteed to perform adequately when the designs become unbalanced.

With notable optimality properties for point estimation, MLE and REML methods are less useful when it comes to confidence interval estimation for small samples because the asymptotic REML-based confidence intervals tend to be liberal (Burch, 2011; Burdick and Graybill, 1992; Searle et al., 1992; Khuri and Sahai, 1985). Bayesian methods, in particular, hierarchical modeling, can be an effective resolution to complicated models, but the delicate question of selecting appropriate prior distributions must be addressed.

There are numerous applications of normal mixed linear models related to topics such as animal breeding studies (Burch and Iyer, 1997; E et al., 2008), multilevel studies (O'Connell and McCoach, 2008)), longitudinal studies (Laird and Ware, 1982), but many methods do not go beyond two variance components or are designed for a very specific setting. We propose a solution based on generalized fiducial inference that easily allows for inference beyond two variance components and for the general normal linear mixed model settings.

A typical form of a normal mixed linear model is

$$Y = \mathbf{X}\beta + \mathbf{VZ} + \epsilon, \tag{2.1}$$

where Y is an $n \times 1$ vector of data, $\mathbf{X}$ is a known $n \times p$ fixed effects design matrix, β is a $p \times 1$ vector of unknown fixed effects, $\mathbf{VZ} = \sum_{i=1}^{r-1} \mathbf{V}_i Z_i$, where Z_i is a vector of effects representing each level of random effect i such that $E(Z_i) = 0$ and $\text{var}(Z_i) = \sigma_i^2 \mathbf{I}_{l_i}$ where l_i is the number of levels of random effect i, $\mathbf{V}_i$ is the known design matrix for random effect i, and ϵ is an $n \times 1$ vector representing the error and $E(\epsilon) = 0$ and $\text{var}(\epsilon) = \sigma_\epsilon^2 \mathbf{I}_n$. Note that there are r total random components in this model. It is often assumed that $\mathbf{Z}$ and ϵ are independent,

$\mathbf{Z} \sim N(\mathbf{0}, \mathbf{D})$, where

$$\mathbf{D} = \begin{bmatrix} \sigma_1^2 \cdot \mathbf{I}_{l_1} & & \\ & \ddots & \\ & & \sigma_{r-1}^2 \cdot \mathbf{I}_{l_{r-1}} \end{bmatrix},$$

and $\epsilon \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_n)$, thus, $Y \sim N(\mathbf{X}\beta, \sum_{i=1}^{r-1} \sigma_i^2 \mathbf{V}_i \mathbf{V}_i^T + \sigma_\epsilon^2 \mathbf{I}_n)$.

An unbalanced design occurs when the number of observations, n_{ij}, $j = 1, \ldots, l_i$ and $i = 1, \ldots, r$, falling within each level of a given effect are not all equal ($n_{ij} \neq n_{ik}$ for some $j \neq k$ with $j, k = 1, \ldots, l_i$).

The focus of this chapter is the construction of confidence intervals for the unknown parameters of (2.1), with emphasis on the variance components σ_i^2 for $i = 1, \ldots, r-1$ and σ_ϵ^2. Inferences are derived from the generalized fiducial distributions of the unknown parameters, and we propose a sequential Monte Carlo (SMC) algorithm to obtain these samples. Like a Bayesian posterior without assuming a prior, this procedure produces a distribution on the parameter space. We evaluate the quality of the simulated generalized fiducial distribution based on the quality of the confidence intervals. We introduce the proposed method using a simple model with two unknown parameters, then explain how it generalizes to all normal linear mixed models, and state and prove a theorem concluding the convergence of the algorithm. To demonstrate small sample performance, we perform a simulation study on two different types of models (two-fold nested models and two-factor cross-classification with interaction models), and include a real-data application for each model.

2.3 Method

In this section, we discuss how the generalized fiducial framework introduced in Section 1.2 can be applied to the normal linear mixed model setting. The proposed generalized fiducial approach is designed specifically for *interval* data (e.g. due to the measuring instrument's resolution, rounding for storage on a computer, or bid-ask spread in financial data). There are several reasons to consider interval data. First, there are many examples where it is critical (or required per the regulations outlined in GUM (1995)) to incorporate all known sources of

uncertainty (e.g. Elster, 2000; Frenkel and Kirkup, 2005; Hannig et al., 2007; Lira and Woger, 1997; Taraldsen, 2006; Willink, 2007). Secondly, our simulation results for the normal linear mixed model show that even when considering this extra source of uncertainty, the proposed method is competitive or better than classical and Bayesian methods that assume the data are exact (see Section 2.4). (The proposed method is also appropriate for *noninterval*, or standard, data simply by artificially defining a narrow interval around each datum; Hannig (2009a) proves that as the interval width decreases to zero the generalized fiducial distribution converges to the generalized fiducial distribution for exact data.) Finally, on a purely philosophical level, all continuous data has some degree of uncertainty, as noted previously, due to the resolution of the measuring instrument or truncation for storage on a computer. Note that we are not suggesting that all methods should incorporate this known uncertainty; however, we were able to appeal to this known uncertainty for the computational aspect of the proposed method.

The form of the normal mixed linear model from equation (2.1) is adapted to work in the generalized fiducial inference setting as

$$\mathbf{Y} = \mathbf{X}\beta + \sum_{i=1}^{r} \sigma_i \sum_{j=1}^{l_i} V_{i,j} z_{i,j}, \tag{2.2}$$

where $\mathbf{X}$ is a known $n \times p$ fixed-effects design matrix, β is the $p \times 1$ vector of fixed effects, $V_{i,j}$ is the $n \times 1$ design vector for level j of random effect i, l_i is the number of levels per random effect i, σ_i^2 is the variance of random effect i, and the $z_{i,j}$ are i.i.d. standard normal random variables. The error variance σ_ϵ^2 from (2.1) is equivalent to σ_r^2 in this model, and $V_{r,j}$ is a vector of zeros with a one at j for $j = 1, \ldots, l_r = n$ (i.e. $\mathbf{V}_r$ is the identity matrix). We will derive a framework for estimation for the unknown parameters (β and $\sigma_i, i = 1, \ldots, r$) of this model, which will be applicable to both the balanced and the unbalanced case. The design is balanced if there is an equal number of observations in each level of each effect, otherwise the design is unbalanced. The unbalanced case is especially important because the classical methods are no longer exact when the designs are unbalanced.

Consider the following examples illustrating the connection between (2.1) and (2.2).

Example 1: One-way random effects model. A one-way random effects model is conventionally written as

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, i = 1, \ldots, a, j = 1, \ldots, n_i, \tag{2.3}$$

with unknown mean μ, random effect $\alpha \sim N(0, \sigma_\alpha^2)$ where σ_α^2 is unknown, a is the number of levels of α, n_i is the number of observations in level i, and error terms $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ where σ_ϵ^2 is also unknown, and α and ϵ are independent (Jiang, 2007). This can be structured in the form of equation (2.2) as

$$\mathbf{Y} = \mathbf{X}\beta + \sigma_1 \sum_{j=1}^{l_1} V_{1,j} z_{1,j} + \sigma_2 \sum_{j=1}^{n} V_{2,j} z_{2,j},$$

where $\beta = \mu$ is the overall mean, $\mathbf{X} = \mathbf{1}_n$ (an $n \times 1$ vector of ones), $l_1 = a$ is the number of levels for the first random effect and $V_{1,j}$ indicates which observations are in level j with random effect variance $\sigma_1^2 = \sigma_\alpha^2$. The second random effect corresponds to the error, and hence $V_{2,\cdot} = \mathbf{I}_n$ with σ_ϵ^2 as the error variance component. The $z_{1,\cdot}$ and $z_{2,\cdot}$ are i.i.d. standard normal random variables.

Example 2: Discretely observed Brownian motion model. Consider the model

$$Y_t = a + bt + \sigma W_t, \tag{2.4}$$

where a, b and σ are unknown parameters, and W_t is a discretely observed standard Brownian motion ($W(0) = 0$ and $W(s+t) - W(s) \sim N(0, t)$ for all $s, t \geq 0$) with discretely observed time points at $t = t_1, \ldots, t_n$ (i.e. $\{Y_t\}_{t \geq 0}$ is a Brownian motion with a drift). This design has been used to model many physical phenomena, and contributed to the development of the famous Black-Scholes model for option pricing. In the generalized fiducial framework, (2.4) is written as

$$\mathbf{Y} = \mathbf{X}\beta + \sigma \sum_{j=1}^{n} V_{1,j} z_{1,t_j},$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix},$$

$\beta = (a, b)^T$, and $V_{1,j}$ is a column vector with $j-1$ zeros followed by j terms equal to $\sqrt{t_j - t_{j-1}}$ for $j = 1, \ldots, n$ letting $t_0 = 0$.

The SMC algorithm presented in this section is seeking a weighted sample of particles $\{Z_{1:t}^{(J)}, W_{1:t}^{(J)}\}_{J=1}^{N}$ (where $W_{1:t}^{(J)}$ is the unnormalized importance weight for particle $Z_{1:t}^{(J)}$) from the generalized fiducial distribution of the unknown parameters in the normal mixed linear model. Once this sample of N weighted particles is obtained, inference procedures such as confidence intervals and parameter estimates can be performed on any of the unknown parameters or functions of parameters. For example, parameter estimates can be determined by taking a weighted average or the median of the particles with the associated (normalized) weights. A $C\%$ confidence interval can be found easily for each parameter by ordering the particles and finding the particle values θ_L and θ_U such that the sum of the normalized weights for the particles between θ_L and θ_U is $C\%$. The weighted approximation of the generalized fiducial distribution on the parameter space has a flexibility similar to a posterior distribution in the Bayesian setting.

2.3.1 Basic model

The proposed method is first introduced using a simple special case of (2.2), and then is generalized in the next section. The purpose of the presented algorithm is to obtain a sample from the generalized fiducial distribution of the unknown parameters. The basic model has the following structural equation $Y = \mu + \sigma Z$ where Y is the data vector, Z is a vector of i.i.d normal random variables, and μ and σ are unknown parameters. As discussed earlier, the data Y are not observed exactly, but rather the interval version $\mathbf{a} < Y \leq \mathbf{b}$ with $\mathbf{a}$ and $\mathbf{b}$, as $n \times 1$ vectors, determined by the level of uncertainty of the measurement (e.g. due to

the resolution of the instrument used). The idea is to generate Z^* such that

$$\mathbf{a} < \mu + \sigma Z^* \leq \mathbf{b}. \tag{2.5}$$

where Z^* are independent copies of Z. For the interval data $(\mathbf{a}, \mathbf{b}] = \{(a_i, b_i], i = 1, \ldots, n\}$, we want to generate z_i^* such that $a_i < \mu + \sigma z_i^* \leq b_i$ for $i = 1, \ldots, n$. The distribution on the parameter space will be derived from the set function $Q((\mathbf{a}, \mathbf{b}], Z^*) = \{(\mu, \sigma) : a_i < \mu + \sigma z_i^* \leq b_i, i = 1, \ldots, n\}$. Hence the generalized fiducial distribution – the target distribution of the simulation – can be defined as

$$V(Q((\mathbf{a}, \mathbf{b}], Z^*)) \mid \{Q((\mathbf{a}, \mathbf{b}], Z^*) \neq \emptyset\}. \tag{2.6}$$

We develop an SMC algorithm where one particle is one realization of Q from (2.6), and $\{Z_{1:t}^{(J)}, W_{1:t}^{(J)}\}_{J=1}^N$ is a collection of N weighted particles where $W_{1:t}^{(J)}$ is the unnormalized importance weight for particle $Z_{1:t}^{(J)} = (z_1^{(J)}, \ldots, z_t^{(J)})^T$. Note that for this basic model, each datum has only one random component and, hence, only one z_i, $i = 1, \ldots, t$, associated with it. We also note here that capital and unbolded Z is the vector, and lowercase z is a single realization; furthermore, though the $\mathbf{a}$ and $\mathbf{b}$ are lowercase and bold, they each represent a vector.

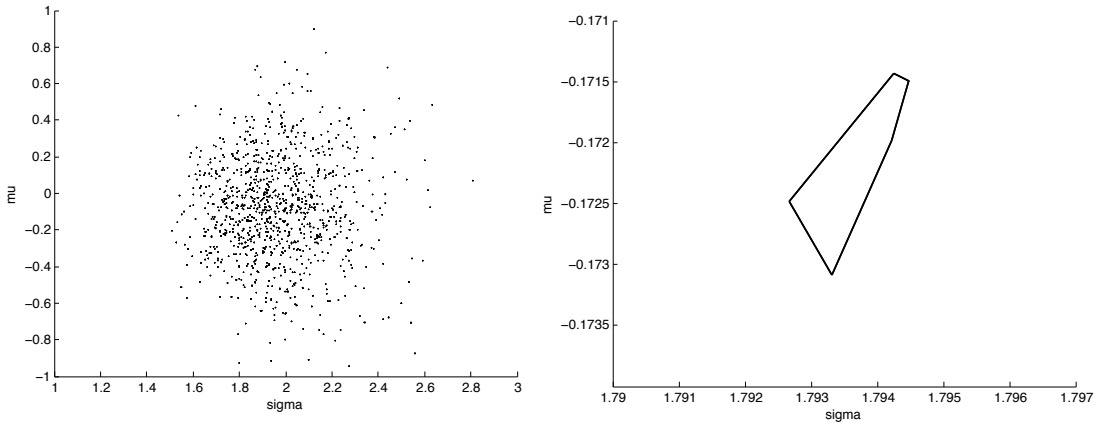
The actual algorithm for the basic model progresses by iteration between the steps listed in Table 2.1, and the steps are described in detail in this section.

Step	Description	Do
FOR	$J = 1, \ldots, N$	
0	Initialization	$t = 1, 2$ Draw $z_t^{(J)} \sim \pi_{1:2}$ (see Equation (2.7))
		Set $W_{1:2}^{(J)} = 1$
		Go to step 1
WHILE	$t \leq n$, $t > 2$	
1	Sampling	Draw $z_t^{(J)} \sim \widetilde{\pi}_{t\|1:t-1}$ (see Equation (2.8))
		Go to step 2
2	Calculate weights	$W_{1:t}^{(J)} = \pi_{1:t}(Z_{1:t}^{(J)}) / \widetilde{\pi}_{1:t}(Z_{1:t}^{(J)})$
		Go to step 3
3	Calculate ESS_t	See equation (1.6)
		If $\text{ESS}_t <$ threshold, go to step 4
		Else, set $t = t + 1$ and go to step 1
4	Resampling	Resample particle system
		Set $W_{1:t}^{(J)} = N^{-1}$
		Go to step 5
5	Alteration	Alter particle system according to Section 2.3.2
		Set $t = t + 1$ and go to step 1

Table 2.1: Algorithm for the basic model.

As an illustration of the basic algorithm, a sample of $N = 5000$ weighted particles were simulated from a data set generated from the basic model of (2.5) with $n = 20$, $\mu = 0$, and $\sigma = 2$. In order to maintain some sense of weight, or importance, of the generated particles, 1000 of the 5000 generated particles were randomly selected based on normalized importance weights and are displayed in Figure 2.1(a) with a single particle shown in Figure 2.1(b). Each particle is represented by a polygon in the figures indicating the possible values of μ and σ based on the sampled Z^*.

The proposal distribution used is the standard Cauchy distribution due to improved computational efficiency of sampling in the tails over the more natural choice of a standard normal



(a) Sample of 1000 particles

(b) One particle

Figure 2.1: (a) A sample of 5000 weighted particles were simulated from a data set generated from the basic model of (2.5) with $n = 20$, $\mu = 0$, and $\sigma = 2$. One thousand of the the 5000 generated particles were randomly selected based on weight and are displayed in this plot; (b) A single particle selected from (a).

distribution. Recalling our target distribution of (2.6), we want a sample $J = 1, \ldots, N$ of $Z_{1:t}^{(J)} | Z_{1:t}^{(J)} \in \mathbf{C}_t$, where $\mathbf{C}_t = \{Z_{1:t}^{(J)} : a_i < \mu + \sigma z_i^{(J)} \leq b_i, i = 1, \ldots, t\}$ is the region of $Z_{1:t}^{(J)}$ such that the set $Q_t^{(J)}$ is not empty ($Q_t^{(J)}$ is defined below). The target density at time t can be written as

$$\pi_{1:t}(Z_{1:t}^{(J)}) = \pi_{1:t}(Z_{1:t}^{(J)} | (\mathbf{a}, \mathbf{b}]_{1:t}) \propto \exp(-\sum_{i=1}^{t} (z_i^{(J)})^2/2) \cdot \mathbf{I}_{\mathbf{C}_t}(Z_{1:t}^{(J)}) \tag{2.7}$$

where $(\mathbf{a}, \mathbf{b}]_{1:t} = ((a_1, b_1], (a_2, b_2], \ldots, (a_t, b_t])$, $Z_{1:t}^{(J)} = (z_1^{(J)}, z_2^{(J)}, \ldots, z_t^{(J)})$ and $\mathbf{I}_{\mathbf{C}_t}(\cdot)$ is an indicator random variable for the set $\mathbf{C}_t$.

Define $Q_t^{(J)}(Z_{1:t}^{(J)}) = \{(\mu, \sigma) : a_i < \mu + \sigma z_i^{(J)} \leq b_i, i = 1, \ldots, t\}$ as the set containing the values of the parameters that satisfy equation (2.5) given the data and generated Z^* up to time t for particle J with $J = 1, \ldots, N$. In order to efficiently locate the regions of $z_t^{(J)}$ for which $Q_t^{(J)}$ is non-empty, define $m_t(Z_{1:t-1}^{(J)}) = \min(\frac{a_t - \mu}{\sigma},\ (\mu, \sigma) \in Q_{t-1}^{(J)}(Z_{1:t-1}^{(J)}))$ and $M_t(Z_{1:t-1}^{(J)}) = \max(\frac{b_t - \mu}{\sigma},\ (\mu, \sigma) \in Q_{t-1}^{(J)}(Z_{1:t-1}^{(J)}))$, which are the minimum and maximum possible values of $z_t^{(J)}$ that satisfy the inequalities of (2.5) up to time $t - 1$. That is, for this basic model, the values $m_t(Z_{1:t-1}^{(J)})$ and $M_t(Z_{1:t-1}^{(J)})$ are found easily by considering the vertices of the interior polygon at time $t - 1$, $Q_{t-1}^{(J)}$, and the lower and upper constraints on the new

data at time t, a_t and b_t.

The initial z_1 and z_2 are sampled from a standard normal distribution without restriction, then the remaining particles Z_t, $t > 2$, are sampled from a standard Cauchy truncated to $(m_t(Z_{1:t-1}^{(J)}), M_t(Z_{1:t-1}^{(J)}))$. Hence the proposal density at time t is

$$\widetilde{\pi}_{1:t}(Z_{1:t}|(\mathbf{a},\mathbf{b}]_{1:t}) = \widetilde{\pi}_{1:t}(Z_{1:t}) \propto \frac{\exp(-((z_1^{(J)})^2 + (z_2^{(J)})^2)/2) \prod_{i=3}^{t} [\mathbf{I}_{(m_i(Z_{1:t-1}^{(J)}), M_i(Z_{1:t-1}^{(J)}))}(z_i^{(J)})]}{\prod_{i=3}^{t} [(1 + (z_i^{(J)})^2)(F(M_i(Z_{1:t-1}^{(J)})) - F(m_i(Z_{1:t-1}^{(J)})))]},$$

where F is the cumulative distribution function of a standard Cauchy distribution. This also results in the conditional proposal distribution

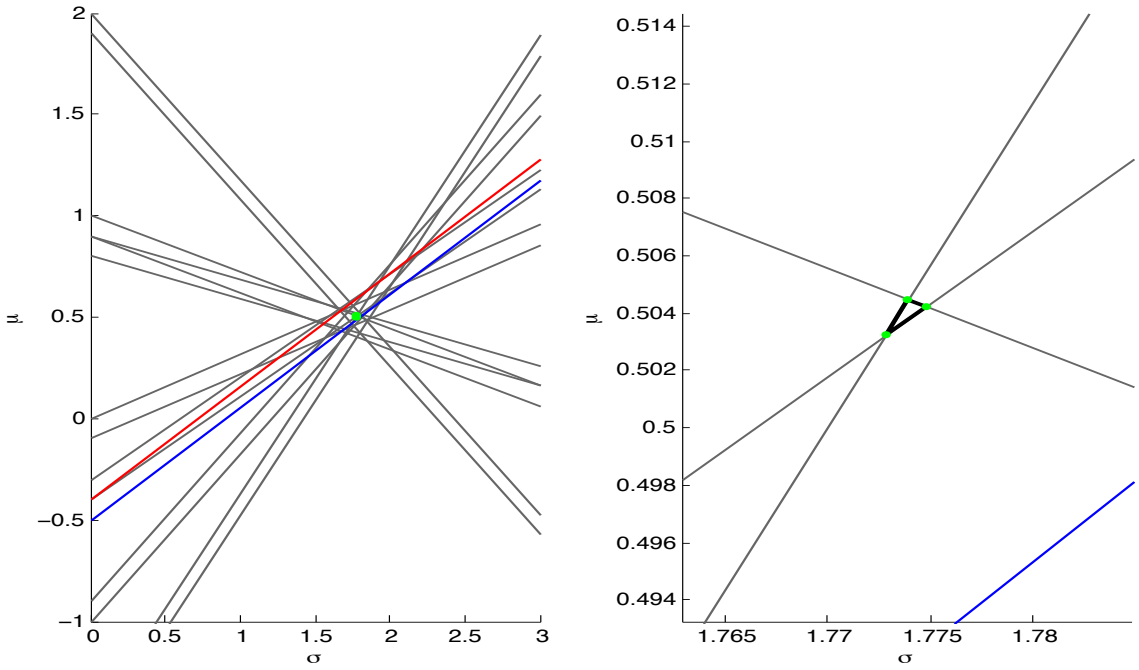
$$\widetilde{\pi}_{t|1:t-1}(z_t|(\mathbf{a},\mathbf{b}]_{1:t}) = \widetilde{\pi}_{t|1:t-1}(z_t) \propto \frac{\mathbf{I}_{(m_t(Z_{1:t-1}^{(J)}), M_t(Z_{1:t-1}^{(J)}))}(z_t^{(J)})}{(1 + (z_t^{(J)})^2)(F(M_t(Z_{1:t-1}^{(J)})) - F(m_t(Z_{1:t-1}^{(J)})))}. \quad (2.8)$$

At time t, the unnormalized importance weights are defined as $W_{1:t} = \pi_{1:t}(Z_{1:t})/\widetilde{\pi}_{1:t}(Z_{1:t})$, and are discussed further in Section 2.3.3.

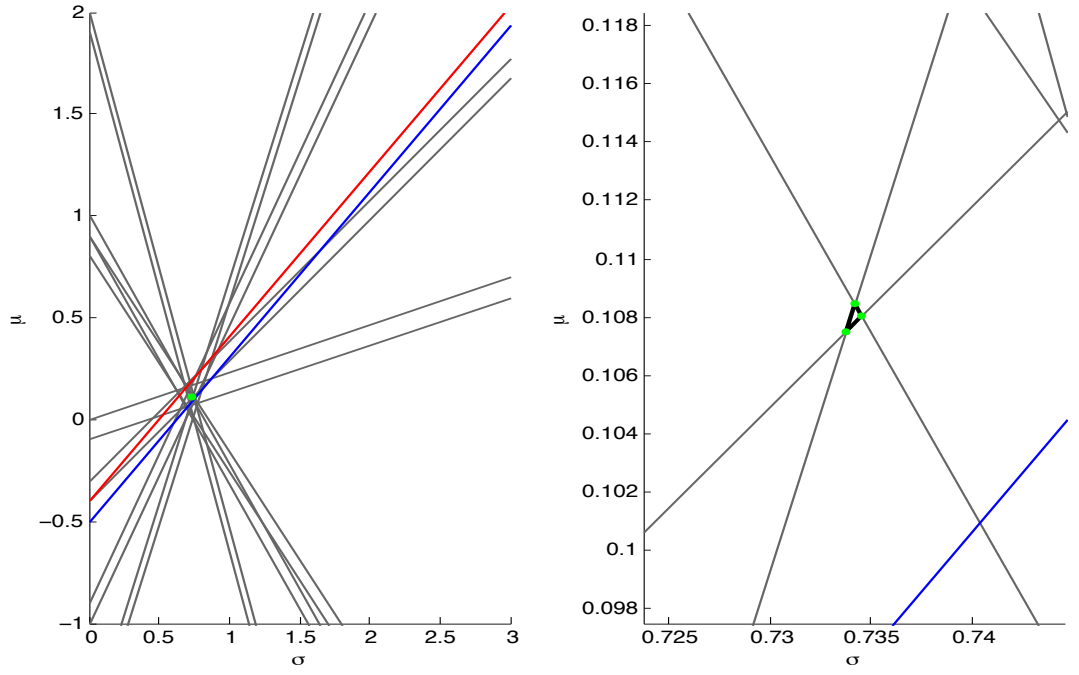
2.3.2 Basic alteration

Standard SMC resampling finds particles according to the distribution of weights at a given time step, copies the particle, and then assigns the resampled particles equal weight. By copying particles in this setting, it is unlikely we would end up with an appropriate distribution on the parameter space. Intuitively, this is because after each time step, each particle contains the set, or geometrically a polygon, of possible values of (μ, σ) given the generated $Z_{1:t}^{(J)}$, $J = 1, \ldots, N$. If the particles are copied, the distribution of polygons will be concentrated in a few areas due to particles with initially higher weight, and will not be able to move from those regions. Hence rather than copy the selected particles exactly, we alter them in a certain way in order to retain properties of heavy particles while still allowing for an appropriate sample of $\{Q_t^{(J)}\}_{J=1}^{N}$. This alteration is performed in such a way that it still solves the system of inequalities up to time t using some known properties of the sampled particles. Below we explain the alteration step for the simple model, but more details are presented in the discussion of the general model.

An illustration of the alteration step is found in Figure 2.2. Figure 2.2(a) displays the original particle selected to be altered, and Figure 2.2(b) displays the particle post-alteration.



(a) Original particle.



(b) Altered particle.

Figure 2.2: This is an illustration of the alteration step of the algorithm for the basic model. Plot (a) displays the original particle, and plot (b) is the result of the alteration step on the original particle. The red and blue lines represent the upper and lower bound on the fattened datum for the current time step. The right figures are close ups of the interior polygon of the figure to its left.

Suppose particle $Z^{(L)} = Z_{1:t}^{(L)}$ is to be copied k times at time step t (i.e. due on the weight of particle L, the resampling scheme designated this particle to be replicated k times). Define

$$Z^{(L)} = \mathbf{1}_t \cdot \bar{Z}^{(L)} + ||Z^{(L)} - \bar{Z}^{(L)}|| \cdot \frac{Z^{(L)} - \bar{Z}^{(L)}}{||Z^{(L)} - \bar{Z}^{(L)}||} = \mathbf{1}_t \cdot \mathbf{C} + \mathbf{D} \cdot \tau, \tag{2.9}$$

where $\bar{Z}^{(L)} = \frac{1}{t}\sum_{i=1}^{t} z_i^{(L)}$, $||\cdot||$ is the L^2 norm, $\mathbf{1}_t$ is a $t \times 1$ vector of ones, then $\mathbf{C} = \bar{Z}^{(L)}$, $\mathbf{D} = ||Z^{(L)} - \bar{Z}^{(L)}||$, and $\tau = \frac{Z^{(L)} - \bar{Z}^{(L)}}{||Z^{(L)} - \bar{Z}^{(L)}||}$. This decomposition of $Z^{(L)}$ breaks it into the orthogonal projection onto the vector $\mathbf{1}_t$ and the part that is orthogonal to $\mathbf{1_t}$, where $\mathbf{1_t}$ is the space of fixed effects (in this case, the overall mean). Furthermore, $\mathbf{C}$ and $\mathbf{D}$ are independent by design, and $\mathbf{C} \sim N(0,1)$ and $\mathbf{D} \sim \sqrt{\chi^2_{t-1}}$. The alteration will result from sampling a new $\mathbf{C}$ and a new $\mathbf{D}$ (denoted $\widetilde{\mathbf{C}}$ and $\widetilde{\mathbf{D}}$, respectively). The resampled $Z^{(L)}$ is thus $\widetilde{Z}^{(l)} = \mathbf{1}_t\widetilde{\mathbf{C}}^{(l)} + \widetilde{\mathbf{D}}^{(l)} \cdot \tau$, for $l = 1, \ldots, k-1$ (where k is the number of times the particle is resampled), and $\widetilde{\mathbf{C}}^{(k)} = \mathbf{C}$ and $\widetilde{\mathbf{D}}^{(k)} = \mathbf{D}$. Furthermore, the set $Q_t^{(L)} = \{(\mu, \sigma) : a_i < \mu + \sigma z_i^{(L)} \leq b_i, i = 1, \ldots, t\}$ can be adjusted noting that if (μ, σ) solves $a_i < \mu + \sigma\big(\mathbf{1}_t \cdot \mathbf{C} + \mathbf{D} \cdot \tau\big) \leq b_i$ for $i = 1, \ldots, t$, then $(\widetilde{\mu}^{(l)}, \widetilde{\sigma}^{(l)})$ can be found such that $a_i < \widetilde{\mu}^{(l)} + \widetilde{\sigma}^{(l)}\big(\mathbf{1}_t \cdot \widetilde{\mathbf{C}}^{(l)} + \widetilde{\mathbf{D}}^{(l)} \cdot \tau\big) \leq b_i$ for $i = 1, \ldots, t$ and $l = 1, \ldots, k-1$. The relation between (μ, σ) and $(\widetilde{\mu}, \widetilde{\sigma})$ is

$$\widetilde{\mu} = \mu + \sigma \cdot \mathbf{1}_t \cdot \big(\mathbf{C} - (\mathbf{D}/\widetilde{\mathbf{D}}) \cdot \widetilde{\mathbf{C}}\big) \tag{2.10}$$

$$\widetilde{\sigma} = \sigma \cdot \big(\mathbf{D}/\widetilde{\mathbf{D}}\big). \tag{2.11}$$

The relations (2.10) and (2.11) follow because $\mathbf{a} < \mu + \sigma Z^* \leq \mathbf{b}$ must hold for the altered particle. The desire is to find $(\widetilde{\mu}, \widetilde{\sigma})$ so that

$$\mu + \sigma Z = \mu + \sigma\big(\mathbf{1}_t \cdot \mathbf{C} + \tau\mathbf{D}\big) = \widetilde{\mu} + \widetilde{\sigma}\big(\mathbf{1}_t \cdot \widetilde{\mathbf{C}} + \tau\widetilde{\mathbf{D}}\big) = \widetilde{\mu} + \widetilde{\sigma}\widetilde{Z}.$$

Because the decomposition of (2.9) breaks the particle into the orthogonal projection onto $\mathbf{1_t}$ and a part that is orthogonal to $\mathbf{1_t}$, we can consider each piece independently. Considering the orthogonal parts first, $\sigma\tau\mathbf{D} = \widetilde{\sigma}\tau\widetilde{\mathbf{D}}$ imples that $\tau\big(\sigma\mathbf{D} - \widetilde{\sigma}\widetilde{\mathbf{D}}\big) = 0$. Hence, the updated $\widetilde{\sigma}$ can be found simply as $\widetilde{\sigma} = \sigma\big(\mathbf{D}/\widetilde{\mathbf{D}}\big)$.

The relation between $\widetilde{\mu}$ and μ follows from the remaining portion. We want $\mu + \sigma\big(\mathbf{1_t} \cdot \mathbf{C}\big) =$

$\widetilde{\mu} + \widetilde{\sigma}\big(\mathbf{1_t} \cdot \widetilde{\mathbf{C}}\big)$. It is then straightforward to see that $\widetilde{\mu} = \mu + \big(\sigma \mathbf{1}_t \cdot \mathbf{C} - \widetilde{\sigma}\mathbf{1}_t \cdot \widetilde{\mathbf{C}}\big)$, and so $\widetilde{\mu}$ is calculated as $\widetilde{\mu} = \mu + \sigma \mathbf{1}_t \cdot \big(\mathbf{C} - \mathbf{D}/\widetilde{\mathbf{D}} \cdot \widetilde{\mathbf{C}}\big)$.

2.3.3 General model

In this section, we generalize the ideas of the method presented in Section 2.3.1 for the general normal linear mixed model of Equation (2.2). Again, the goal of the algorithm is to obtain a generalized fiducial sample for the unknown parameters of normal mixed models. As discussed earlier, the data Y are not observed exactly, but rather intervals around the data are determined by the level of uncertainty of the measurement (e.g. due to the resolution of the instrument used). The structural equation formed as interval data for $t = 1, \ldots, n$ with $i = 1, \ldots, r$ random effects is

$$a_t < Y_t = X_t\beta + \sum_{i=1}^{r} \sigma_i \sum_{j=1}^{l_i} v_{i,j,t} z_{i,j} \leq b_t, \tag{2.12}$$

where the random effect design vector component $v_{i,j,t}$ indicates the j^{th} level of a random effect i for the t^{th} element of the data vector, X_t is the t^{th} row of the fixed effect design matrix $\mathbf{X}$, and is $z_{i,j}$ is a standard normal random variable for level j of random effect i. Each datum can have one or more random components and so we write $\mathbf{Z}_{1:t} = (Z_1, \ldots, Z_t)$ with capital letters to indicate the possible vector nature of each Z_k for $k = 1, \ldots, t$ with $\mathbf{Z}_{1:t}$ bolded. In the case that $r > 1$, $\mathbf{Z}_{1:t}$ is vectorized so that the mathematical operations make sense (i.e. $\mathbf{Z}_{1:t}$ is a $\sum_{i=1}^{r} l_i \times 1$ vector). Also, for notational convenience, Z_k will represent all $z_{i,j}$ for $i = 1, \ldots, r$ and $j = 1, \ldots, l_i$ *that are not present or shared with any datum less than* k, which will always at least include the error effect denoted $z_{k,r}$. It will be necessary at times to reference the non-error random effects, and they will be denoted $Z_{k,1:r-1}$ representing all the non-error random effects first introduced at time k. The goal is to generate a sample of the $z_{i,j}$ for $i = 1, \ldots, r$ and $j = 1, \ldots, l_i$ such that $a_t < Y_t \leq b_t$ for $t = 1, \ldots, n$.

Similar to Table 2.1, Table 2.2 displays the algorithm for the general model with the differences being the form of the particles, the proposal and target distributions, and the referenced sections and equations.

Step	Description	Do
FOR	$J = 1, \ldots, N$	
0	Initialization	$t = 1, \ldots, p + r$ Draw $\mathbf{Z}^{(J)}_{1:p+r,1:r} \sim \pi_{1:p+r}$ (see Equation (2.14))
		Set $W^{(J)}_{1:p+r} = 1$
		Go to step 1
WHILE	$t \leq n$, $t > p + r$	
1	Sampling	Draw $Z^{(J)}_t \sim \widetilde{\pi}_{t\|1:t-1}$ (see Equation (2.16))
		Go to step 2
2	Calculate weights	$W^{(J)}_{1:t} = \pi_{1:t}(\mathbf{Z}^{(J)}_{1:t})/\widetilde{\pi}_{1:t}(\mathbf{Z}^{(J)}_{1:t})$
		Go to step 3
3	Calculate ESS_t	See equation (1.6)
		If $\text{ESS}_t <$ threshold, go to step 4
		Else, set $t = t + 1$ and go to step 1
4	Resampling	Resample particle system
		Set $W^{(J)}_{1:t} = N^{-1}$
		Go to step 5
5	Alteration	Alter particle system according to Section 2.3.4
		Set $t = t + 1$ and go to step 1

Table 2.2: Algorithm for normal linear mixed models.

The generalized fiducial distribution on the parameter space is written as

$$V(Q((\mathbf{a}, \mathbf{b}], \mathbf{Z}))|\{Q((\mathbf{a}, \mathbf{b}], \mathbf{Z}) \neq \emptyset\}. \tag{2.13}$$

where we define the set function $Q^{(J)}_t$ as the set containing the values of the parameters that satisfy equation (2.12) given the data and generated $\mathbf{Z}^{(J)}_{1:t}$ up to time t for particle J, $J = 1, \ldots, N$. Generating a sample from (2.13) is equivalent to simulating the $\mathbf{Z}^{(J)}$ such that

the set $Q^{(J)}$ is non-empty, and this results in the *target distribution* at time t written as

$$\pi_{1:t}(\mathbf{Z}_{1:t}|(\mathbf{a},\mathbf{b}]_{1:t}) = \pi_{1:t}(\mathbf{Z}_{1:t}) \propto \exp(-(\mathbf{Z}_{1:t}^T \cdot \mathbf{Z}_{1:t})/2) \cdot \mathbf{I}_{\mathbf{C}_t}(\mathbf{Z}_{1:t}). \tag{2.14}$$

where $\mathbf{I}_{\mathbf{C}_t}(\cdot)$ is an indicator random variable for the set $\mathbf{C}_t$, with $\mathbf{C}_t$ as the set of $\mathbf{Z}_{1:t}$ such that Q_t is not empty. This is equivalent to $\mathbf{C}_t = \{\mathbf{Z}_{1:t} : a_k < X_k\beta + \sum_{i=1}^{r} \sigma_i \sum_{j=1}^{l_i} v_{i,j,k} z_{i,j} \leq b_k, k = 1, \ldots, t\}$. As with the basic model, the restriction that $Q_t^{(J)}$ is non-empty in the general setting can be translated into truncating the possible values of the particle corresponding to the error random effect to the interval defined by

$$\begin{aligned}
&m_t(Z_{1:t-1}^{(J)}, Z_{t,1:r-1}^{(J)}) = \\
&\qquad \min\left(\frac{a_t^{(J)} - \left(X_t\beta + \sum_{i=1}^{r-1} \sigma_i \sum_{j=1}^{l_i} v_{i,j,t} z_{i,j}^{(J)}\right)}{\sigma_r},\ (\beta,\sigma) \in Q_{t-1}^{(J)} \right), \\
&M_t(Z_{1:t-1}^{(J)}, Z_{t,1:r-1}^{(J)}) = \\
&\qquad \max\left(\frac{b_t^{(J)} - \left(X_t\beta + \sum_{i=1}^{r-1} \sigma_i \sum_{j=1}^{l_i} v_{i,j,t} z_{i,j}^{(J)}\right)}{\sigma_r},\ (\beta,\sigma) \in Q_{t-1}^{(J)} \right),
\end{aligned}$$

where that $\sigma = (\sigma_1, \ldots, \sigma_r)^T$. Verbally, $m_t(\mathbf{Z}_{1:t-1}^{(J)}, Z_{t,1:r-1}^{(J)})$ and $M_t(\mathbf{Z}_{1:t-1}^{(J)}, Z_{t,1:r-1}^{(J)})$ are the minimum and maximum possible values of $z_{t,r}^{(J)}$. As with the basic model, the proposal distribution uses the Cauchy distribution and the *full proposal distribution* at time t is

$$\begin{aligned}
\widetilde{\pi}_{1:t} &\propto \exp(-(\mathbf{Z}_{1:p+r}^T \cdot \mathbf{Z}_{1:p+r})/2) \times \\
&\prod_{i=p+r+1}^{t} \left(\frac{\exp(-(Z_{i,1:r-1}^T \cdot Z_{i,1:r-1})/2) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r})}{(1+z_{i,r}^2)(F(M_i) - F(m_i))} \right)
\end{aligned} \tag{2.15}$$

where the $\star$ in $\mathbf{I}_\star(\mathbf{Z}_{t,1:r-1})$ indicates the lack of restriction to a specific set of values for $\mathbf{Z}_{t,1:r-1}$,F is the cdf of the standard Cauchy distribution and $m_t = m_t(\mathbf{Z}_{1:t-1}, Z_{t,1:r-1})$ and $M_t = M_t(\mathbf{Z}_{1:t-1}, Z_{t,1:r-1})$. Note that the reference to $p+r$ (i.e. the dimension of the parameter space) in Equation (2.15) is because the first $p+r$ sampled $\mathbf{Z}$'s are not restricted and can be freely generated from a standard normal distribution ($p+r$ equations solve $p+r$ unknowns).

And the *conditional proposal distribution* at time t is

$$\widetilde{\pi}_{t|1:t-1} \propto \frac{\exp(-(Z_{t,1:r-1}^T \cdot Z_{t,1:r-1})/2) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r})}{(1+z_{r,t}^2)(F(M_t)-F(m_t))} \tag{2.16}$$

where F is the standard Cauchy cdf.

The proposed algorithm targets the generalized fiducial distribution of the unknown parameters of a normal mixed linear model of (2.13) displayed more explicitly in (2.14). Using the derivation of the proposal and target distributions above, and noting that

$$\mathbf{I}_{\mathbf{C}_t}(\mathbf{Z}_{1:t}) = \mathbf{I}_{\mathbf{C}_{t-1}}(\mathbf{Z}_{1:t-1}) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r}). \tag{2.17}$$

Equation (2.17) can easily be understood by considering the case when $r = 1$ (i.e. the basic model with only the error random effect) by noting that (i) if $\mathbf{Z}_{1:t} \in C_t$, then $\mathbf{Z}_{1:t-1} \in C_{t-1}$ and $m_t \le Z_t \le M_t$, and (ii) if $\mathbf{Z}_{1:t-1} \in C_{t-1}$ and $m_t \le Z_t \le M_t$, then $\mathbf{Z}_{1:t} \in C_t$. Then the marginal target distribution at time t is

$$\begin{aligned}
\widehat{\pi}_{1:t-1} &= \int \pi_{1:t}(\mathbf{Z}_{1:t})dZ_t = \int \frac{\exp(-(\mathbf{Z}_{1:t}^T \cdot \mathbf{Z}_{1:t})/2) \cdot \mathbf{I}_{\mathbf{C}_t}(\mathbf{Z}_{1:t})}{\Omega_{1:t}} dZ_t \\
&= \int \frac{\exp(-(\mathbf{Z}_{1:t}^T \cdot \mathbf{Z}_{1:t})/2) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r})}{\Omega_{1:t}} dZ_t \\
&= \frac{\exp(-(\mathbf{Z}_{1:t-1}^T \cdot \mathbf{Z}_{1:t-1})/2)}{\Omega_{1:t}} \int \exp(-(Z_t^T \cdot Z_t)/2) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r})dZ_t \\
&= \frac{\pi_{1:t-1} \cdot \Omega_{1:t-1}}{\Omega_{1:t}} \int \exp(-(Z_t^T \cdot Z_t)/2) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r})dZ_t \\
&\propto \pi_{1:t-1} \cdot (\Phi(M_t) - \Phi(m_t)),
\end{aligned} \tag{2.18}$$

where Φ is the standard normal cumulative distribution functions, and $\Omega_{1:t}$ is the normalization factor at time t. The extra $\Phi(M_t) - \Phi(m_t)$ term in (2.18) makes the *marginal* of the target distribution at time t different from the *target* distribution at time $t-1$, and contributes to the extra steps that are needed in the proof of Theorem 2.3.1. It then follows that

the *conditional target distribution* at time t is

$$\pi_{t|1:t-1} = \frac{\pi_{1:t}(\mathbf{Z}_{1:t})}{\widehat{\pi}_{1:t}(\mathbf{Z}_{1:t})} \propto \frac{\exp(-(Z_t^T \cdot Z_t)/2) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r})}{\Phi(M_t) - \Phi(m_t)}.$$

Finally, the derivation of the weights at time t is

$$W_{1:t} = \frac{\pi_{1:t}}{\widetilde{\pi}_{1:t}} = \frac{\pi_{t|1:t-1} \cdot \widehat{\pi}_{1:t-1}}{\widetilde{\pi}_{t|1:t-1} \cdot \widetilde{\pi}_{1:t-1}} \propto$$
$$\exp(-z_{t,r}^2/2)(1 + z_{t,r}^2)(F(M_t) - F(m_t)) \cdot \frac{\pi_{1:t-1}}{\widetilde{\pi}_{1:t-1}} \propto W_t \cdot W_{1:t-1},$$

where W_t is the sequential update factor.

2.3.4 General alteration

The same issue with resampling addressed for the basic model holds in this general setting: due to the nature of the generalized fiducial framework and the inference goals, the particles selected to be resampled cannot simply be copied. By copying particles in this setting, we would not end up with an appropriate distribution on the parameter space because all the sets $Q^{(J)}$, $J = 1, \ldots, N$, would end up in the same location. An extra step to alter the particles must be incorporated into the algorithm after reampling. If the particles are simply copied, the distribution of the sets $Q^{(J)}$, $J = 1, \ldots, N$, or geometrically the polyhedrons, will be concentrated in a few areas on the parameter space due to particles with initially higher weight. There would be no way to move the particles from those regions because subsequent particles would continue to define subsets of the copied polyhedrons as outlined in the algorithm presented above. The alteration procedure described below is a conceptually similar to the alteration step presented in Section 2.3.2 for the basic model: the particle to be resampled is decomposed into an orthogonal projection onto a certain space and a part orthogonal to that space, and then distributional properties that arise from the decomposition are used to alter the resampled particle. The procedure can be thought of as a Gibbs-sampling step in a non-coordinate direction determined by the selected particle.

We begin by defining terms that will be used in describing the methodology. Consider

some subspace $\mathcal{A} \subset \mathbb{R}^t$ with orthogonal basis vectors $\mathbf{V} = \{V_1, \ldots, V_d\}$. Any vector $Y \in \mathbb{R}^t$ can be written in terms of a projection onto $\mathcal{A}$ plus a projection onto $\mathcal{A}^\perp$ (the subspace orthogonal to $\mathcal{A}$). That is, for any $Y \in \mathbb{R}^t$, $Y = \Pi_{\mathcal{A}} Y + \Pi_{\mathcal{A}^\perp} Y$ where $\Pi_{\mathcal{A}} Y \in \mathcal{A}$ is the projection of Y onto $\mathcal{A}$, and $\Pi_{\mathcal{A}^\perp} Y \in \mathcal{A}^\perp$ with $\Pi_{\mathcal{A}^\perp} Y = Y - \Pi_{\mathcal{A}} Y$. The *projection* of Y onto $\mathcal{A}$ is

$$\Pi_{\mathcal{A}} Y = \frac{Y \cdot V_1}{V_1 \cdot V_1} V_1 + \ldots + \frac{Y \cdot V_d}{V_d \cdot V_d} V_d = \mathbf{V}(\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T Y.$$

Of course if the basis vectors are orthonormal ($\mathbf{V}^T \mathbf{V} = \mathbf{I}_d$), then projection of Y onto $\mathcal{A}$ can be written $\Pi_{\mathcal{A}} Y = \mathbf{V}\mathbf{V}^T Y$. The *null space* of a $t \times d$ matrix $\mathbf{A}$ is defined as $\mathcal{N}(\mathbf{A}) = \{X : X \in \mathbf{R}^d \text{ and } \mathbf{A}X = \mathbf{0}\}$. Also note that the $\text{rank}(\mathbf{A}) + \dim(\mathcal{N}(\mathbf{A})) = d$.

The alteration step of the proposed algorithm is performed in such a way that it still solves the system of inequalities of (2.12) up to time t using the following idea (to ease the notational complexity, we do not include the dependence of the variables on t). Suppose particle L is selected to be resampled (for an L between 1 and N). For each random effect, e, let

$$Y = \mathbf{X}'\beta' + \sigma \mathbf{V}_e Z^{(L)},$$

where $\mathbf{X}' = [\mathbf{X}, \{\sum_{j=1}^{l_i} V_{i,j} z_{i,j}^{(L)}\}_{i \neq e}]$, $\beta' = (\beta, \{\sigma_i\}_{i \neq e})'$, $\sigma = \sigma_e$, and $\mathbf{V}_e Z^{(L)} = \sum_{j=1}^{l_e} V_{e,j} z_{e,j}^{(L)}$. That is, other than random effect e, the random effects design matrices and sampled values are concatenated with the fixed effects.

In order to alter $Z^{(L)}$, we first find the basis vectors, η, for the null space of the matrix $\mathbf{A} = [-\mathbf{X}', \mathbf{V}_e]$. Let the basis vectors of $\mathcal{N}(\mathbf{A}) = \{\eta : \mathbf{A}\eta = \mathbf{0}\}$. Furthermore, let $\eta = (\eta_1, \eta_2)^T$ such that (i) $\mathbf{X}' \cdot \eta_1 + \mathbf{V}_e \cdot \eta_2 = \mathbf{0}$, and (ii) $\eta_2^T \cdot \eta_2 = \mathbf{I}$ (i.e. η_2 is orthonormal). The matrix $\mathbf{A}$ is defined as $[-\mathbf{X}', \mathbf{V}_e]$ because we want the alteration to be such that, after resampling the particles according to the decomposition defined below in Equation (2.19), the result of the changes in the "fixed" (i.e. not random effect e parameters) and changes in σ to have a canceling effect. Since $-\mathbf{X}'\eta_1 + \mathbf{V}_e\eta_2 = \mathbf{0}$, working with the basis for the null space of $\mathbf{A}$ will have that desired consequence.

We perform the following decomposition of the particle $Z^{(L)}$ selected to be resampled:

$$Z^{(L)} = \Pi Z^{(L)} + ||Z^{(L)} - \Pi Z^{(L)}|| \cdot \frac{Z^{(L)} - \Pi Z^{(L)}}{||Z^{(L)} - \Pi Z^{(L)}||} \tag{2.19}$$

where $\Pi Z^{(L)}$ is the projection onto the null space $\mathcal{N}(\mathbf{A})$ (i.e. $\Pi Z^{(L)} = \eta_2 \cdot \eta_2^T \cdot Z^{(L)}$), and $||\cdot||$ is the L^2 norm. Define $\mathbf{C} = \eta_2^T \cdot Z^{(L)}$ (so that, $\eta_2 \cdot \mathbf{C} = \Pi Z^{(L)}$), $\mathbf{D} = ||Z^{(L)} - \Pi Z^{(L)}||$, and $\tau = \frac{Z^{(L)} - \Pi Z^{(L)}}{||Z^{(L)} - \Pi Z^{(L)}||}$ so that $Z^{(L)} = \eta_2 \mathbf{C} + \mathbf{D} \cdot \tau$. Then if $Z^{(L)}$ is standard normal, $\mathbf{C} \sim N_{l_e}(0, \mathbf{I})$ where l_e is the number of levels of random effect e of $Z^{(L)}$ to be resampled at time t, and $\mathbf{D} \sim \sqrt{\chi^2_{l_e - d}}$ where $d = \text{rank}(\mathcal{N}(\mathbf{A}))$; $\mathbf{C}$ and $\mathbf{D}$ are independent by design. The alteration of $Z^{(L)}$ is accomplished by sampling new values of $\mathbf{C}$ and $\mathbf{D}$ (denoted $\widetilde{\mathbf{C}}$ and $\widetilde{\mathbf{D}}$, respectively) according to their distributions determined by the decomposition of Equation (2.19), and the altered particle is

$$\widetilde{Z} = \eta_2 \cdot \widetilde{\mathbf{C}} + \widetilde{\mathbf{D}} \cdot \tau. \tag{2.20}$$

Notice that if $Z^{(L)}$ is a standard normal conditioned on $\mathbf{C}_t$, then so is $\widetilde{Z}$.

Furthermore, the set $Q^{(L)} = \{(\beta', \sigma) : a < \mathbf{X}'\beta' + \sigma \mathbf{V}_Z^{(L)} \leq b\}$ can be easily updated by noting that if (β', σ) solves $a < \mathbf{X}'\beta' + \sigma \mathbf{V}_e(\eta_2 \cdot \mathbf{C} + \mathbf{D} \cdot \tau) \leq b$, then $(\widetilde{\beta}', \widetilde{\sigma})$ can be found such that $a < \mathbf{X}'\widetilde{\beta}' + \widetilde{\sigma} \mathbf{V}_e(\eta_2 \cdot \widetilde{\mathbf{C}} + \widetilde{\mathbf{D}} \cdot \tau) \leq b$ by considering $\mathbf{X}'\beta' + \sigma \mathbf{V}_e Z^{(L)} = \mathbf{X}\beta' + \sigma \mathbf{V}_e\big(\eta_2 \cdot \mathbf{C} + \tau \mathbf{D}\big) = \mathbf{X}'\widetilde{\beta}' + \widetilde{\sigma} \mathbf{V}_e\big(\eta_2 \cdot \widetilde{\mathbf{C}} + \tau \widetilde{\mathbf{D}}\big) = \mathbf{X}'\widetilde{\beta}' + \widetilde{\sigma} \mathbf{V} \widetilde{Z}$.

Examining the orthogonal parts first, $\sigma \mathbf{V}_e(\tau \mathbf{D}) = \widetilde{\sigma} \mathbf{V}_e(\tau \widetilde{\mathbf{D}})$ implies that $\mathbf{V}\tau\big(\sigma \mathbf{D} - \widetilde{\sigma} \widetilde{\mathbf{D}}\big) = 0$ so that

$$\widetilde{\sigma} = \sigma\big(\mathbf{D}/\widetilde{\mathbf{D}}\big).$$

The relation between $\widetilde{\beta}$ and β' follows from the remaining portion as

$$\begin{aligned} \mathbf{X}'\beta' + \sigma \mathbf{V}_e\big(\eta_2 \cdot \mathbf{C}\big) = \mathbf{X}'\widetilde{\beta}' + \widetilde{\sigma} \mathbf{V}_e\big(\eta_2 \cdot \widetilde{\mathbf{C}}\big) \text{ implies} \\ \mathbf{X}'(\widetilde{\beta}' - \beta') + \sigma \mathbf{V}_e \eta_2\big(\widetilde{\mathbf{C}} \cdot \mathbf{D}/\widetilde{\mathbf{D}} - \mathbf{C}\big) = 0. \end{aligned} \tag{2.21}$$

Noting by definition $-\mathbf{X}'\eta_1 + \mathbf{V}_e\eta_2 = 0$, then

$$-\mathbf{X}'\sigma\eta_1\big(\widetilde{\mathbf{C}} \cdot \mathbf{D}/\widetilde{\mathbf{D}} - \mathbf{C}\big) + \sigma \mathbf{V}_e \eta_2\big(\widetilde{\mathbf{C}} \cdot \mathbf{D}/\widetilde{\mathbf{D}} - \mathbf{C}\big) = 0. \tag{2.22}$$

By combining (2.21) and (2.22), we see that $\widetilde{\beta} - \beta' = -\sigma\eta_1 \cdot \big(\widetilde{\mathbf{C}} \cdot \mathbf{D}/\widetilde{\mathbf{D}} - \mathbf{C}\big)$, and hence

$$\widetilde{\beta} = \beta' - \sigma\eta_1 \cdot \big(\widetilde{\mathbf{C}} \cdot \mathbf{D}/\widetilde{\mathbf{D}} - \mathbf{C}\big).$$

Hence the sets $Q_t^{(J)}(Z_{1:t}^{(J)})$ are easily updated to $Q_t^{(J)}(\widetilde{Z}_{1:t}^{(J)})$. This procedure is repeated for each random effect $e = 1, \ldots, r$.

2.3.5 Convergence of algorithm

The proposed algorithm targets the generalized fiducial distribution of the unknown parameters of a normal linear mixed model of (2.13) displayed more explicitly in (2.14). The following theorem confirms that the weighted particle system from the proposed algorithm achieves consistency of the weighted particles as the particle sample size N approaches infinity. One of the difficulties with convergence for the weighted particles of an SMC algorithm is due to the heavy interaction, and thus dependence, of the particles throughout the algorithm. Douc and Moulines (2008) provide a nice general result for the consistency of weighted particle systems, which were adapted and applied to work in our generalized fiducial setting. We begin by stating the theorem, then discuss more details about our generalized fiducial methodology, present a brief introduction to the relevant ideas of Douc and Moulines (2008), and finish this section with the proof of the theorem.

Theorem 2.3.1. *Given a weighted sample* $\{\mathbf{Z}_{1:n}^{(J)}, W_{1:n}^{(J)}\}_{J=1}^{N}$ *obtained using the algorithm presented above targeting* (2.14)*, then for any bounded, measurable function* f*,*

$$\Big(\sum_{I=1}^{N} W_{1:n}^{(I)}\Big)^{-1} \sum_{J=1}^{N} f(\mathbf{Z}_{1:n}^{(J)}) W_{1:n}^{(J)} \xrightarrow{P} \int f(\mathbf{Z}_{1:n})\pi_{1:n} = \pi_{1:n} f(\mathbf{Z}_{1:n}), \text{ as } N \longrightarrow \infty.$$

This result holds for slightly weaker conditions, which are addressed below in the proof. The proof of the convergence of the proposed SMC algorithm follows from ideas presented in Douc and Moulines (2008). Theorem 2.3.1 will follow from proving the convergence of the generated particles after each stage of the algorithm: sampling, resampling, and alteration. The development of the particle system using the proposed algorithm does not follow the

traditional SMC algorithm as presented in Douc and Moulines (2008, Section 2), and requires some extra steps to validate the result.

Notation and definitions The proof uses the full, conditional and marginal proposal and target distributions derived earlier in this section, and the following notation and definitions. A particle system is defined as $\{\mathbf{Z}_{1:t}^{(J)}, W_{1:t}^{(J)}\}_{J=1}^{N}$ with $\mathbf{Z}_{1:t}^{(J)}$ sampled from the proposal distribution $\widetilde{\pi}_{1:t}$ of Equation (2.15) and un-normalized weights $W_{1:t}^{(J)}$ with normalizing factor at time t defined as $\Omega_t = \sum_{J=1}^{N} W_{1:t}^{(J)}$. The following definitions are from Douc and Moulines (2008) with modification for our notation and proposed algorithm. First, we define two sigma-fields $\mathcal{F}_0 \triangleq \sigma(\{\mathbf{Z}_{1:t-1}^{(J)}\}_{J=1}^{N}, (a,b]_{1:t})$ and $\mathcal{F}_J \triangleq \mathcal{F}_0 \vee \sigma(\{\mathbf{Z}_{1:t}^{(K)}\}_{1\leq K\leq J}, (a,b]_{1:t})$, for $J = 1, \ldots, N$. That is, we have a sigma field of all the generated particles up to time $t-1$ and all the data up to time t, and a triangular array of sigma fields of the particles generated up to time t after sampling Z_t from the conditional proposal distribution along with the data to time t. Though the sigma fields are dependent on t, we suppress the notation because the focus is on the transition from $t-1$ to t for sampling, and then at t for resampling and alteration. Specifically, the current state of the particle system will be at time $t-1$ and the concern is for what happens going from time $t-1$ to time t. At time $t-1$, the particle system is targeting probability measure $\pi_{1:t-1}$ on $(\Theta_{1:t-1}, \mathcal{B}(\Theta_{1:t-1}))$. This measure is approximated by our particle system, $\{\mathbf{Z}_{1:t-1}^{(J)}, W_{1:t-1}^{(J)}\}_{J=1}^{N}$. Furthermore, $\mathbf{Z}_{1:t-1}^{(J)} \in \Theta_{1:t-1}$ and $W_{1:t-1}^{(J)} \geq 0$ such that $\Omega_{t-1} \sum_{J=1}^{N} W_{1:t-1}^{(J)} = 1$. Note that π will, with the appropriate subscript indexing the time, be used to denote both the measure and density. The importance sampling step transforms the current particle system to one that is targeting $\pi_{1:t}$ on $(\Theta_{1:t}, \mathcal{B}(\Theta_{1:t}))$, denoted by $\{\mathbf{Z}_{1:t}^{(J)}, W_{1:t}^{(J)}\}_{J=1}^{N}$.

We will also use the notion of a *proper* set from (Douc and Moulines, 2008, Section 2.1). They define a proper set to be a subset B of a general state space X where (i) B is a linear space (closed under addition and scalar multiplication), (ii) if some function $g \in B$ with measurable function f such that $|f| \leq |g|$, then it follows that $|f| \in B$, and (iii) all constants

are in B. We define the proper set $B_t \triangleq \{f \in L^1(\Theta_{1:t}, \pi_{1:t}), F(\cdot, |f|) \in B_{t-1}\}$, where

$$F(\mathbf{Z}_{1:t-1}, f) = \int_{\Theta_t} f(\mathbf{Z}_{1:t-1}, Z_t) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t, M_t)}(z_{t,r})(\Phi(M_t) - \Phi(m_t))\pi_{t|1:t-1}(dZ_t) \tag{2.23}$$

and can be thought of as the transition kernel from time $t-1$ to t.

The following definition is from Douc and Moulines (2008), which defines the notion of weighted sample consistency that will be used in the proof.

Definition 2.3.1. *A weighted sample* $\{\mathbf{Z}_{1:t}^{(J)}, W_{1:t}^{(J)}\}_{J=1}^N$ *is* <u>consistent</u> *for the probability measure* $\pi_{1:t}$ *and the proper set* B_t *if, for any* $f \in B_t$, $\Omega_t^{-1} \sum_{J=1}^N W_{1:t}^{(J)} f(\mathbf{Z}_{1:t}^{(J)}) \xrightarrow{P} \int f(\mathbf{Z}_{1:t})\pi_{1:t}(d\mathbf{Z}_{1:t}) \triangleq \pi_{1:t}(f)$, *and* $\Omega_t^{-1} \max_{J=1}^N W_{1:t}^{(J)} \xrightarrow{P} 0$.

We see that this definition includes a weighted law of large numbers for the particle system, and a restriction on the asymptotic contribution of the individual weights.

The theorem needed from Douc and Moulines (2008) is reproduced below as Theorem 2.3.2 for the convenience of the reader with minor modifications for our notation. Note that $\{U_{N,J}\}_{J=1}^N$ is a triangular array of random variables, and will be used in the proof as our weighted particle system. A triangular array of sub-sigma field of the general sigma field $\mathcal{F}$ is defined as $\{\mathcal{F}_J\}_{J=1}^N$ such that (i) $\mathcal{F}_{J-1} \subset \mathcal{F}_J$, and (ii) $U_{N,J}$ is measurable with respect to $\mathcal{F}_J$. We note that ideas of our proof follow from ideas from the proofs of Douc and Moulines (2008) adapted for our setting.

Theorem 2.3.2. (Douc and Moulines, 2008, Theorem A.1)
Assume that $E[|U_{N,J}||\mathcal{F}_{J-1}] < \infty$ *P-a.s. for any* N *and any* $J = 1, \ldots, N$, *and*

$$\sup_N P\big(\sum_{J=1}^N E[|U_{N,J}||\mathcal{F}_{J-1}] \geq \lambda\big) \to 0 \text{ as } \lambda \to \infty \tag{2.24}$$

$$\sum_{J=1}^N E[|U_{N,J}| \cdot \mathbf{1}\{|U_{N,J}| \geq \epsilon\}|\mathcal{F}_{J-1}] \xrightarrow{P} 0 \text{ for any } \epsilon > 0 \tag{2.25}$$

Then, $\max_{1 \leq I \leq N} |\sum_{J=1}^I U_{N,J} - \sum_{J=1}^I E[U_{N,J}|\mathcal{F}_{J-1}]| \xrightarrow{P} 0$.

Sampling The first step in proving the asymptotic consistency of the algorithm is to show that the importance sampling scheme is asymptotically consistent, which is the statement of Lemma 2.3.1. The importance sampling step in the algorithm maintains the following relation. Given $f \in B_t$,

$$\begin{aligned} E(W_{1:t}^{(J)} f(\mathbf{Z}_{1:t}^{(J)})|\mathcal{F}_{J-1}) &= W_{1:t-1}^{(J)} E(W_t^{(J)} f(\mathbf{Z}_{1:t}^{(J)})|\mathcal{F}_{J-1}) \\ &= W_{1:t-1}^{(J)} \cdot \int f(\mathbf{Z}_{1:t}^{(J)}, Z_t) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r}) W_t \widetilde{\pi}_{t|1:t-1}(dZ_t) \\ &= W_{1:t-1}^{(J)} \cdot \int f(\mathbf{Z}_{1:t}^{(J)}, Z_t) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r}) (\Phi(M_t^{(J)}) - \Phi(m_t^{(J)})) \pi_{t|1:t-1}(dZ_t) \\ &= W_{1:t-1}^{(J)} \cdot F(\mathbf{Z}_{1:t-1}, f), \end{aligned} \tag{2.26}$$

for $J = 1, \ldots, N$.

Lemma 2.3.1. *(Sampling) Assume that the weighted sample $\{\mathbf{Z}_{1:t-1}^{(J)}, W_{1:t-1}^{(J)}\}_{J=1}^N$ is consistent for $(\pi_{1:t-1}, B_{t-1})$ and that $F(\Theta_{1:t}, \cdot)$ belongs to B_{t-1} where F is defined in Equation (2.23). Then the set B_t defined above is a proper set, and for $f \in B_t$, the weighted sample $\{\mathbf{Z}_{1:t}^{(J)}, W_{1:t}^{(J)}\}_{J=1}^N$ defined above is consistent for $(\pi_{1:t}, B_t)$.*

Proof. First to show B_t is a proper set. If a and b are constants and functions $f, g \in B_t$, then $af + bg \in B_t$ since $f, g \in B_t$ implies that $\int |f| d\pi_{1:t} < \infty$ and $\int |g| d\pi_{1:t} < \infty$. It follows that $\int |af + bg| d\pi_{1:t} < |a| \int |f| d\pi_{1:t} + |b| \int |g| d\pi_{1:t} < \infty$ by the triangle inequality. Next, if function $g \in B_t$ with measurable f such that $|f| < |g|$, then $\int |f| d\pi_{1:t} < \int |g| d\pi_{1:t} < \infty$, so $f \in B_t$. Because $\pi_{1:t}(\Theta_{1:t}) = 1$, all constants $\in B_t$.

Because $\{\mathbf{Z}_{1:t-1}^{(J)}, W_{1:t-1}^{(J)}\}_{J=1,\ldots,N}$ is consistent for $(\pi_{1:t-1}, B_{t-1})$, $f \in B_t$, and the function

$F(\cdot, f) \in B_{t-1}$, then for any $f \in B_t$,

$$\begin{aligned}
&\Omega_{t-1}^{-1} \sum_{J=1}^{N} E[f(\mathbf{Z}_{1:t}^{(J)}) W_{1:t}^{(J)} | \mathcal{F}_{J-1}] \\
&\quad = \Omega_{t-1}^{-1} \sum_{J=1}^{N} W_{1:t-1}^{(J)} F(\mathbf{Z}_{1:t-1}^{(J)}, f) \text{ by Equation (2.26)} \\
&\quad \longrightarrow \int \cdots \int f(\mathbf{Z}_{1:t}) \mathbf{I}_\star(\mathbf{Z}_{t,1:r-1}) \mathbf{I}_{(m_t, M_t)}(z_{t,r}) \pi_{t|1:t-1}(dZ_t) \pi_{1:t-1}(d\mathbf{Z}_{1:t-1}) \\
&\quad \triangleq \pi_{1:t-1} F(\cdot, f), \qquad (2.27)
\end{aligned}$$

where the last step follows because $\{\mathbf{Z}_{1:t-1}^{(J)}, W_{1:t-1}^{(J)}\}_{J=1,\ldots,N}$ is consistent for $(\pi_{1:t-1}, B_{t-1})$, and $F(\cdot, f) \in B_{t-1}$.

We next show the following holds

$$\max_{1 \le I \le N} \left| \Omega_{t-1}^{-1} \sum_{J=1}^{I} \{ W_{1:t}^{(J)} f(\mathbf{Z}_{1:t}^{(J)}) - E[f(\mathbf{Z}_{1:t}^{(J)}) W_{1:t}^{(J)} | \mathcal{F}_{J-1}] \} \right| \xrightarrow{P} 0. \qquad (2.28)$$

by appealing to Theorem 2.3.2. Beginning with Condition (2.24), let $U_{N,J} = \Omega_{t-1}^{-1} W_{1:t}^{(J)} f(\mathbf{Z}_{1:t}^{(J)})$ for $J = 1, \ldots, N$, then

$$\begin{aligned}
&\sum_{J=1}^{N} E[|U_{N,J}||\mathcal{F}_{J-1}] \\
&\quad = \Omega_{t-1}^{-1} \sum_{J=1}^{N} E(|W_{1:t}^{(J)} f(\mathbf{Z}_{1:t}^{(J)})||\mathcal{F}_{J-1}) \\
&\quad = \Omega_{t-1}^{-1} \sum_{J=1}^{N} W_{1:t-1}^{(J)} F(\mathbf{Z}_{1:t-1}, |f|) \xrightarrow{P} \pi_{1:t-1} F(\cdot, |f|). \qquad (2.29)
\end{aligned}$$

The tightness condition, Condition (2.24), follows because $F \in L^1(\Theta_{1:t-1}, \pi_{1:t-1})$.

Next we show that Condition (2.25) holds. For any $\epsilon > 0$, consider $\sum_{J=1}^{N} E[|U_{N,J}| \mathbf{1}\{|U_{N,J}| \ge \epsilon\} | \mathcal{F}_{N,J-1}]$, where $U_{N,J} = \Omega_{t-1}^{-1} W_{1:t}^{(J)} f(\mathbf{Z}_{1:t}^{(J)})$ for $J = 1, \ldots, N$ (same as above). Define $R(\mathbf{Z}_{1:t-1}, f) \triangleq \int f(\mathbf{Z}_{1:t-1}, Z_t) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t, M_t)}(z_{t,r}) d\widetilde{\pi}_{t|1:t-1}(Z_t)$, which is analogous to $F(\mathbf{Z}_{1:t-1}, f)$, but the integral is with respect to the proposal measure. For $K > 0$, $R(\mathbf{Z}_{1:t-1}, W_t \cdot |f| \mathbf{1}\{W_t \cdot |f| \ge K\}) \le R(\mathbf{Z}_{1:t-1}, W_t \cdot |f|) = F(\mathbf{Z}_{1:t-1}, |f|)$ because

$W_t\widetilde{\pi}_{t|1:t-1} = (\Phi(M_t) - \Phi(m_t))\pi_{t|1:t-1}$ so that $R(\mathbf{Z}_{1:t-1}, W_t \cdot |f| \cdot \mathbf{1}\{W_t \cdot |f| \geq K\}) \in B_{t-1}$. For $\epsilon > 0$,

$$
\begin{aligned}
&\sum_{J=1}^{N} E[|U_{N,J}|\mathbf{1}\{|U_{N,J}| \geq \epsilon\}|\mathcal{F}_{J-1}] \cdot \mathbf{1}\{\Omega_{t-1}^{-1} \max_{1\leq I\leq N} W_{1:t-1}^{(I)} \leq \epsilon/K\} \\
&= \sum_{J=1}^{N} E[|\Omega_{t-1}^{-1} W_{1:t}^{(J)} \cdot f(\mathbf{Z}_{1:t}^{(J)})| \cdot \mathbf{1}\{\Omega_{t-1}^{-1} W_{1:t}^{(J)} \cdot |f(\mathbf{Z}_{1:t}^{(J)})| \geq \epsilon\}|\mathcal{F}_{J-1}] \\
&\qquad\qquad \times \mathbf{1}\{\Omega_{t-1}^{-1} \max_{1\leq I\leq N} W_{1:t-1}^{(I)} \leq \epsilon/K\} \\
&= \Omega_{t-1}^{-1} \sum_{J=1}^{N} W_{1:t-1}^{(J)} \cdot \int W_t^{(J)} \cdot |f(\mathbf{Z}_{1:t-1}^{(J)}, Z_t)| \cdot \mathbf{1}\{\Omega_{t-1}^{-1} W_{1:t}^{(J)} \cdot |f| \geq \epsilon\} \\
&\qquad\qquad \times \mathbf{1}\{\Omega_{t-1}^{-1} \max_{1\leq I\leq N} W_{1:t-1}^{(I)} \leq \epsilon/K\} \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r}) d\widetilde{\pi}_{t|1:t-1}(Z_t) \\
&\leq \Omega_{t-1}^{-1} \sum_{J=1}^{N} W_{1:t-1}^{(J)} \cdot \int W_t^{(J)} \cdot |f(\mathbf{Z}_{1:t-1}^{(J)}, Z_t)| \cdot \mathbf{1}\{\Omega_{t-1}^{-1} W_{1:t}^{(J)} \cdot |f| \geq \Omega_{t-1}^{-1} \max_{1\leq I\leq N} W_{1:t-1}^{(I)} \cdot K\} \\
&\qquad\qquad \times \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r}) d\widetilde{\pi}_{t|1:t-1}(Z_t) \\
&= \Omega_{t-1}^{-1} \sum_{J=1}^{N} W_{1:t-1}^{(J)} \int W_t^{(J)} \cdot |f(\mathbf{Z}_{1:t-1}^{(J)}, Z_t)| \cdot \mathbf{1}\left\{ \frac{W_{1:t}^{(J)}}{\max_{1\leq I\leq N} W_{1:t-1}^{(I)}} \cdot |f| \geq K \right\} \\
&\qquad\qquad \times \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r}) d\widetilde{\pi}_{t|1:t-1}(Z_t) \\
&\leq \Omega_{t-1}^{-1} \sum_{J=1}^{N} W_{1:t-1}^{(J)} \int W_t^{(J)} \cdot |f(\mathbf{Z}_{1:t-1}^{(J)}, Z_t)| \cdot \mathbf{1}\{W_t^{(J)} \cdot |f| \geq K\} \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \\
&\qquad\qquad \times \mathbf{I}_{(m_t,M_t)}(z_{t,r}) d\widetilde{\pi}_{t|1:t-1}(Z_t) \\
&= \Omega_{t-1}^{-1} \sum_{J=1}^{N} W_{1:t-1}^{(J)} \cdot R(\mathbf{Z}_{1:t-1}^{(J)}, W_t^{(J)} \cdot |f| \cdot \mathbf{1}\{W_t^{(J)} \cdot |f| \geq K\}) \\
&\xrightarrow{P} \int R(\mathbf{Z}_{1:t-1}, W_t^{(J)} \cdot |f| \cdot \mathbf{1}\{W_t \cdot |f| \geq K\}) d\pi_{1:t-1}.
\end{aligned}
$$

Now we have

$$
\begin{aligned}
&\sum_{J=1}^{N} E[|\Omega_{t-1}^{-1} W_{1:t}^{(J)} \cdot f(\mathbf{Z}_{1:t}^{(J)})| \cdot \mathbf{1}\{\Omega_{t-1}^{-1} W_{1:t}^{(J)} \cdot |f(\mathbf{Z}_{1:t}^{(J)})| \geq \epsilon\}|\mathcal{F}_{J-1}] \cdot \mathbf{1}\{\Omega_{t-1}^{-1} \max_{1\leq I\leq N} W_{1:t-1}^{(I)} \leq \epsilon/K\} \\
&\quad \xrightarrow{P} \int R(\mathbf{Z}_{1:t-1}, W_t \cdot |f| \cdot \mathbf{1}\{W_t \cdot |f| \geq K\}) d\pi_{1:t-1}.
\end{aligned}
$$

In order for Condition (2.25) to be satisfied, (i) $\sum_{J=1}^{N} E[|\Omega_{t-1}^{-1} W_{1:t}^{(J)} \cdot f(\mathbf{Z}_{1:t}^{(J)})| \cdot \mathbf{1}\{\Omega_{t-1}^{-1} W_{1:t}^{(J)} \cdot |f(\mathbf{Z}_{1:t}^{(J)})| \geq \epsilon\} | \mathcal{F}_{J-1}] \cdot \mathbf{1}\{\Omega_{t-1}^{-1} \max_{1 \leq I \leq N} W_{1:t-1}^{(I)} \leq \epsilon / K\} \xrightarrow{P} \sum_{J=1}^{N} E[|\Omega_{t-1}^{-1} W_{1:t}^{(J)} \cdot f(\mathbf{Z}_{1:t}^{(J)})| \cdot \mathbf{1}\{\Omega_{t-1}^{-1} W_{1:t}^{(J)} \cdot |f(\mathbf{Z}_{1:t}^{(J)})| \geq \epsilon\} | \mathcal{F}_{J-1}]$ and (ii) $\int R(\mathbf{Z}_{1:t-1}, W_t \cdot |f| \cdot \mathbf{1}\{W_t \cdot |f| \geq K\}) d\pi_{1:t-1} \xrightarrow{P} 0$. Point (i) holds by noting that since $\Omega_{t-1}^{-1} \max_{1 \leq J \leq N} W_{t-1}^{(J)} \xrightarrow{P} 0$ by assumption, so it follows that $\mathbf{1}\{\Omega_{t-1}^{-1} \max_{1 \leq J \leq N} W_{1:t-1}^{(J)} \leq \epsilon / K\} \xrightarrow{P} 1$. Point (ii) holds by the Dominated Convergence Theorem as $K \longrightarrow \infty$ since the integral is bounded by an L_1 function.

Since both Conditions (2.24) and (2.25) of Theorem 2.3.2 hold, so Equation (2.28) follows. Also recalling Equation (2.27), we conclude that

$$\Omega_{t-1}^{-1} \sum_{J=1}^{N} f(\mathbf{Z}_{1:t}^{(J)}) W_{1:t}^{(J)} \xrightarrow{P} \pi_{1:t-1} F(\cdot, f). \tag{2.30}$$

Next note, by assumption, $F \in B_{t-1}$, and as a proper set, B_{t-1} includes the constant function $f \equiv 1$, which implies

$$\begin{aligned}
\Omega_{t-1}^{-1} \Omega_t &= \Omega_{t-1}^{-1} \sum_{J=1}^{N} W_{1:t}^{(J)} \\
&\xrightarrow{P} \int \cdots \int_{\Theta_t} \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t, M_t)}(z_{t,r}) (\Phi(M_t) - \Phi(m_t)) \pi_{t|1:t-1}(dZ_t) \pi_{1:t-1}(d\mathbf{Z}_{1:t-1}) \\
&= \pi_{1:t-1} F(\cdot, 1).
\end{aligned} \tag{2.31}$$

Combining Equations (2.30) and (2.31) and noting that $\pi_{1:t} = (\Phi(M_t) - \Phi(m_t)) \cdot \pi_{t|1:t-1} \cdot \pi_{1:t-1}$, we get our desired result

$$\begin{aligned}
&\Omega_t^{-1} \sum_{J=1}^{N} f(\mathbf{Z}_{1:t}^{(J)}) W_{1:t}^{(J)} = \frac{\Omega_{t-1}^{-1} \sum_{J=1}^{N} f(\mathbf{Z}_{1:t}^{(J)}) W_{1:t}^{(J)}}{\Omega_t \Omega_{t-1}^{-1}} \xrightarrow{P} \frac{\pi_{1:t-1} F(\cdot, f)}{\pi_{1:t-1} F(\cdot, 1)} \\
&= \frac{\int \cdots \int f(\mathbf{Z}_{1:t-1}, Z_t) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t, M_t)}(z_{t,r}) (\Phi(M_t) - \Phi(m_t)) \pi_{t|1:t-1}(dZ_t) \pi_{1:t-1}(d\mathbf{Z}_{1:t-1})}{\int \cdots \int \cdot \int_{\Theta_t} \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t, M_t)}(z_{t,r}) (\Phi(M_t) - \Phi(m_t)) \pi_{t|1:t-1}(dZ_t) \pi_{1:t-1}(d\mathbf{Z}_{1:t-1})} \\
&= \frac{\int \cdots \int f(\mathbf{Z}_{1:t-1}, Z_t) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t, M_t)}(z_{t,r}) \pi_{1:t}(d\mathbf{Z}_{1:t})}{\int \cdots \int \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t, M_t)}(z_{t,r}) \pi_{1:t}(d\mathbf{Z}_{1:t})} \\
&= \pi_{1:t} f.
\end{aligned}$$

That is,

$$\Omega_t^{-1} \sum_{J=1}^{N} f(\mathbf{Z}_{1:t}^{(J)}) W_{1:t}^{(J)} \xrightarrow{P} \pi_{1:t} f.$$

The final step of the proof for this lemma – showing that $\Omega^{-1} \max_{J=1}^{N} W_{1:t}^{(J)} \xrightarrow{P} 0$ – follows directly from the result in Douc and Moulines (2008). □

Resampling The purpose of this section is to prove that the particle system is still consistent after the resampling step of the algorithm. After resampling, the unnormalized importance weights are set to one.

Lemma 2.3.2. *(Resampling) Assume that the weighted sample* $\{(\mathbf{Z}_{1:t}^{(J)}, W_{1:t}^{(J)})\}_{J=1}^{N}$ *is consistent for* $(\pi_{1:t}, B_t)$*. Then, the uniformly weighted sample* $\{(\widetilde{\mathbf{Z}}_{1:t}^{(J)}, 1)\}_{J=1}^{N}$ *obtained using multinomial or deterministic-plus-residual multinomial resample is consistent for* $(\pi_{1:t}, B_t)$*.*

Proof. This holds directly from (Douc and Moulines, 2008, Theorem 3). □

Alteration The final step of the proof of Theorem 2.3.1 is to prove the consistency of the particle system after the alteration step. The lemma is stated below followed by the proof.

Lemma 2.3.3. *(Alteration) Assuming the uniformly weighted sample* $\{(\mathbf{Z}_{1:t}^{(J)}, 1)\}_{J=1}^{N}$ *is consistent for* $(\pi_{1:t}, B_t)$*, then the altered uniformly weighted sample* $\{(\widetilde{\mathbf{Z}}_{1:t}^{(J)}, 1)\}_{J=1}^{N}$ *is consistent for* $(\pi_{1:t}, \widetilde{B}_t)$ *(*$\widetilde{B}_t$ *defined below).*

Proof. Note that the $\{(\widetilde{\mathbf{Z}}_{1:t}^{(J)}, 1)\}_{J=1}^{N}$ is the *altered* particle system, while $\{(\mathbf{Z}_{1:t}^{(J)}, 1)\}_{J=1}^{N}$ are the resampled particles from Lemma 2.3.2. Letting $f \in B_t$ with $\widetilde{C}$, $\widetilde{D}$, and projection matrix Π as defined below Equation (2.19), we define the following conditional expectation

$$\begin{aligned} E[f(\widetilde{\mathbf{Z}}_{1:t}^{(J)}) | \widetilde{\mathcal{F}}_{J-1}] &= \int f\Big(\eta_2 \cdot \widetilde{C} + \widetilde{D} \frac{(\mathbf{Z}_{1:t}^{(J)} - \Pi \mathbf{Z}_{1:t}^{(J)})}{||\mathbf{Z}_{1:t}^{(J)} - \Pi \mathbf{Z}_{1:t}^{(J)}||}\Big) d\pi_{\widetilde{C},\widetilde{D}} \\ &= \int f(\eta_2 \cdot \widetilde{C} + \widetilde{D} \tau_{1:t}^{(J)}) d\pi_{\widetilde{C},\widetilde{D}} \triangleq h_f(\mathbf{Z}_{1:t}^{(J)}), \end{aligned}$$

where $\widetilde{\mathcal{F}}_J \triangleq \widetilde{\mathcal{F}}_0 \vee \sigma(\{\widetilde{\mathbf{Z}}_{1:t}^{(K)}\}_{1 \le K \le J}, (a,b]_{1:t})$, and $\widetilde{\mathcal{F}}_0 \triangleq \sigma(\{\mathbf{Z}_{1:t}^{(J)}\}_{J=1}^{N}, (a,b]_{1:t})$ for $J = 1, \ldots, N$. That is, $h_f(\widetilde{\mathbf{Z}}_{1:t}^{(J)})$ is the conditional expectation of the altered particles with respect to the

joint distribution of $\widetilde{C}$ and $\widetilde{D}$ denoted $\pi_{\widetilde{C},\widetilde{D}}$. We note that h_f is a function of $\mathbf{Z}_{1:t}$ because $\tau_{1:t}$ is a function of $\mathbf{Z}_{1:t}$, and, at times, it will be necessary to write $\tau_{1:t}^{(J)} = \tau_{1:t}(\mathbf{Z}_{1:t}^{(J)})$. Recall that C and D are defined by a decomposition of the original particle, call it L, selected to be resampled, $\mathbf{Z}_{1:t}^{(L)}$, and are independent by design. The $\widetilde{C}$ and $\widetilde{D}$ are the random variables to be resampled according to the target distributions of C and D with the $\tau_{1:t}^{(J)}$ considered fixed so that $\widetilde{\mathbf{Z}}_{1:t} = \eta_2 \cdot \widetilde{C} + \widetilde{D}\tau_{1:t}$.

The lemma will follow once we show

$$N^{-1}\sum_{J=1}^{N} E[f(\widetilde{\mathbf{Z}}_{1:t}^{(J)})|\widetilde{\mathcal{F}}_{J-1}] = N^{-1}\sum_{J=1}^{N} h_f(\mathbf{Z}_{1:t}^{(J)}) \longrightarrow \int h_f(\mathbf{Z}_{1:t})d\pi_{1:t}(\mathbf{Z}_{1:t}) = \int f(\mathbf{Z}_{1:t})d\pi_{1:t}(\mathbf{Z}_{1:t}).$$

This is because trivially $E[f(\widetilde{\mathbf{Z}}_{1:t}^{(J)})|\widetilde{\mathcal{F}}_{J-1}] = f(\widetilde{\mathbf{Z}}_{1:t}^{(J)})$, so all that is needed is for (i) $N^{-1}\sum_{J=1}^{N} h_f(\mathbf{Z}_{1:t}^{(J)}) \longrightarrow \int h_f(\mathbf{Z}_{1:t})d\pi_{1:t}(\mathbf{Z}_{1:t})$ and (ii) $\int h_f(\mathbf{Z}_{1:t})d\pi_{1:t}(\mathbf{Z}_{1:t}) = \int f(\mathbf{Z}_{1:t})d\pi_{1:t}(\mathbf{Z}_{1:t})$. The proof of Lemma 2.3.1 required more steps because the importance weights were not all N^{-1} as is the case here. The goal is to show that after the alteration to the particle system as described in Section 2.3.4, the altered particle system still targets the correct distribution $\pi_{1:t}$.

First, let us consider point (i). For h_f to be in B_t, f must be selected so that the following two conditions hold for any direction $\mathbf{Z}_{1:t}$:

$$\int |h_f(\widetilde{Z}_{1:t})|d\pi_{1:t} = \int \Big| \int f(\eta_2 \cdot \widetilde{C} + \widetilde{D}\tau_{1:t})d\pi_{\widetilde{C},\widetilde{D}}\Big| d\pi_{1:t} < \infty, \text{ and} \tag{2.32}$$

$$\begin{aligned} F(\mathbf{Z}_{1:t-1}, h_f) &= \\ &= \int_{\Theta_t} h_f(\mathbf{Z}_{1:t-1}, Z_t) \cdot \mathbf{I}_\star(Z_{t,1:r-1}) \cdot \mathbf{I}_{(m_t,M_t)}(z_{t,r})(\Phi(M_t) - \Phi(m_t))d\pi_{t|1:t-1}(Z_t) \\ &= \int_{\Theta_t} \Big(\int f\big(\eta_2 \cdot \widetilde{C} + \widetilde{D}\tau_{1:t}(\mathbf{Z}_{1:t-1}, Z_t)\big)d\pi_{\widetilde{C},\widetilde{D}}\Big) \cdot \mathbf{I}_\star(Z_{t,1:r-1})\times \\ &\qquad \mathbf{I}_{(m_t,M_t)}(z_{t,r})(\Phi(M_t) - \Phi(m_t))d\pi_{t|1:t-1}(Z_t) < \infty. \end{aligned} \tag{2.33}$$

These conditions are requiring that f is picked so that $h_f \in B_t$; let $\widetilde{B}_t$ be the set of $f \in B_t$ such that (2.32) and (2.33) hold. Then, $\widetilde{B}_t \subset B_t$, and since, for example, all bounded functions

would satisfy (2.32) and (2.33), $\widetilde{B}_t$ is non-empty. An example of an *unbounded* function f that satisfies the conditions, since t is finite, is the identity function $f(x) = x$. Since h_f would then be taking an expected value of two variables with finite expected values – $\widetilde{C}$, which follows a normal distribution centered at zero, and $\widetilde{D}$, which follows a chi-squared distribution with degrees of freedom, and hence a mean, connected to the number of levels of the random effect being resampled – it would exist on B_t. This will be finite as long as $t < \infty$. (More precisely, as long as the number of levels within the random effects is finite.)

Now we only need to show that point (ii) that $\int h_f(\mathbf{Z}_{1:t})d\pi_{1:t}(\mathbf{Z}_{1:t}) = \int f(\mathbf{Z}_{1:t})d\pi_{1:t}(\mathbf{Z}_{1:t})$. This follows because

$$\begin{aligned}
\int h_f(\mathbf{Z}_{1:t})d\pi_{1:t} &= \int h_f^*(\tau_{1:t})d\pi_\tau = \int \Big(\int f(\eta_2 \cdot \widetilde{C} + \widetilde{D}\tau_{1:t})d\pi_{\widetilde{C},\widetilde{D}} \Big) d\pi_\tau \\
&= \int \Big(\int f(\eta_2 \cdot C + D\tau_{1:t})d\pi_{C,D} \Big) d\pi_\tau = \int f(\eta_2 \cdot C + D\tau_{1:t})d\pi_{C,D} \times d\pi_\tau \\
&= \int f(\mathbf{Z}_{1:t})d\pi_{1:t} = E_{\pi_{1:t}}[f(\mathbf{Z}_{1:t})].
\end{aligned}$$

where $h_f^*(\tau_{1:t}) = h_f(\mathbf{Z}_{1:t})$, the equality from line one to line two follows because $\tau(\mathbf{Z}_{1:t}) = \tau(\widetilde{\mathbf{Z}}_{1:t})$, and the equality in the second line follows by Fubini's Theorem because $\widetilde{C}$ and $\widetilde{D}$ are independent of τ. □

2.4 Simulation study and applications

This simulation study has two parts. In the first part, we consider balanced and unbalanced designs for the two-fold nested model with model designs selected from Hernandez et al. (1992). In the second part, we use balanced and unbalanced two-factor crossed designs with an interaction term with model designs selected from Hernandez and Burdick (1993); both sets of designs include varying levels of imbalance. In addition to the classical, ANOVA-based methods proposed in Hernandez et al. (1992) and Hernandez and Burdick (1993), we compare the performance of our method to a Bayesian method proposed in Gelman (2006). The purpose of this study is to compare the performance of the proposed method with current methods using models with varying levels of imbalance. The proposed, classical and

Bayesian methods were compared using frequentist repeated sampling properties. Specifically, performance was compared based on empirical coverage of confidence intervals and average confidence interval length of the variance components. We focus on the confidence intervals of the variance components because that is where the classical methods have some issues when the designs become imbalanced, which are discussed below.

It is understood that the selection of a prior distribution influences the behavior of the posterior; the priors were selected based on recommendations in the literature for normal mixed linear models (Gelman, 2006). While Bayesian methods do not necessarily maintain frequentist properties, many practitioners interpret results from Bayesian analyses as approximately frequentist (i.e. they expect repeated-sampling properties to approximately hold) due to Bernstein-von Mises Theorem (Le Cam, 1986; van der Vaart, 2007), and so performing well in a frequentist sense has appeal. There are a number of examples investigating frequentist performance of Bayesian methodology such as Diaconis and Freedman (1986b), Diaconis and Freedman (1986a), Ghosal et al. (2000), and Mossman and Berger (2001).

It is important to note that the proposed method and the Bayesian methods are not restricted to the model designs selected for this study, and can be applied to any normal linear mixed model that satisfies the assumptions from previous sections, while the included ANOVA methods were developed specifically for the model designs used in this study. A more efficient algorithm than the proposed method may be possible for *specific* model designs, but one of our goals was to present a mode of inference for a general normal linear mixed model design. This is particularly helpful for a non-statistician practitioner who may not be familiar with the best way of selecting an appropriate approximate ANOVA method for an unbalanced design, or how to select a prior and implement a computational algorithm to approximate the posterior distribution.

As presented below, the proposed method tends to be conservative with comparable or shorter intervals than the classical and Bayesian methods used in the study.

2.4.1 Two-fold nested model

For the first part of the simulation study, we consider the unbalanced two-fold nested linear model

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk} \tag{2.34}$$

for $i = 1, \ldots, I$, $j = 1, \ldots, J_i$, and $k = 1, \ldots, K_{ij}$, where μ is an unknown constant and $\alpha_i \sim N(0, \sigma^2_\alpha)$, $\beta_{ij} \sim N(0, \sigma^2_\beta)$, and $\epsilon_{ijk} \sim N(0, \sigma^2_\epsilon)$.

Table 2.3 displays the model designs used in this part of the simulation study. Five model designs of Hernandez et al. (1992) were selected to cover different levels of imbalance both in the number of nested groups (J_i) and the number of observations within each group (K_{ij}). The parameters ϕ_1 and ϕ_2 reflect the degree of imbalance due to J_i and K_{ij}, respectively. The measures of imbalance listed is based on methods presented in Khuri (1987) where values range from 0 to 1, and smaller values suggest a greater degree of imbalance. The parameters values used in this part of the study are displayed in Table 2.4 with $\mu = 0$.

Design	ϕ_1	ϕ_2	ϕ	I	J_i	K_{ij}	n
MI - 1	0.9000	0.8889	0.8090	5	2,1,1,1,1	4,4,2,2,2,2	16
MI - 2	0.7778	0.7337	0.6076	3	4,2,1	1,5,5,5,1,5,1	23
MI - 3	1.0000	1.0000	1.0000	3	3,3,3	2,2,2,2,2,2,2,2,2	18
MI - 4	0.4444	1.0000	0.4444	6	1,1,1,1,1,7	2,2,2,2,2,2	24
						2,2,2,2,2,2	
MI - 5	1.0000	0.4444	0.4444	3	2,2,2	1,1,1,1,1,7	12

Table 2.3: Model designs used in the two-fold nested model of (2.34).

NOTE: ϕ_1 and ϕ_2 reflect the degree of imbalance due to J_i and K_{ij}, respectively, and $\phi = \phi_1 \cdot \phi_2$ is an overall measure of imbalance. See (2.34) for definitions of I, J_i, and K_{ij}; note sample size $(n) = \sum_i \sum_j K_{ij}$.

Design	**Parameter values** $(\sigma^2_\alpha, \sigma^2_\beta, \sigma^2_\epsilon)$
PI -1	$(0.2, 0.1, 0.7)$
PI -2	$(0.4, 0.3, 0.3)$
PI -3	$(0.2, 0.7, 0.1)$
PI -4	$(25, 4.0, 16)$
PI -5	$(1.0, 1.0, 1.0)$

Table 2.4: Model designs used in the two-fold nested model of (2.34).

For each model and parameter design combination, 2000 independent data sets were generated, and 5000 particles were simulated for the proposed method.

ANOVA-based methods

Hernandez et al. (1992) present two methods for determining confidence intervals for σ_α^2 and σ_β^2. In fact, they propose a third method for σ_β^2, but we do not include it because there is not an analogous method for σ_α^2. The first method proposed by Hernandez et al. (1992) uses, what they refer to as, the unweighted sum of squares. This is denoted as **USS**. This decomposition for the unbalanced two-fold nested model has been employed in other model settings noted in their manuscript. The other method proposed by Hernandez et al. (1992) relies on a decomposition that they developed from Ting et al. (1990)'s exact method for balanced designs. This is denoted by **TYPEI**. In general for unbalanced designs, these decompositions do not result in independent nor chi-squared random variables; however, for certain model designs, **USS** and **TYPEI** do have these desired properties. They also note that the performance of these methods will depend, for better or for worse, on the ratios $\sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\epsilon^2)$ or $\sigma_\alpha^2/\sigma_\epsilon^2$ for the intervals on σ_α^2, and $\sigma_\beta^2/\sigma_\epsilon^2$ for the intervals on σ_β^2.

Bayesian method

A Bayesian method is also considered for comparison. Bayesian hierarchical models provide a means of constructing confidence intervals for random effects models. Part of the art of the Bayesian methodology is in selecting appropriate prior distributions for the unknown parameters. For inference on the unknown variance component parameters when there is not prior information available (i.e. when seeking a noninformative prior), Gelman (2006) recommends employing a uniform prior distribution on the standard deviation parameters when there are a sufficient number of groups; otherwise, a half-t distribution is suggested. Gelman (2006) notes that often an inverse-gamma prior ($IG(\epsilon, \epsilon)$ as $\epsilon \longrightarrow 0$) is considered, but that limiting posterior distribution is not, in fact, proper. Gelman shows that if the variance component is close to 0, the resulting posterior becomes dependent on this prior. Gelman also points out that the uniform prior on the standard deviation parameter, as long

as there are more than two levels, results in a proper limiting posterior distribution; however, he notes that if the scale parameter a is truly finite, considering $a \longrightarrow \infty$ results in a positive miscalibration of the posterior mean (i.e. a positive bias). Gelman attributes this to the fact that the variance components must be positive, which he says is similar to the issue the classical methods face with the possibility of negative estimators of the variance components.

Per the recommendation of Gelman (2006), both uniform and half-t priors are considered (denoted $\mathbf{BAY1}_{1.5}$ and $\mathbf{BAY1}_3$, and $\mathbf{BAY2}_{1.5}$ and $\mathbf{BAY2}_3$, respectively, where the subscripts 1.5 and 3 specify the prior scale variable). The R package 'rjags' was used to implement the Bayesian methods used in the simulation study and applications. Gelman (2006) suggests using a uniform prior (i.e. $U(0, a)$) on the standard deviation parameters when there are at least 5 groups and explains that fewer than 3 groups results in an improper posterior distribution. Calibration is necessary in selecting the parameter a in the prior distribution; we use 1.5 and 3 times the range of the data (per the recommendation in Gelman (2006, p. 528) to use a value that is "high but not off the scale"), which appears reasonable when reviewing the resulting posterior distributions. For example, the hierarchical model for the two-fold nested model of (2.34) for $\mathbf{BAY1}_{1.5}$ and $\mathbf{BAY1}_3$ is

$$
\begin{aligned}
Y_{ijk} &\sim N(\mu + \alpha_i + \beta_{ij}, \sigma_\epsilon^2),\ i = 1, \ldots, I,\ j = 1, \ldots, J_i,\ k = 1, \ldots, K_{ij} \\
\alpha_i &\sim N(0, \sigma_\alpha),\ i = 1, \ldots, I,\ \beta_{ij} \sim N(0, \sigma_\beta),\ j = 1, \ldots, J_i, \\
\sigma_\alpha &\sim U(0, a),\ \sigma_\beta \sim U(0, a),\ \mu \sim N(0, 10^8)
\end{aligned}
$$

For the second Bayesian method, a similar hierarchical model is used. Instead of a uniform distribution on the non-error variance components, a half-Cauchy distribution with scale parameter a set as 1.5 or 3 times the range of the data is used. Six thousand particles were generated with the first 1000 discarded (i.e. a "burn-in" of 1000 particles).

Results

Performance is based on the empirical coverage of $(1-\alpha)100\%$ confidence intervals and average interval length for the parameters of interest θ. We define a lower-tailed $(1-\alpha)100\%$ confidence

interval on θ as the interval $(-\infty, U_\alpha]$ such that $P(-\infty < \theta \leq U_\alpha) = 1-\alpha$, an upper-tailed $(1-\alpha)100\%$ confidence interval on θ as the interval $[L_\alpha, \infty)$ such that $P(L_\alpha \leq \theta < \infty) = 1-\alpha$, and a two-sided equal-tailed confidence interval on θ as the interval $[L_{\alpha/2}, U_{\alpha/2}]$ such that $P(L_{\alpha/2} \leq \theta \leq U_{\alpha/2}) = 1-\alpha$. Based on the normal approximation to the binomial distribution, we will consider empirical coverage between 94% and 96% appropriate for 95% two-sided confidence intervals.

The simulation results for Model (2.34) are displayed in Figures 2.3 - 2.10 with the corresponding five-number summaries displayed in Tables A.1 - A.4 in the appendix. The results are also summarized in Appendix A.1 as box plots for 95% two-sided confidence intervals and average lengths combined for parameters σ^2_α and σ^2_β. Though there are some differences in performance for each model and particular variance component, the box plots in the Appendix provide a convenient summary for each model used in the study. One point in each box plot represents a model and parameter combination from Tables 2.4 and 2.3. We note that Hernandez et al. (1992) focuses on confidence intervals for the non-error variance components and do not derive confidence intervals for μ or σ^2_ϵ, so there are not results for **USS** or **TYPEI** for Figures 2.7, 2.8, 2.9, or 2.10.

The focus of the study is on the performance of the non-error variance components σ^2_α and σ^2_β. The results for σ^2_α are found in Figures 2.3 and 2.4. Overall, **USS** and **TYPEI** tend to maintain the stated coverage. $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$, and **FID** are conservative for the two-sided intervals, and very conservative for the one-sided lower confidence intervals. $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, and $\mathbf{BAY2}_3$ range from mildly liberal to very conservative for the upper confidence intervals with median coverages that are mildly conservative. **FID** is conservative to very conservative for the upper confidence intervals. Overall, **FID** tends to be conservative to very conservative. However, even given its conservative performance, the average interval lengths for **FID** are competitive with **USS**, **TYPEI**, and $\mathbf{BAY1}_{1.5}$, but beats $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, and $\mathbf{BAY2}_3$. The fact that the average interval lengths of σ^2_α for $\mathbf{BAY1}_{1.5}$ performs better than the other Bayesian intervals is somewhat surprising. A similar result appears for σ^2_α and σ^2_β, for the two-way crossed with interaction model discussed in Section 2.4.2, which are analogous parameters to σ^2_α of the two-fold nested

model of this section in that they are the main random effects (i.e. not nested). For model designs MI - 1 and 4, all the Bayesian interval lengths are comparable; however, it is with model designs MI - 2, 3, and 5 that $\mathbf{BAY1}_{1.5}$ has the shortest average interval lengths among the Bayesian intervals. The reason this is rather surprising is that designs MI - 2, 3, and 5 have only three levels for the main random effect ($I = 3$), and Gelman (2006) suggests the use of a uniform prior when the number of levels is at least 5.

The results for σ^2_β are found in Figures 2.5 and 2.6. **USS** and **TYPEI** tend to maintain the stated level of coverage for all the confidence intervals. $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$, and **FID** have similar conservative empirical coverages for all the confidence intervals with **FID** slightly more conservative for the upper confidence intervals. Several of the average interval lengths of **USS** and **TYPEI** for σ^2_β are much wider than $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$, and **FID**. These wide intervals are from MI - 1 (revealed in the middle plot of Figure 2.5). This appears to be due to the model design; the derivation of the confidence intervals results in the degrees of freedom of 1 for the nested factor. We note that the wide confidence intervals for σ^2_β also appear in the simulation results for model MI-1 in Hernandez et al. (1992). $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$, and **FID** do not appear to have this issue. Also, $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, and $\mathbf{BAY2}_3$ have very similar performances unlike with σ^2_α where $\mathbf{BAY1}_{1.5}$ had shorter intervals than the other Bayesian methods.

$\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$, and **FID** all tend to have conservative coverage for the two-sided confidence intervals for μ as displayed in Figures 2.7 and 2.8, but **FID** has the shortest average confidence interval lengths. **FID** tends to be within the stated coverage for σ^2_ϵ as displayed in Figures 2.9 and 2.10, with $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$ having correct coverage for the two-sided confidence intervals, liberal to conservative coverage for the upper confidence intervals, and slightly conservative coverage for the lower confidence intervals. Overall for σ^2_ϵ, **FID** has the shortest confidence intervals.

The proposed method, while maintaining conservative coverage, has average interval lengths that are competitive or better than the other methods used in this part of the study. The proposed method offers an easily generalizable framework and provides intervals for fixed effects and the error variance component, unlike the methods presented in Hernandez et al.

(1992). Furthermore, Hernandez et al. (1992) can have intervals that become excessively wide with certain model designs as seen in the middle plot of Figure 2.5. While the conservative coverage for the Bayesian methods can be deemed acceptable, their average interval lengths tend to be wider than the proposed method.

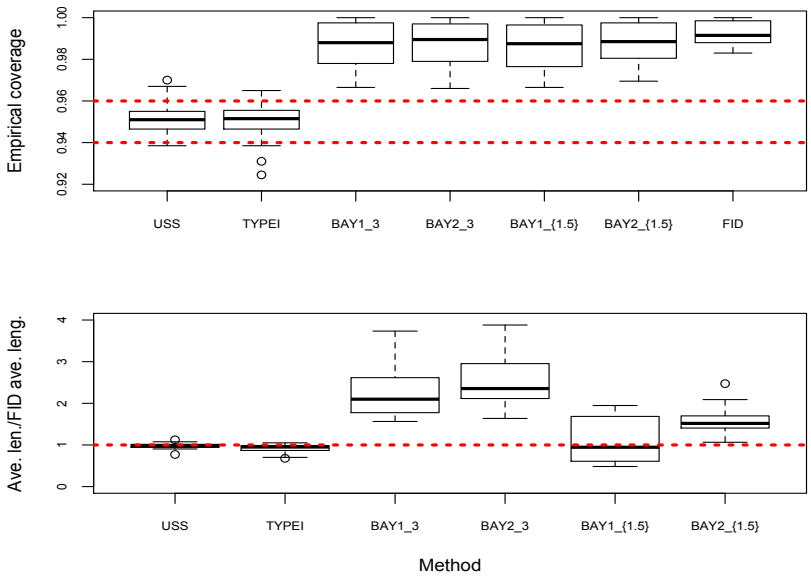


Figure 2.3: Simulation results for σ_α^2 from the two-fold nested model of (2.34). The top plot is the empirical coverage probabilities for 95% two-sided confidence intervals. The bottom plot is the average interval lengths divided by the average interval lengths of **FID**. Each value used in a box plot corresponds to a particular model design and parameter combination.

Application

In addition to the simulation study, we consider the application of Model (2.34) presented in Hernandez et al. (1992) concerning the blood pH of female mice offspring. Fifteen dams were mated with 2 or 3 sires where each sire was only used for one dam (i.e. 37 sires were used in the experiment), and the purpose of the study was to determine if the variability in the blood pH of the female offspring is in part due to the variability in the mother. There is imbalance in the data due to the number of sires mated with each dam (2 or 3); also note the natural imbalance in the data resulting from the number of female offspring.

The 95% confidence intervals based on the real data are presented in Table 2.5. An example of the generalized fiducial distribution for σ_α^2 is displayed in Figure 2.11. This

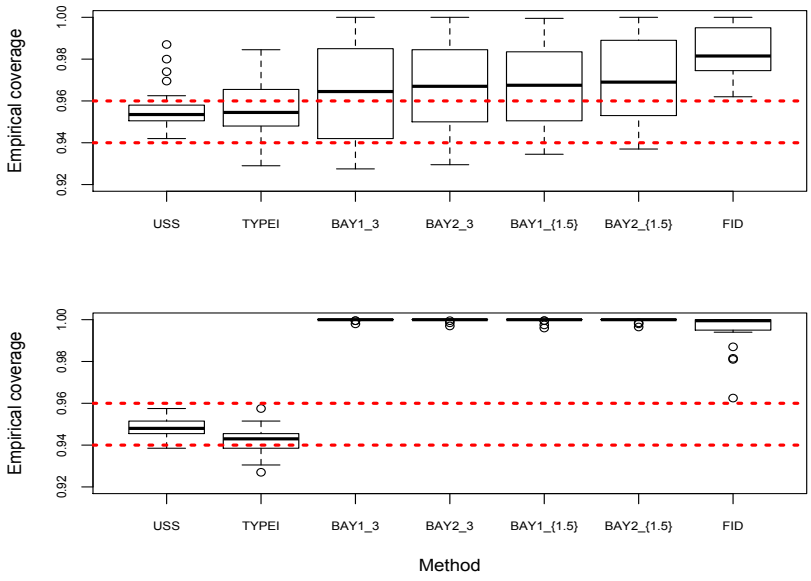


Figure 2.4: Simulation results for σ_α^2 from the two-fold nested model of (2.34). The top plot is the empirical coverage probabilities for 95% one-sided upper confidence intervals. The bottom plot is the empirical coverage probabilities for 95% one-sided lower confidence intervals. Each value used in a box plot corresponds to a particular model design and parameter combination.

highlights one of the advantages of the proposed method over classical methods (and shared with Bayesian methods), which is a distribution on the parameter space allowing for inferences similar to those made using Bayesian posterior distributions.

In order to evaluate the empirical coverage of the proposed method in this real-data example, we perform a simulation study using the REML estimates for all the parameters ($\mu = 44.92$, $\sigma_\alpha^2 = 8.90$, $\sigma_\beta^2 = 2.65$, and $\sigma_\epsilon^2 = 24.81$). Simulating 2000 independent data sets with the noted parameter values, we find the empirical coverage using **USS**, **TYPEI**, $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$, and **FID**, and the average lengths of the two-sided intervals. The results of the simulation study are also found in Table 2.5.

The third column of Table 2.5 displays the 95% two-sided confidence intervals based on the real data. The remaining columns display the results based on the 2000 independent data sets generated using the REML estimates. The empirical coverage of the intervals for σ_α^2 are within the stated coverage for all the methods, but we see that **FID** has the shortest interval lengths. The intervals for σ_β^2 are within the stated coverage to conservative, with $\mathbf{BAY1}_{1.5}$,

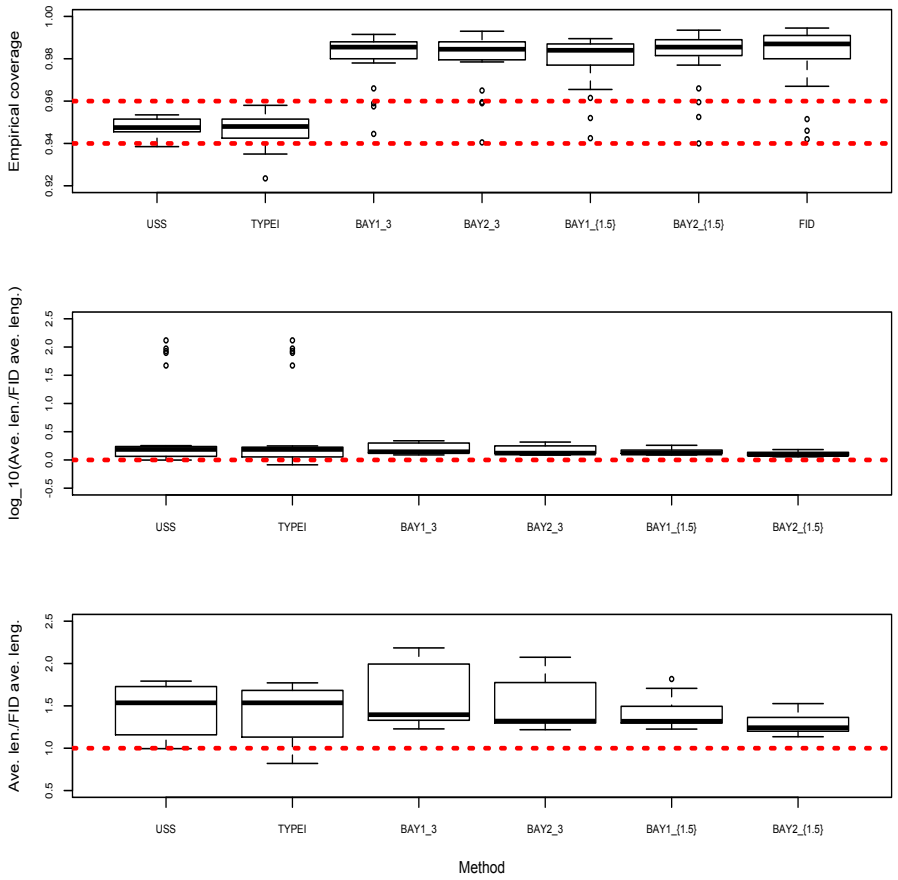


Figure 2.5: Simulation results for σ_β^2 from the two-fold nested model of (2.34). The top plot is the empirical coverage probabilities for 95% two-sided confidence intervals. The middle plot is the average interval lengths divided by the average interval lengths of **FID** on the $\log_{10}$ scale. The bottom plot is the average interval lengths divided by the average interval lengths of **FID** cutting out the outliers of **USS** and **TYPEI** displayed in the middle plot. Each value used in a box plot corresponds to a particular model design and parameter combination.

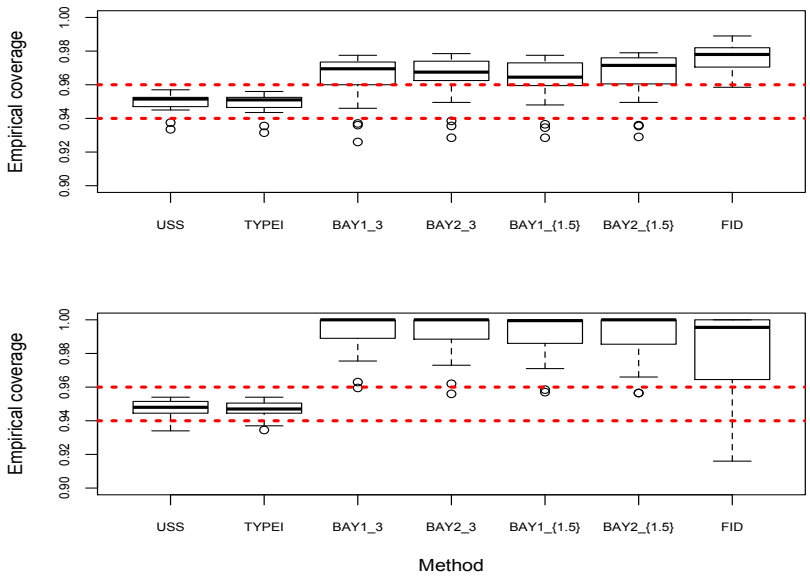


Figure 2.6: Simulation results for σ_β^2 from the two-fold nested model of (2.34). The top plot is the empirical coverage probabilities for 95% one-sided upper confidence intervals. The bottom plot is the empirical coverage probabilities for 95% one-sided lower confidence intervals. Each value used in a box plot corresponds to a particular model design and parameter combination.

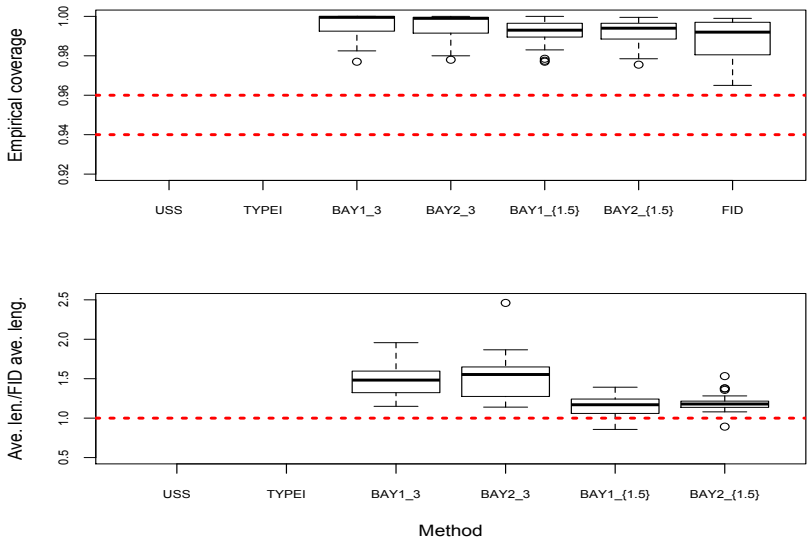


Figure 2.7: Simulation results for μ from the two-fold nested model of (2.34). The top plot is the empirical coverage probabilities for 95% two-sided confidence intervals. The bottom plot is the average interval lengths divided by the average interval lengths of **FID**. Each value used in a box plot corresponds to a particular model design and parameter combination. The paper proposing **USS** and **TYPEI** (Hernandez et al., 1992) focuses on confidence intervals for the non-error variance components and do not derive confidence intervals for μ.

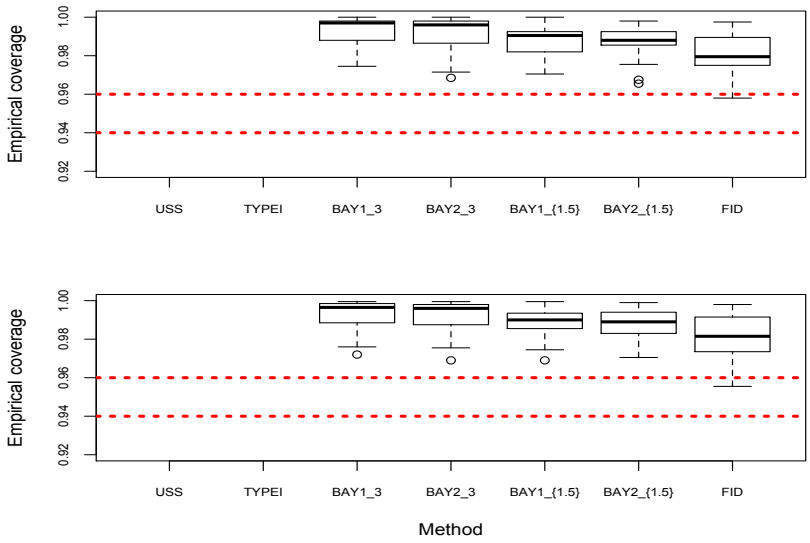


Figure 2.8: Simulation results for μ from the two-fold nested model of (2.34). The top plot is the empirical coverage probabilities for 95% one-sided upper confidence intervals. The bottom plot is the empirical coverage probabilities for 95% one-sided lower confidence intervals. Each value used in a box plot corresponds to a particular model design and parameter combination. The paper proposing **USS** and **TYPEI** (Hernandez et al., 1992) focuses on confidence intervals for the non-error variance components and do not derive confidence intervals for μ.

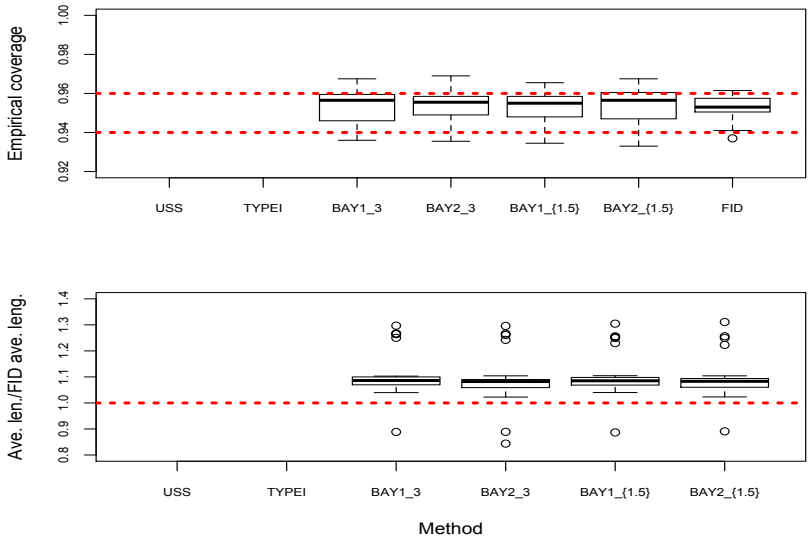


Figure 2.9: Simulation results for σ_ϵ^2 from the two-fold nested model of (2.34). The top plot is the empirical coverage probabilities for 95% two-sided confidence intervals. The bottom plot is the average interval lengths divided by the average interval lengths of **FID**. Each value used in a box plot corresponds to a particular model design and parameter combination. The paper proposing **USS** and **TYPEI** (Hernandez et al., 1992) focuses on confidence intervals for the non-error variance components and do not derive confidence intervals for σ_ϵ^2.

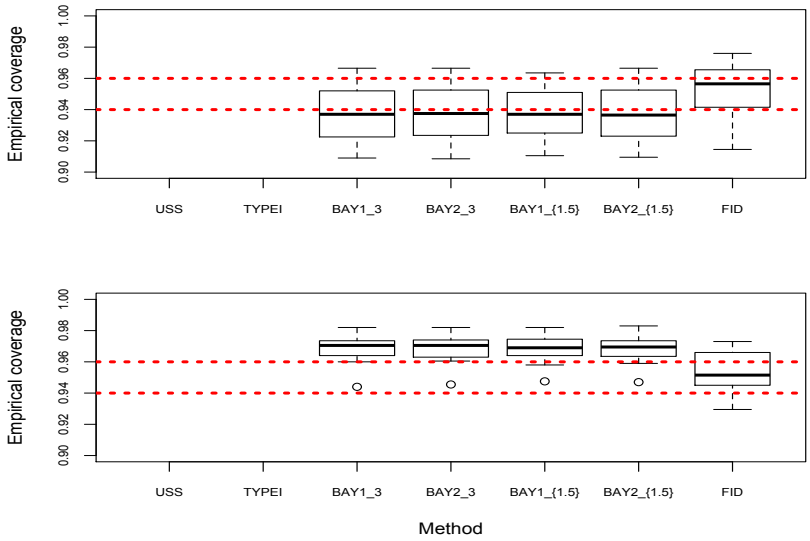


Figure 2.10: Simulation results for σ_ϵ^2 from the two-fold nested model of (2.34). The top plot is the empirical coverage probabilities for 95% one-sided upper confidence intervals. The bottom plot is the empirical coverage probabilities for 95% one-sided lower confidence intervals. Each value used in a box plot corresponds to a particular model design and parameter combination. The paper proposing **USS** and **TYPEI** (Hernandez et al., 1992) focuses on confidence intervals for the non-error variance components and do not derive confidence intervals for σ_ϵ^2.

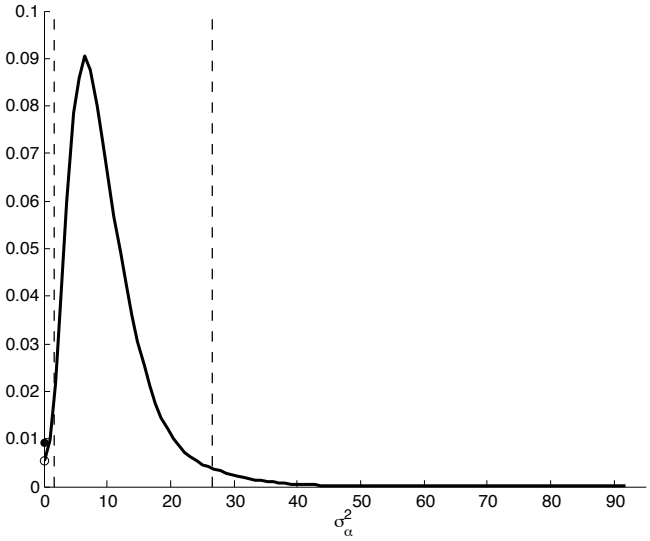


Figure 2.11: The generalized fiducial distribution of 5000 generated particles for σ_α^2 using the real data described in Section 2.4.1. The solid line is the normal kernel density estimate of the distribution with a point mass at zero. The two vertical dashed lines represent the lower and upper bounds for 95% confidence intervals based on the weights of the generated particles.

BAY1$_3$, and **BAY2**$_{1.5}$, **BAY2**$_3$ having the most conservative coverage. Again, we see that **FID** has the shortest overall intervals.

Var. comp.	Method	95% 2-sided C.I.	2-sided/ave.len.	Upper/lower
σ^2_α	**USS**	(2.30, 28.56)	0.953/25.1	0.949/0.958
	TYPEI	(1.94, 26.23)	0.950/25.2	0.949/0.957
	BAY1$_3$	(1.51, 30.21)	0.956/37.2	0.948/0.962
	BAY2$_3$	(1.76, 30.04)	0.955/27.5	0.950/0.961
	BAY1$_{1.5}$	(1.73, 30.72)	0.955/29.2	0.948/0.959
	BAY2$_{1.5}$	(1.56, 30.02)	0.955/27.3	0.948/0.960
	FID	(1.53, 26.67)	0.948/24.5	0.958/0.947
σ^2_β	**USS**	(0.00, 11.56)	0.961/10.9	0.952/0.952
	TYPEI	(0.00, 11.26)	0.964/12.4	0.953/0.953
	BAY1$_3$	(0.04, 12.55)	0.980/11.3	0.959/0.996
	BAY2$_3$	(0.01, 11.92)	0.983/11.3	0.958/0.995
	BAY1$_{1.5}$	(0.17, 11.81)	0.976/11.2	0.956/0.994
	BAY2$_{1.5}$	(0.01, 12.49)	0.982/11.3	0.958/0.994
	FID	(0.19, 10.54)	0.974/10.2	0.951/0.986

Table 2.5: Two-fold nested model: real data example

NOTE: The 95% intervals are based on the actual data while the remaining information are the empirical results from 2000 independently generated data set using the REML estimates for each parameter. The results are the empirical coverage and average interval lengths of 95% confidence intervals.

2.4.2 Two-factor crossed design with interaction

In this part of the simulation study, we consider the two-factor crossed designs with interaction written as

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \tag{2.35}$$

for $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, K_{ij}$, where μ is an unknown constant and $\alpha_i \sim N(0, \sigma^2_\alpha)$, $\beta_j \sim N(0, \sigma^2_\beta)$, $(\alpha\beta)_{ij} \sim N(0, \sigma^2_{\alpha\beta})$, and $\epsilon_{ijk} \sim N(0, \sigma^2_\epsilon)$, all independent of each other.

Table 2.6 displays the model designs used in this part of the study, and, again, the overall measure of imbalance (ϕ) proposed in Khuri (1987) is displayed for each design. The model designs are taken from Hernandez and Burdick (1993), where they propose an ANOVA-based approach for confidence intervals on the variance components. The parameters values used

in this part of the study are $\mu = 0$, and the following combinations displayed in Table 2.7.

Design	ϕ	I	J	K_{ij}	n
MII - 1	.8768	4	3	2, 1, 3/ 2, 1, 1/ 2, 2, 2/ 1, 2, 3	22
MII - 2	.6667	3	3	4, 1, 1/ 4, 1, 1/ 4, 1, 1	18
MII - 3	.6667	3	3	4, 4, 4/ 1, 1, 1/ 1, 1, 1	18
MII - 4	.4011	3	4	8, 1, 1/ 1, 1, 1, 1/ 1, 1, 1, 1	19
MII - 5	.7619	5	3	1,2,2/5,2,7/2,2,2/2,4,2/3,2,2	40
MII - 6	1.000	3	3	2,2,2/2,2,2/2,2,2	18

Table 2.6: Model designs used in the two-factor crossed design with interaction model of (2.35).

NOTE: The parameter ϕ is an overall measure of imbalance of the model. See (2.35) for definitions of I, J, and K_{ij}; note sample size $n = \sum_i \sum_j K_{ij}$.

Design	Parameter values $(\sigma^2_\alpha, \sigma^2_\beta, \sigma^2_{\alpha\beta}, \sigma^2_\epsilon)$
PII -1	$(0.1, 0.5, 0.1, 0.3)$
PII -2	$(0.1, 0.3, 0.1, 0.5)$
PII -3	$(0.1, 0.1, 0.3, 0.5)$
PII -4	$(0.1, 0.1, 0.5, 0.3)$
PII -5	$(1.0, 1.0, 1.0, 1.0)$

Table 2.7: Parameter values used in the two-factor crossed design with interaction model of (2.35).

For each design and set of parameter values, 2000 independent data sets were generated, and 5000 particles were simulated for the proposed method. As before, performance is based on the empirical coverage of $(1-\alpha)100\%$ confidence intervals and average interval length for the parameter of interest θ. Based on the normal approximation to the binomial distribution, we will consider empirical coverage between 94% and 96% appropriate for 95% two-sided confidence intervals.

ANOVA-based method

The ANOVA-based method of Hernandez and Burdick (1993) is based on what they refer to as the unweighted sum of squares in order to construct confidence intervals for σ^2_α, σ^2_β, and $\sigma^2_{\alpha\beta}$. (Note that this is the same unweighted sum-of-squares approach used in Hernandez et al. (1992) for the two-fold nested model.) The method they propose is based on intervals

for balanced designs presented by Ting et al. (1990). The usual issues with the ANOVA-based methods occur with this design when the data are unbalanced: non-unique partitioning of the total sum-of-squares, lack of independence, and lack of desired distributions (no chi-squared random variables). The claim they make is that the lack of independence has a "canceling effect" with the lack of chi-squaredness. In the simulation study, this method will be called **HB**.

Bayesian method

The same prior distributions were used for the two-factor crossed design with interaction as with the two-fold nested model. The details of the Bayesian method are found in Gelman (2006), and outlined above in Section 2.4.1. As with Model (2.34), $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_{3}$, $\mathbf{BAY2}_{1.5}$, and $\mathbf{BAY2}_{3}$ are used to denote the results using the Bayesian framework.

Results

The simulation results for Model (2.35) are displayed in Figures 2.12 - 2.21 with corresponding five-number summaries displayed in Tables A.5 - A.9 found in the appendix. As with Model (2.34), the results are further summarized in Appendix A.1 as box plots for 95% two-sided confidence intervals and average interval lengths combined for parameters σ^2_α, σ^2_β, and $\sigma^2_{\alpha\beta}$. Though there are some differences in performance for each model and particular variance component, the box plots in the Appendix provide a convenient summary for each model used in the study. One point in each box plot represents a model and parameter combination from Tables 2.7 and 2.6. We note that Hernandez and Burdick (1993) focuses on confidence intervals for the non-error variance components and do not derive confidence intervals for μ or σ^2_ϵ, so there are not results for **HB** for Figures 2.18 - 2.21.

HB maintains correct empirical coverage for the 95% confidence two-sided, upper and lower interval coverage on σ^2_α, σ^2_β, and $\sigma^2_{\alpha\beta}$ as displayed in Figures 2.12 - 2.17.

$\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_{3}$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_{3}$ and **FID** have comparable empirical coverages for the two-sided intervals on σ^2_α with **FID** having slightly more conservative intervals. For the upper confidence intervals, **FID** remains conservatives, but the $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_{3}$, $\mathbf{BAY2}_{1.5}$,

and $\mathbf{BAY2}_3$ are within the stated coverage to conservative. The lower confidence interval empirical coverages are very conservative for $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$ and **FID**.

The performance for $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$ and **FID** for σ^2_β is very similar to that of σ^2_α except the empirical coverage on the upper confidence intervals for $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, and $\mathbf{BAY2}_3$ range from slightly liberal to correct coverage, and **FID** is less conservative.

$\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$ and **FID** for $\sigma^2_{\alpha\beta}$ have conservative two-sided intervals, moderately conservative upper confidence intervals, and correct to very conservative coverage on the lower intervals.

The average interval lengths for **HB** are competitive with **FID** for σ^2_α and σ^2_β, but otherwise **FID** tends to have the shortest average confidence intervals for σ^2_α, σ^2_β, and $\sigma^2_{\alpha\beta}$.

The empirical coverage for all the intervals on μ for $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$, and **FID** tend to be conservative as displayed in Figures 2.18 and 2.19. However, the average interval lengths on μ for **FID** are shorter than the average interval lengths for $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, and $\mathbf{BAY2}_3$.

The empirical coverage of all intervals on σ^2_ϵ for **FID** tend to be within the stated coverage, while the intervals for $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, and $\mathbf{BAY2}_3$ range from within the stated coverage for the two-sided intervals to slightly liberal for the upper intervals and slightly conservative for the upper intervals. **FID** produces the shortest average confidence intervals for σ^2_ϵ.

As with the previous simulation study, **FID**, while maintaining conservative coverage, has average interval lengths that are generally competitive or better than the competing methods used in this study.

Application

An application of Model (2.35) is discussed in Khuri and Littell (1987) where they analyze the variation in fusiform rust due to the family (classified by the female parent) and test location in Southern pine tree plantations. The data are the proportion of trees from each family and test location that has fusiform rust, and the variance stabilizing transformation

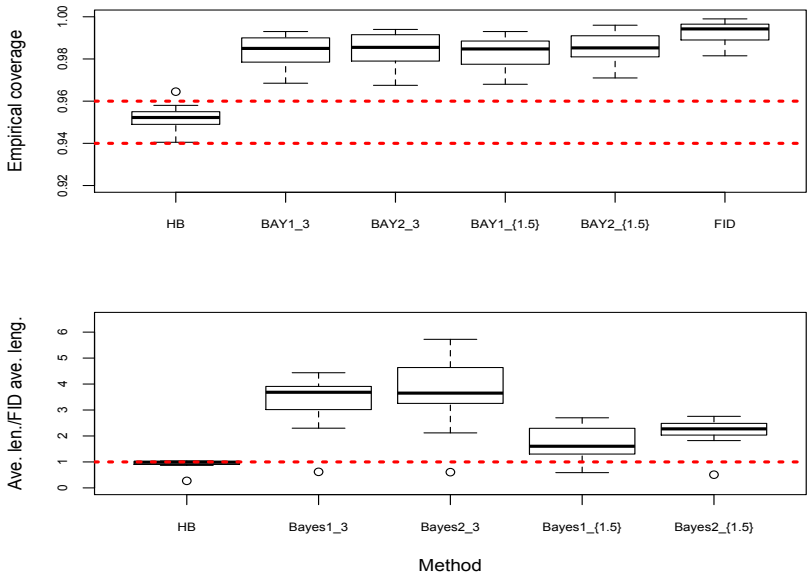


Figure 2.12: Simulation results for σ^2_α for the two-factor crossed design with interaction. The top plot is the empirical coverage probabilities for 95% two-sided confidence intervals. The bottom plot is the average interval lengths divided by the average interval lengths of **FID**. Each value used in a box plot corresponds to a particular model design and parameter combination.

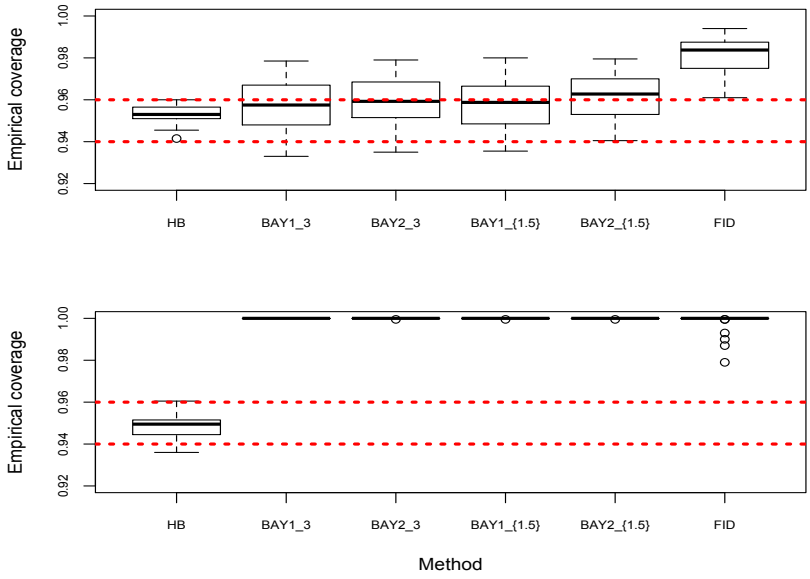


Figure 2.13: Simulation results for σ^2_α for the two-factor crossed design with interaction. The top plot is the empirical coverage probabilities for 95% one-sided upper confidence intervals. The bottom plot is the empirical coverage probabilities for 95% one-sided lower confidence intervals. Each value used in a box plot corresponds to a particular model design and parameter combination.

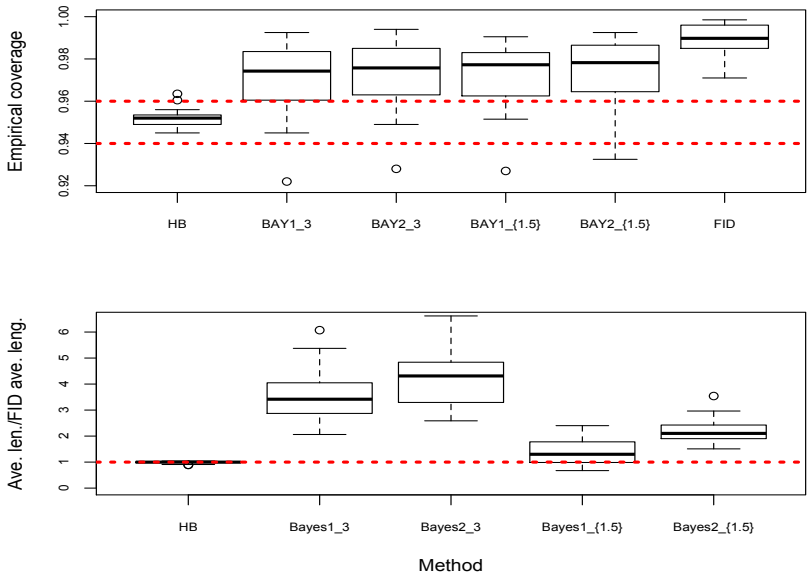


Figure 2.14: Simulation results for σ^2_β for the two-factor crossed design with interaction. The top plot is the empirical coverage probabilities for 95% two-sided confidence intervals. The bottom plot is the average interval lengths divided by the average interval lengths of **FID**. Each value used in a box plot corresponds to a particular model design and parameter combination.

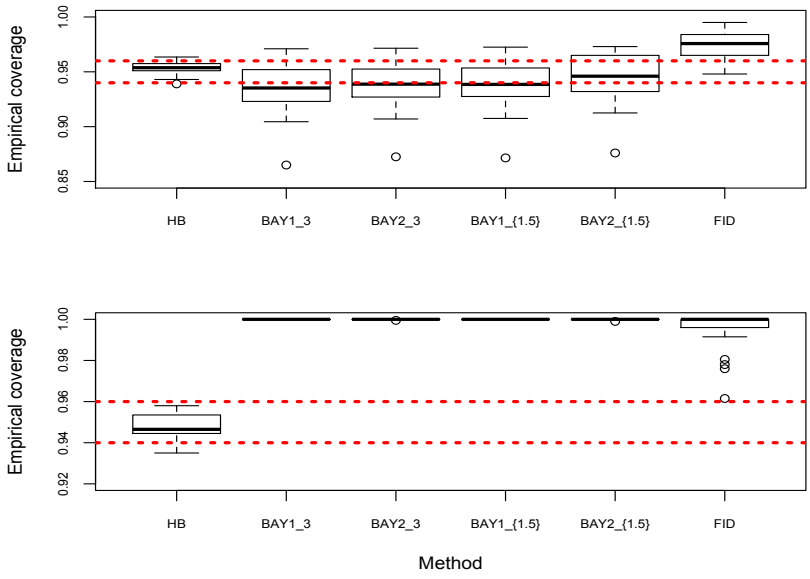


Figure 2.15: Simulation results for σ^2_β for the two-factor crossed design with interaction. The top plot is the empirical coverage probabilities for 95% one-sided upper confidence intervals. The bottom plot is the empirical coverage probabilities for 95% one-sided lower confidence intervals. Each value used in a box plot corresponds to a particular model design and parameter combination.

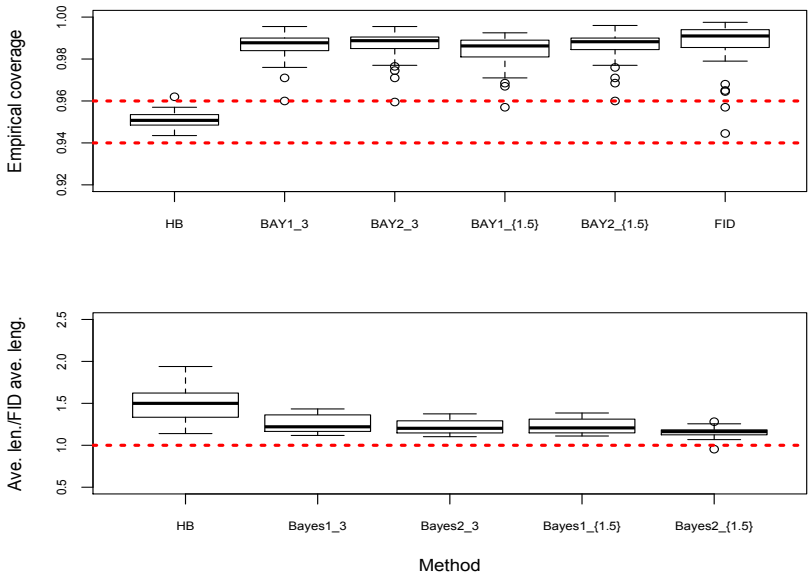


Figure 2.16: Simulation results for $\sigma^2_{\alpha\beta}$ for the two-factor crossed design with interaction. The top plot is the empirical coverage probabilities for 95% two-sided confidence intervals. The bottom plot is the average interval lengths divided by the average interval lengths of **FID**. Each value used in a box plot corresponds to a particular model design and parameter combination.

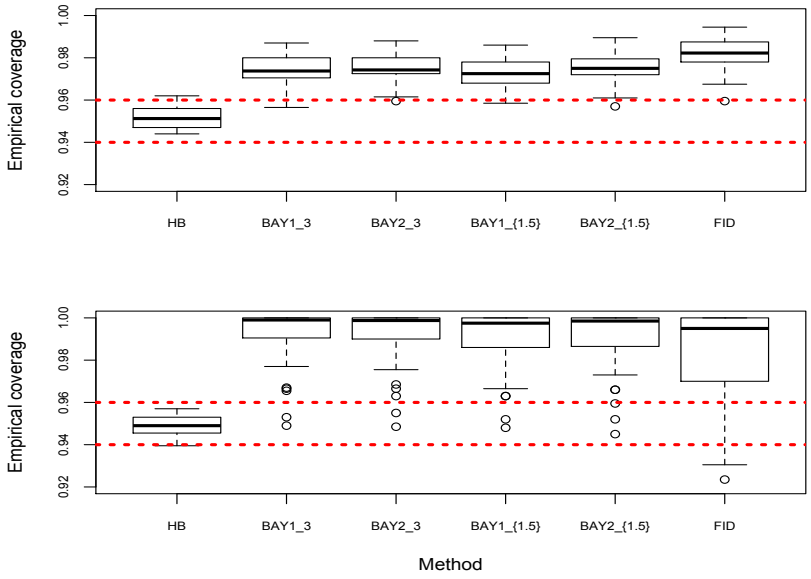


Figure 2.17: Simulation results for $\sigma^2_{\alpha\beta}$ for the two-factor crossed design with interaction. The top plot is the empirical coverage probabilities for 95% one-sided upper confidence intervals. The bottom plot is the empirical coverage probabilities for 95% one-sided lower confidence intervals. Each value used in a box plot corresponds to a particular model design and parameter combination.

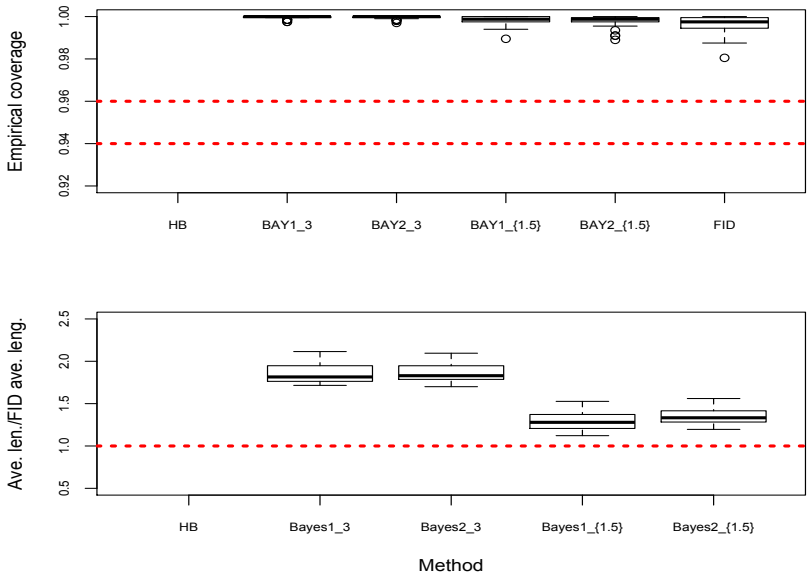


Figure 2.18: Simulation results for μ for the two-factor crossed design with interaction. The top plot is the empirical coverage probabilities for 95% two-sided confidence intervals. The bottom plot is the average interval lengths divided by the average interval lengths of **FID**. Each value used in a box plot corresponds to a particular model design and parameter combination.

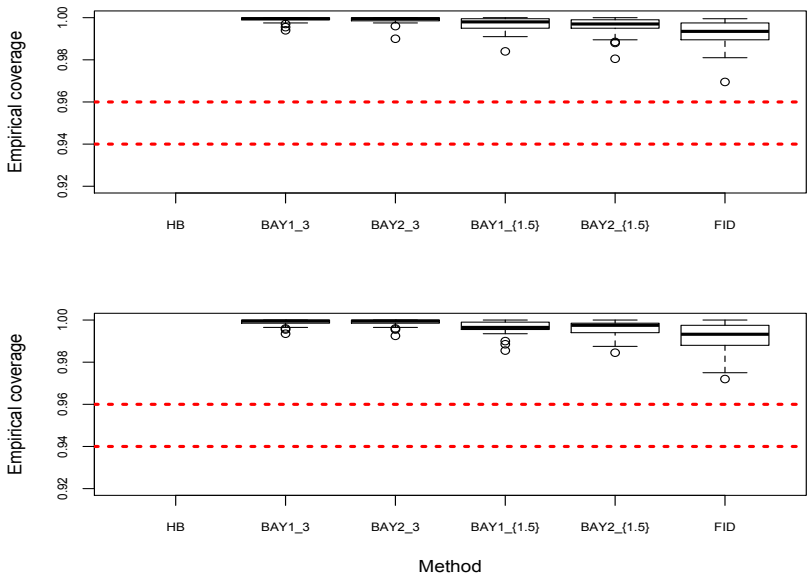


Figure 2.19: Simulation results for μ for the two-factor crossed design with interaction. The top plot is the empirical coverage probabilities for 95% one-sided upper confidence intervals. The bottom plot is the empirical coverage probabilities for 95% one-sided lower confidence intervals. Each value used in a box plot corresponds to a particular model design and parameter combination.

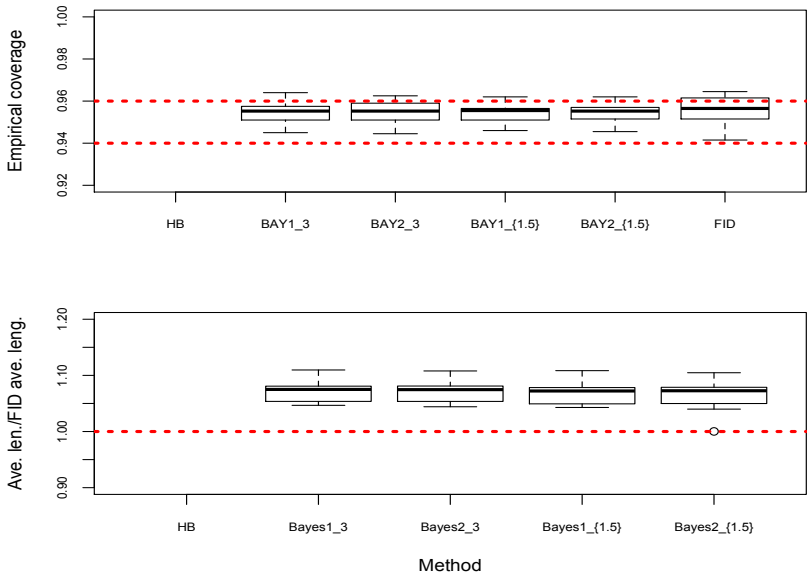


Figure 2.20: Simulation results for σ_ϵ^2 for the two-factor crossed design with interaction. The top plot is the empirical coverage probabilities for 95% two-sided upper confidence intervals. The bottom plot is the average interval lengths divided by the average interval lengths of **FID**. Each value used in a box plot corresponds to a particular model design and parameter combination.

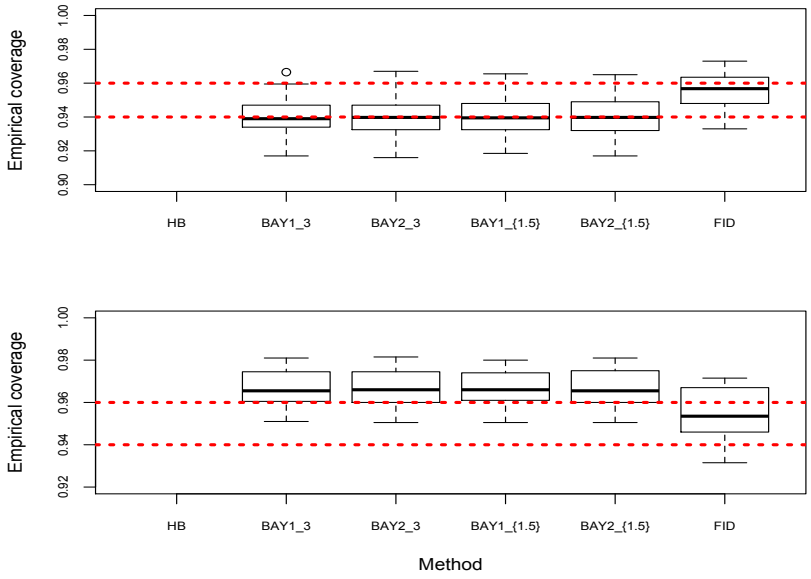


Figure 2.21: Simulation results for σ_ϵ^2 for the two-factor crossed design with interaction. The top plot is the empirical coverage probabilities for 95% one-sided upper confidence intervals. The bottom plot is the empirical coverage probabilities for 95% one-sided lower confidence intervals. Each value used in a box plot corresponds to a particular model design and parameter combination.

for proportions, the inverse sine of the square root of the data, has been taken for the data set. At each test location, data were collected for trees with female parents from different families with one to four measurements for each family and test combination.

Table 2.8 lists 95% confidence intervals for **HB**, $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$, and **FID** based on the actual data. In order to evaluate the empirical coverage of these methods, we perform a simulation study using the REML estimates for all the parameters ($\mu = 0.9349$, $\sigma^2_\alpha = 0.0224$, $\sigma^2_\beta = 0.0274$, $\sigma^2_{\alpha\beta} = 0.000$, and $\sigma^2_\epsilon = 0.0340$). Simulating 2000 independent data sets with the noted parameter values, we find the empirical coverage using **HB**, $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, $\mathbf{BAY2}_3$, and **FID**, and the average lengths of the two-sided intervals. These results are also presented in Table 2.8. All 95% intervals based on the real data capture the REML estimates to two decimal places of rounding. The issue with the Bayesian methods for a variance component at, or very close to, zero is that the possible simulated values will be bounded away from zero unless an adjustment is made to the algorithm. An adjustment such as an MCMC algorithm allowing for point masses at zero would be necessary to alleviate this problem.

HB is within the stated coverage except it becomes slightly liberal for the upper intervals for σ^2_α and the lower intervals for σ^2_β. $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, and $\mathbf{BAY2}_3$ within the stated coverage for the two-sided intervals on σ^2_α and σ^2_β, liberal for the upper intervals, and very conservative for the lower intervals. Because the REML estimate for $\sigma^2_{\alpha\beta}$ was zero, the empirical coverage for the Bayesian intervals would not be expected to perform well without a special adjustment to the simulation procedure (adding the possibility of a point mass at zero). **FID** is within the stated coverage for all intervals except it is slightly liberal for the lower intervals on σ^2_α, and liberal for the upper interval on $\sigma^2_{\alpha\beta}$, which is not surprising with the value of zero on a parameter restricted to be nonnegative. This also highlights the importance of model selection since it is unlikely the interaction term should have been included in the model. Model selection is not addressed in this dissertation, but is a subject of future research. $\mathbf{BAY1}_{1.5}$, $\mathbf{BAY1}_3$, $\mathbf{BAY2}_{1.5}$, and $\mathbf{BAY2}_3$ have the longest average interval lengths. **HB** and **FID** have comparable average interval lengths for $\sigma^2_{\alpha\beta}$, **FID** is shortest for σ^2_α and **HB** is shortest for σ^2_β.

Var. comp.	Method	95% 2-sided C.I.	2-sided./ave. len.	Upper/lower
σ^2_α	**HB**	(0.007, 0.528)	0.940/0.420	0.918/0.958
	BAY1_3	(0.004, 0.330)	0.963/0.513	0.932/0.996
	BAY1_3	(0.004, 0.461)	0.965/0.495	0.937/0.993
	BAY$1_{1.5}$	(0.005, 0.337)	0.963/0.427	0.934/0.993
	BAY$2_{1.5}$	(0.005, 0.312)	0.965/0.382	0.936/0.994
	FID	(0.004, 0.184)	0.961/0.162	0.963/0.937
σ^2_β	**HB**	(0.0002, 0.186)	0.957/0.206	0.976/0.931
	BAY1_3	(0.008, 1.465)	0.957/1.845	0.914/1.000
	BAY2_3	(0.007, 1.023)	0.959/1.630	0.919/0.999
	BAY$1_{1.5}$	(0.008, 0.723)	0.958/1.018	0.918/0.999
	BAY$2_{1.5}$	(0.007, 0.683)	0.961/1.029	0.921/1.000
	FID	(0.006, 0.426)	0.968/0.331	0.963/0.953
$\sigma^2_{\alpha\beta}$	**HB**	(0.000, 0.043)	0.967/0.030	0.942/1.000
	BAY1_3	$(1.41 \times 10^{-6}, 0.030)$	0.000/0.031	0.000/1.000
	BAY2_3	$(5.21 \times 10^{-6}, 0.029)$	0.000/0.031	0.000/1.000
	BAY$1_{1.5}$	$(1.10 \times 10^{-5}, 0.025)$	0.000/0.031	0.000/1.000
	BAY$2_{1.5}$	$(4.62 \times 10^{-6}, 0.028)$	0.000/0.031	0.000/1.000
	FID	(0.000, 0.022)	0.945/0.026	0.887/1.000

Table 2.8: Two-factor crossed with interaction model: real data example

NOTE: The 95% intervals are based on the actual data while the remaining information are the empirical results from 2000 independently generated data set using the REML estimates for each parameter. The results are the empirical coverage and average interval lengths of 95% confidence intervals.

2.5 Conclusion and future research

Even with the long history of inference procedures for normal linear mixed models, a good-performing, unified inference method is lacking. ANOVA-based methods offer, what tends to be, model-specific solutions that may not perform well if the design is unbalanced. While Bayesian methods allow for solutions to very complex models, determining an appropriate prior distribution may not be straightforward, and can be confusing for the non-statistician practitioner. In addition, for the models considered in the simulation study and the priors selected based on recommendations in the literature, the Bayesian interval lengths were not generally competitive with the other methods used in the study. The proposed method allows for confidence interval estimation for all parameters of balanced and unbalanced normal mixed linear models. It is interesting to note that even though more variation was incorporated into the data for the proposed method due to its acknowledgement of known uncertainty using intervals, in the simulation study, the proposed method tended to have conservative coverage, but the average interval lengths were comparable or shorter than the other methods that assumed the data were observed exactly. The currently implemented algorithm is suitable for 9 or fewer parameters, but the method does not limit the number of parameters.

When the data are i.i.d (e.g. when the error effect is the only random component), the confidence intervals based on the generalized fiducial distribution are asymptotically correct (Hannig, 2009a). When the data are not i.i.d., previous experience and simulation results suggests that the generalized fiducial method presented above still leads to asymptotically correct inference as the sample size n increases. This is a topic of future research, and requires a Bernstein-von-Mises type result. To the best of our knowledge, an analogous Bayesian result for the normal linear mixed model setting is not currently available in the literature.

Some other interesting extensions of this work would be to consider spatial dependence in the data. The current algorithm relies on the linearity of the variance components and random effects (see Equation (2.2)). A similar algorithm could be developed where, rather than considering interior polyhedrons of intersecting hyperplanes, we would need to consider hypersurfaces. Another extension would be to allow more choices for the distribution of the

random effects. The main difficulty in implementing a model with non-Gaussian random effects would be in determining an appropriate alteration method.

Chapter 3

Logistic regression with mixed effects

3.1 Summary

In this chapter, the generalized fiducial framework is applied to logistic regression models with both fixed and random effects. After introducing the model and outlining the methodology, a simulation study is described to compare the performance of the proposed methodology with competing methodology.

3.2 Introduction

The logistic regression model is a special case of the generalized linear model introduced in Nelder and Wedderburn (1972) when the response is binary. The logistic regression model has many uses such as inference on the median lethal dose of a drug (Williams, 1986; Harris et al., 1999), prediction of failure (Dalal et al., 1989), and animal mating studies (Schall, 1991). The focus of this chapter is on the logistic regression model with additional random effects of the form,

$$\text{logit}(\mathbf{p}) = \mathbf{X}\beta + \mathbf{VZ} \tag{3.1}$$

with binary response vector $Y = (y_1, \ldots, y_n)^T$ where $\mathbf{p} = (p_1, \ldots, p_n)^T$ is a vector of unknown probabilities of success corresponding to the observations with $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$, $\mathbf{X}$ is the $n \times f$ fixed effects design matrix, β is the f-dimensional vector of unknown fixed effects, $\mathbf{VZ} = \sum_{i=1}^{r} \mathbf{V}_i Z_i$, where Z_i is an l_i-dimensional vector representing each level of random

effect i such that $E(Z_i) = 0$ and, for example, $\text{var}(Z_i) = \sigma_i \mathbf{I}_{l_i}$ where l_i is the number of levels of random effect i, $\mathbf{V}_i$ is the random effects design matrix corresponding to random effect i. It is often assumed that the random effects, Z_i, $i = 1, \ldots, r$ are independent and normally distributed (Jiang, 2007). We have developed the proposed generalized fiducial methodology using the canonical link function, the logit link, but generalized fiducial inference would be applicable to other link functions.

In this chapter, we propose a generalized fiducial approach for inference on the unknown parameters of the equation

$$\text{logit}(\mathbf{p}) = \mathbf{X}\beta + \sum_{i=1}^{r} \sigma_i \sum_{j=1}^{l_i} V_{ij} z_{ij} \tag{3.2}$$

following the notation as Equation (3.1) except that the z_{ij} are i.i.d. $N(0,1)$ and V_{ij} is the random effects design vector corresponding to level j of random effect i. As in the previous chapter, the application of the generalized fiducial paradigm in this setting requires a computational algorithm to obtain a sample from the fiducial distribution of the unknown parameters β and $\sigma = (\sigma_1, \ldots, \sigma_r)^T$. In the following sections, we introduce the generalized fiducial framework and the computational methodology developed, prove the asymptotic consistency of the algorithm, perform a simulation study, and finish with concluding remarks. The simulation study compares the proposed fiducial methodology with competing methodology found in the literature.

3.3 Method

3.3.1 Basic model: estimation of p

In order to understand the generalized fiducial methodology for logistic regression, we begin by introducing one of the main ideas employed by considering the probability of success for a binomial random variable (a special case of Equation (3.2)). Suppose $\mathbf{y}$ is a Bernoulli random variables with success probability $p \in [0,1]$. Then $\mathbf{y}$ can be written as $\mathbf{y} = \mathbf{I}_{[0,p)}(\mathbf{u})$ where $\mathbf{I}_A$ is the indicator random variable for the set A, and $\mathbf{u}$ is a uniform random variable on the

interval from 0 to 1 (written $U(0,1)$). If we now have n i.i.d. $\mathbf{y}$'s, this can be represented as

$$\mathbf{y}_i = \mathbf{I}_{[0,p)}(\mathbf{u}_i)$$

for $i = 1, \ldots, n$ where the $\mathbf{u}_i$'s are i.i.d. $U(0,1)$. If $Y = (y_1, \ldots, y_n)$ is a realization of $\mathbf{y}$ with $s = \sum_{i=1}^{n} y_i$ as the observed number of successes, we note that the true, unobserved p must be within the interval $[u_{s:n}, u_{s+1:n}]$ where $u_{i:m}$ is the ith of m ordered values. More precisely, and following the ideas presented in Example 2, Section 4 of Hannig (2009b), we define the sets $Q(y,u)$ as

$$Q(y,u) = \begin{cases} [0, u_{1:n}] & \text{if } s = 0 \text{ (no successes)}, \\ (u_{n:n}, 1] & \text{if } s = n \text{ (all successes)}, \\ (u_{s:n}, u_{s+1,n}] & \text{if } 0 < s < n, \\ & \text{with } \sum_{i=1}^{n} \mathbf{I}(y_i = 1)\mathbf{I}_{[0,u_{s:n}]}(u_i) = s, \\ \emptyset & \text{otherwise.} \end{cases}$$

The proposed methodology will rely on generating independent copies of the $\mathbf{u}$'s so that the sets $Q(y,u)$ are not empty.

3.3.2 General model: logistic regression with mixed effects

The purpose of the following methodology is to develop an algorithm of which to obtain an approximation to the generalized fiducial distribution of the unknown parameters of Equation (3.2). We use the sequential Monte Carlo framework introduced in Section 1.3. The structural equation for the general model is

$$\begin{aligned} \text{logit}(p_k) &= X_k\beta + \sum_{i=1}^{r}\sum_{j=1}^{l_i} \sigma_i v_{ijk} z_{ij}, \text{ and} \\ y_k &= \mathbf{I}_{[0,p_k)}(u_k), \end{aligned} \tag{3.3}$$

for $k = 1, \ldots, n$, X_t is the kth row of fixed effects design matrix $\mathbf{X}$, and v_{ijk} is the coefficient of the random effects design matrix corresponding to level j of effect i for the kth observation,

$z_{ij} \sim N(0,1)$, and $u_k \sim U(0,1)$. Rather than explicitly generating the u_k's, we will use the distribution of $a_k = \text{logit}(u_k)$ to obtain our particle approximation. Since $a_k = \log(\frac{u_k}{1-u_k})$ with $u_k \sim U(0,1)$, the density of a_k is $f_a(x) = \frac{e^x}{(1+e^x)^2}$.

The goal is to sample the the z_{ij}'s and a_k's so that the structural equation holds. That is, sample $A_{1:t}^{(J)} = (a_1^{(J)}, \ldots, a_t^{(J)})$ and $\mathbf{Z}_{1:t}^{(J)} = (Z_1^{(J)}, \ldots, Z_t^{(J)})$, for $J = 1, \ldots, N$ so that Equation (3.3) is satisfied. Using the set notation developed in the previous chapter, we are seeking a sample of the sets

$$Q_t^{(J)}(Y_{1:t}, A_{1:t}^{(J)}, \mathbf{Z}_{1:t}^{(J)}) = \Bigg\{ (\beta, \sigma) :$$

$$\text{logit}(p_k) = X_k\beta + \sum_{i=1}^{r}\sum_{j=1}^{l_i} \sigma_i v_{ijk} z_{ij}^{(J)}, \text{ and } a_k^{(J)} \begin{bmatrix} \leq \text{logit}(p_k) & \text{if } y_k = 1 \\ > \text{logit}(p_k) & \text{if } y_k = 0 \end{bmatrix}, \text{ for } k = 1, \ldots, t \Bigg\}$$

such that $Q_t^{(J)}$, $J = 1, \ldots, N$ are non-empty and letting $\sigma = (\sigma_1, \ldots, \sigma_r)$.

At time t, the full target distribution can be written as

$$\pi_{1:t} \propto \prod_{i=1}^{t} \left\{ \frac{e^{a_t}}{(1+e^{a_t})^2} \right\} \cdot e^{-\frac{1}{2}\mathbf{Z}_{1:t}^T \cdot \mathbf{Z}_{1:t}} \cdot \mathbf{I}_{\mathbf{C}_t}(A_{1:t}, \mathbf{Z}_{1:t}) \tag{3.4}$$

where $\mathbf{C}_t$ is the set of $A_{1:t}$ and $\mathbf{Z}_{1:t}$ such that there is a solution to the structural equation of (3.3). Since we do not know $\mathbf{C}_t$, we cannot sample directly from our target distribution and must define a proposal distribution. Noting that the $\mathbf{Z}_{1:t}$ can be sampled without restriction, our full proposal distribution at time i is

$$\widetilde{\pi}_{1:t} \propto e^{-\frac{1}{2}\mathbf{Z}_{1:t}^T \cdot \mathbf{Z}_{1:t}} \prod_{j=1}^{f+r} \left\{ \frac{e^{a_j}}{(1+e^{a_j})^2} \right\} \prod_{k=f+r+1}^{t} \left\{ \frac{e^{a_k}}{(1+e^{a_k})^2} \cdot \frac{[\mathbf{I}_{(-\infty, M_k]}(a_k)]^{y_k} \cdot [\mathbf{I}_{[m_k, \infty)}(a_k)]^{1-y_k}}{[F_a(M_k)]^{y_k} \cdot [1 - F_a(m_k)]^{1-y_k}} \right\}$$

where F_a is the cdf of $\mathbf{a}_k$'s, $M_t = \max(\gamma_t)$ and $m_t = \min(\gamma_t)$ with

$$\gamma_t = \Big\{ X_t\beta + \sum_{i=1}^{r}\sum_{j=1}^{l_i} \sigma_i v_{ijt} z_{ij} : (\beta, \sigma) \in Q_{t-1} \Big\}.$$

The conditional proposal distribution at time t is

$$\widetilde{\pi}_{t|1:t-1} \propto e^{-\frac{1}{2}Z_t^T \cdot Z_t} \cdot \frac{e^{a_t}}{(1+e^{a_t})^2} \cdot \left(\frac{\mathbf{I}_{(-\infty, M_t]}(a_t)}{F_a(M_t)} \right)^{y_t} \cdot \left(\frac{\mathbf{I}_{[m_t, \infty)}(a_t)}{1 - F_a(m_t)} \right)^{1-y_t}.$$

The marginal target distribution at time t is

$$\widehat{\pi}_{1:t-1} = \int \pi(A_{1:t}, \mathbf{Z}_{1:t}) d(a_t, Z_t) \propto \pi_{1:t-1} \cdot (F_a(a_t))^{y_t} \cdot (1 - F_a(a_t))^{1-y_t},$$

and it follows that the conditional target distribution at time t is

$$\pi_{t|1:t-1} = \frac{\pi_{1:t}}{\widehat{\pi}_{1:t-1}} \propto e^{-\frac{1}{2}Z_t^T \cdot Z_t} \cdot \frac{e^{a_t}}{(1+e^{a_t})^2} \cdot \left(\frac{\mathbf{I}_{(-\infty, M_t]}(a_t)}{F_a(M_t)} \right)^{y_t} \cdot \left(\frac{\mathbf{I}_{[m_t, \infty)}(a_t)}{1 - F_a(m_t)} \right)^{1-y_t}.$$

The resulting importance weights at time t are

$$\begin{aligned} W_{1:t} = \frac{\pi_{1:t}}{\widetilde{\pi}_{1:t}} &= \frac{\pi_{t|1:t-1} \cdot \widehat{\pi}_{1:t-1}}{\widetilde{\pi}_{t|1:t-1} \cdot \widetilde{\pi}_{1:t-1}} \\ &= \frac{\pi_{1:t-1}}{\widetilde{\pi}_{1:t-1}} \cdot \frac{\pi_{t|1:t-1}(F_a(M_t))^{y_t} \cdot (1 - F_a(m_t))^{1-y_t}}{\widetilde{\pi}_{t|1:t-1}} \\ &= W_{1:t-1} \cdot (F_a(M_t))^{y_t} \cdot (1 - F_a(m_t))^{1-y_t} = W_{1:t-1} \cdot W_t. \end{aligned}$$

3.3.3 General alteration

After each step of the algorithm, the ESS_t from Equation (1.6) is calculated and checked to see if it falls below the designated threshold T (e.g. $T = N/2$). If ESS_t is under the threshold, the particle system is resampled, and the particles are selected for resampling based on their normalized importance weight. Each particle will have a designated number of replications from 0 to N such that the total number of replications assigned to the particle system is N (particles with higher weights are more likely to be replicated). Suppose particle L is selected to be resampled K times.

We begin by defining a decomposition analogous to the decomposition of the particles in the normal linear mixed model setting defined in Equation (2.19). The particle $A_{1:t}$ (sup-

pressing the dependency on J) is decomposed as

$$A_{1:t} = P(P^T A_{1:t}) + (A_{1:t} - PP^T A_{1:t}) = P(P^T A_{1:t}) + Q(Q^T A_{1:t}) = PB + Q\tau, \tag{3.5}$$

where P is the orthogonal basis of $D = [X_{1:t}, \sum_{i=1}^{r} \mathbf{V}_i Z_i]$ that spans the subspace $\mathcal{A}$. Then PP^T is the orthogonal projection onto $\mathcal{A}$ with QQ^T as the projection onto the subspace orthogonal to $\mathcal{A}$. The goal of the alteration step is to resample the variable $B = P^T A_{1:t}$ conditional on $\tau = Q^T A_{1:t}$. The desired sampling distribution is then

$$f_{B|\tau}(B|\tau) = \frac{f_{B,\tau}(B,\tau)}{f_\tau(\tau)} \propto f_{B,\tau}(B,\tau),$$

where f_G is the density for the variable G. Noting that

$$\begin{pmatrix} B \\ \tau \end{pmatrix} = \begin{pmatrix} P^T \\ Q^T \end{pmatrix} A_{1:t},$$

so that

$$A_{1:t} = \begin{pmatrix} P^T \\ Q^T \end{pmatrix}^{-1} \begin{pmatrix} B \\ \tau \end{pmatrix} = \begin{pmatrix} P & Q \end{pmatrix} \begin{pmatrix} B \\ \tau \end{pmatrix} = PB + Q\tau$$

since

$$\begin{pmatrix} P^T \\ Q^T \end{pmatrix}^{-1} = \begin{pmatrix} P & Q \end{pmatrix} \tag{3.6}$$

because P and Q are orthogonal. Hence, the density of $B|\tau$ can be written

$$f_{B|\tau} \propto f_{B,\tau}(B,\tau) \propto f_{A_{1:t}}((P,Q)\cdot(B,\tau)^T |\det(P,Q)| = f_{A_{1:t}}(PB + Q\tau). \tag{3.7}$$

The fact that $|\det(P,Q)| = 1$ follows from two properties of determinants: (i) $\det(X) = \det(X^T)$, and (ii) $\det(X) = 1/\det(X^{-1})$ (when the inverse exists). Because of Equation (3.6), the result follows.

Now $A_{1:t}$ can be altered by generating a new B, say $\widetilde{B}$, from the density of Equation (3.7)

by a Monte Carlo sampling method such as importance sampling. Using importance sampling requires the generation of a sample of $\widetilde{B}$ from some proposal distribution, and then selecting a new $\widetilde{B}$ according to the determined importance weights. For example, if $\widetilde{B}$ is generated from a Cauchy distribution, the unnormalized importance weight for particles $K = 1, \ldots, N^*$ would be

$$W^{(K)} = \left(e^{P\widetilde{B}^{(K)}+Q\tau}/(1+e^{P\widetilde{B}^{(K)}+Q\tau})^2\right)\left(1+(P\widetilde{B}^{(K)}+Q\tau)^2\right),$$

where P, Q, and τ are known from the decomposition of Equation (3.5).

After an updated $A_{1:t}$ is selected, denoted $\widetilde{A}_{1:t} = P\widetilde{B} + Q\tau$, the parameters (β, σ) can be updated to $(\widetilde{\beta}, \widetilde{\sigma})$ using $(\widetilde{\beta}, \widetilde{\sigma})^T = (\beta, \sigma)^T - (D^TD)^{-1}D^TP(B - \widetilde{B})$. This easily follows by

$$\begin{aligned}
D \cdot (\beta, \sigma)^T - A_{1:t} &= D \cdot (\widetilde{\beta}, \widetilde{\sigma})^T - \widetilde{A}_{1:t} \\
D \cdot (\beta, \sigma)^T - PB - Q\tau &= D \cdot (\widetilde{\beta}, \widetilde{\sigma})^T - P\widetilde{B} - Q\tau \\
D \cdot (\beta, \sigma)^T - PB + P\widetilde{B} &= D \cdot (\widetilde{\beta}, \widetilde{\sigma})^T \\
(\beta, \sigma)^T - (D^TD)^{-1}D^TP(B - \widetilde{B}) &= (\widetilde{\beta}, \widetilde{\sigma})^T,
\end{aligned}$$

where $D = [X_{1:t}, \sum_{i=1}^{r} \mathbf{V}_i Z_i]$ as above. For particle, $J = 1, \ldots, N$, after resampling the $A_{1:t}^{(J)}$, the random effects $Z_j^{(J)}$, $j = 1, \ldots, r$ are resampled in the same manner as the particles in the normal linear mixed model setting described in Section 2.3.4.

3.3.4 Convergence of algorithm

The proposed algorithm targets the generalized fiducial distribution of the unknown parameters of the logistic regression with mixed effects model of (3.2) with the target distribution displayed in (3.4). The following theorem confirms that the weighted particle system from the proposed algorithm is consistent as the particle sample size N approaches infinity. The proof relies on the ideas employed in the proof of Theorem 2.3.1 from Chapter 2.

Theorem 3.3.1. *Given a weighted sample* $\{A_{1:t}^{(J)}, \mathbf{Z}_{1:n}^{(J)}, W_{1:n}^{(J)}\}_{J=1}^{N}$ *obtained using the algorithm*

presented above targeting (3.4)*, then for any bounded, measurable function* f*,*

$$\Big(\sum_{I=1}^{N} W_{1:n}^{(I)}\Big)^{-1} \sum_{J=1}^{N} f(A_{1:n}^{(J)}, \mathbf{Z}_{1:n}^{(J)}) W_{1:n}^{(J)} \xrightarrow{P} \int f(A_{1:n}, \mathbf{Z}_{1:n}) \pi_{1:n} = \pi_{1:n} f(A_{1:n}, \mathbf{Z}_{1:n}),$$

as $N \longrightarrow \infty$*.*

Notation and definitions The following assumes the notation developed in Section 3.3.2. Because this result is analogous to Theorem 2.3.1, we begin by modifying some definitions used previously. First, we define two sigma-fields $\mathcal{F}_0 \triangleq \sigma(\{A_{1:t-1}^{(J)}, \mathbf{Z}_{1:t-1}^{(J)}\}_{J=1}^{N}, Y_{1:t})$ and $\mathcal{F}_J \triangleq \mathcal{F}_0 \vee \sigma(\{A_{1:t}^{(K)}, \mathbf{Z}_{1:t}^{(K)}\}_{1 \leq K \leq J}, Y_{1:t})$, for $J = 1, \ldots, N$. Though the sigma fields are dependent on t, we suppress the notation because the current state of the particle system will be at time t. At time $t-1$, the particle system is targeting probability measure $\pi_{1:t-1}$ on $(\Theta_{1:t-1}, \mathcal{B}(\Theta_{1:t-1}))$. This measure is approximated by our particle system, $\{A_{1:t}^{(J)}, \mathbf{Z}_{1:t-1}^{(J)}, W_{1:t-1}^{(J)}\}_{J=1}^{N}$ for $J = 1, \ldots, N$. Furthermore, $(A_{1:t}^{(J)}, \mathbf{Z}_{1:t-1}^{(J)}) \in \Theta_{1:t-1}$, and $W_{1:t-1}^{(J)} \geq 0$ and such that $\Omega_{t-1} \sum_{J=1}^{N} W_{1:t-1}^{(J)} = 1$. Note that π will, with the appropriate subscript indexing the time, be used to denote both the measure and density. The importance sampling step transforms the current particle system to one that is targeting $\pi_{1:t}$ on $(\Theta_{1:t}, \mathcal{B}(\Theta_{1:t}))$, denoted by $\{A_{1:t}^{(J)}, \mathbf{Z}_{1:t}^{(J)}, W_{1:t}^{(J)}\}_{J=1}^{N}$.

We define the proper set $B_t \triangleq \{f \in L^1(\Theta_{1:t}, \pi_{1:t}), F(\cdot, |f|) \in B_{t-1}\}$, where

$$\begin{aligned} &F(A_{1:t-1}, \mathbf{Z}_{1:t-1}, f) \\ &\quad = \int_{\Theta_t} f(A_{1:t-1}, A_t, \mathbf{Z}_{1:t-1}, Z_t) \cdot \mathbf{I}_{\star}(Z_t) \times \\ &\qquad [\mathbf{I}_{(-\infty, M_t]}(a_t)]^{y_t} [\mathbf{I}_{[m_t, \infty)}(a_t)]^{1-y_t} \cdot [F_a(M_t)]^{y_t} \cdot [1 - F_a(m_t)]^{1-y_t} d(a_t, Z_t) \end{aligned}$$

where $\mathbf{I}_{\star}$ indicates the lack of restriction on the Z_t's.

We must show that the particle system is consistent after each stage of the algorithm: sampling, resampling, and alteration. Consistency after sampling and resampling follow for the same reasons as in Lemma's 2.3.1 and 2.3.2, so here we focus on proving consistency after the alteration step described in Section 3.3.3.

Proof. Let the altered particles be represented as $\{\widetilde{A}_{1:t}^{(J)}, \widetilde{\mathbf{Z}}_{1:t}^{(J)}, 1\}_{J=1}^{N}$ and $\{A_{1:t}^{(J)}, \mathbf{Z}_{1:t}^{(J)}, 1\}_{J=1}^{N}$

as the "original" particles after resampling the system. Furthermore, since the alteration of the $\mathbf{Z}_{1:t}^{(J)}$ is completed in the same way as addressed in Lemma 2.3.3, and the two alteration steps are done independently, we only have left to prove consistency after altering the $A_{1:t}^{(J)}$'s, $J = 1, \ldots, N$; reference to the altered and unaltered $\mathbf{Z}_{1:t}$ is suppressed for the remainder of this proof.

Letting $f \in B_t$ with $\widetilde{B}$, P, Q, and τ as defined below Equation (3.5), we define the following conditional expectation

$$E[f(\widetilde{A}_{1:t}^{(J)})|\widetilde{\mathcal{F}}_{J-1}] = \int f\big(P\widetilde{B} + QQ^T A_{1:t}\big) d\pi_{\widetilde{B}|\tau} = \int f\big(P\widetilde{B} + Q\tau\big) d\pi_{\widetilde{B}|\tau} \triangleq h_f(A_{1:t}^{(J)}),$$

where $\widetilde{\mathcal{F}}_J \triangleq \widetilde{\mathcal{F}}_0 \vee \sigma(\{\widetilde{A}_{1:t}^{(K)}\}_{1 \le K \le J}, Y_{1:t})$, and $\widetilde{\mathcal{F}}_0 \triangleq \sigma(\{A_{1:t}^{(J)}\}_{J=1}^N, Y_{1:t})$ for $J = 1, \ldots, N$. That is, $h_f(\widetilde{A}_{1:t}^{(J)})$ is the conditional expectation of the altered particles with respect to the conditional distribution of $\widetilde{B}|\tau$ denoted $\pi_{\widetilde{B}|\tau}$. Recall that B is defined by a decomposition of the original particle, call it L, selected to be resampled, $A_{1:t}^{(L)}$. The $\widetilde{B}$ is a random variable to be resampled according to the target distribution of $B|\tau$ with the $\tau_{1:t}^{(J)}$ considered fixed so that $\widetilde{A}_{1:t} = P\widetilde{B} + Q\tau_{1:t}$.

The proof will follow once we show

$$\int h_f(A_{1:t}) d\pi_{1:t}(A_{1:t}) = \int f(A_{1:t}) d\pi_{1:t}(A_{1:t}), \tag{3.8}$$

because everything else, including how to pick f, will follow as before given the updated definitions. Relation (3.8) holds because

$$\begin{aligned}
\int h_f(A_{1:t}) d\pi_{1:t} &= \int h_f^*(\tau_{1:t}) d\pi_\tau = \int \Big(\int f(P\widetilde{B} + Q\tau_{1:t}) d\pi_{\widetilde{B}|\tau} \Big) d\pi_\tau \\
&= \int \Big(\int f(PB + Q\tau_{1:t}) d\pi_{B|\tau} \Big) d\pi_\tau = \int f(PB + Q\tau_{1:t}) d\pi_{B|\tau} \times d\pi_\tau \\
&= \int f(A_{1:t}) d\pi_{1:t} = E_{\pi_{1:t}} f(A_{1:t}).
\end{aligned}$$

where $h_f^*(\tau_{1:t}) = h_f(A_{1:t})$. □

3.4 Simulation study and applications

We have performed a short simulation study to reveal some of the benefits of the methodology outlined in the previous section. First we consider a logistic dose-response model for inference on the median lethal dose. We compare the empirical performance of two-sided confidence intervals using the generalized fiducial method with several popular, competing methods analyzed in the literature (Williams, 1986; Sitter and Wu, 1993; Harris et al., 1999; Huang et al., 2002; Huang, 2001). We also consider two models each with two fixed effects and one random effect.

3.4.1 Median lethal dose

The median lethal dose (LD50) is the amount of a substance, such as a drug, that is expected to kill half of its users. This measure is important when studying the level of toxicity in biological experiments that are attempting to understand the effects a drug or other substance has on some defined living population. Trevan (1951) has a brief account of the early history of toxicity experiments including the introduction of the study of the LD50 in quantal assays, which are studies with a death or survival response. In fact, the LD50 was introduced in Trevan (1927), in which it was first suggested to use the LD50 rather than other measures of toxicity. To underscore the relevance of inference on the LD50, Kreger (1992), which is a special reference brief on how these tests can be administered, lists over 250 citations where the LD50 or an alternative was studied (including citations addressing animal testing laws) between January of 1980 and March of 1992.

The results of LD50 experiments can be studied and modeled using logistic regression as

$$\text{logit}(p_i) = \beta_1(x_i - \mu) \tag{3.9}$$

where p_i is the probability of a response (e.g. death) at a particular dose administration x_i for $i = 1, \ldots, k$, and μ is the median lethal dose (i.e. the dosage, x_j, such that $p_j = 0.5$). The x_i, $i = 1, \ldots, k$ are known while β_1 and μ are fixed effects that are unknown. This can be

written in the form of Equation (3.1) as

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_i \tag{3.10}$$

for $i = 1, \ldots, k$ where the LD50 is $\mu = -\beta_0/\beta_1$.

In this section, we study the performance of 95% two-sided confidence intervals on the median lethal dose using the proposed generalized fiducial method compared to three competing methods found in the literature. The generalized fiducial method will be denoted by **FID**, and the three competing methods for defining two-sided confidence intervals used in this study are the delta method (**DELTA**), Fieller's method (**FIELLER**), and the likelihood ratio method (**LR**). As Faraggi et al. (2003) notes, these methods tend to be the recommended approaches for confidence intervals on the LD50. Intervals based on adjustments to the competing methods, including the Bartlett adjustment to the likelihood ratio method, selected for our study have also been attempted by others, but have not been found to have a significant impact on performance (Harris et al., 1999; Huang et al., 2002). A bootstrapping procedure was studied in Huang (2001), but found to be liberal for small sample sizes, and an approximate bootstrap procedure was analyzed in Faraggi et al. (2003), but found to be overly conservative. In this simulation study, we do not compare the proposed method to a bootstrapping procedure, but is a future interest to find a better performing bootstrapping procedure.

The model form of Equation (3.10) will be used to find 95% two-sided confidence intervals on the LD50, μ, explicitly defined in Equation (3.9). **DELTA** and **FIELLER** use maximum likelihood estimators in order to define their confidence intervals. Let $\widehat{\beta}_0$ and $\widehat{\beta}_1$ represent the maximum likelihood estimates for β_0 and β_1, respectively, when they exist. Define $\widehat{\mu} = -\widehat{\beta}_0/\widehat{\beta}_1$. Let $v_{11} = \text{var}(\widehat{\beta}_0)$, $v_{22} = \text{var}(\widehat{\beta}_1)$, and $v_{12} = \text{cov}(\widehat{\beta}_0, \widehat{\beta}_1)$ represent the asymptotic variances and covariance of the corresponding estimators.

Delta method (DELTA)

A $100(1-\alpha)\%$ two-sided confidence interval using the delta method is constructed using the relation $\widehat{\mu} = g(\widehat{\beta}_0, \widehat{\beta}_1) = -\widehat{\beta}_0/\widehat{\beta}_1$, where $\text{var}(\widehat{\mu}) = \text{var}(g(\widehat{\beta}_0, \widehat{\beta}_1))$ is approximated using a first - order Taylor series approximation. This results in the interval

$$\widehat{\mu} \pm z_{1-\alpha/2}\widehat{\beta}_1^{-2}(v_{22} + 2\widehat{\mu}v_{12} + \widehat{\mu}^2 v_{22})$$

where z_c is the c-th quantile of the standard normal distribution.

Fieller's method (FIELLER)

Fieller's Theorem (Fieller, 1954) provides a method for constructing confidence regions on the ratio of normal variables. Since $\widehat{\mu}$ is the ratio of two maximum likelihood estimators, Fieller's Theorem can be used as an approximate procedure. The resulting confidence region is defined as the set

$$\left\{\mu_0 : \left|\frac{\widehat{\beta}_0 - \widehat{\beta}_1\mu_0}{v_{22} + 2\mu_0 v_{12} + \mu_0^2 v_{22}}\right| < z_{1-\alpha/2}\right\}, \tag{3.11}$$

which is derived from noting the equivalence of the null hypothesis $H_0 : \mu = \mu_0$ to $H_0 : \beta_0 + \beta_1\mu_0 = 0$. This can be implemented by following Huang (2001)'s notation for the set in Equation (3.11) as

$$\frac{\widehat{\mu} + gv_{12}/v_{22}}{1-g} \pm \frac{z_{1-\alpha/2}}{\widehat{\beta}_1(1-g)}\Big(v_{11} + 2\widehat{\mu}v_{12} + \widehat{\mu}^2 v_{22} - g(v_{11} - v_{12}^2/v_{22})\Big)^2$$

where $g = z_{1-\alpha/2}^2 v_{22}/\widehat{\beta}_1^2$.

Likelihood ratio method (LR)

Letting L_0 be the likelihood of the data under the null hypothesis $H_0 : \mu = \mu_0$, and L_1 be the likelihood of the data under the alternative hypothesis $H_1 : \mu \neq \mu_0$, then $-2\log\left(\frac{L_0}{L_1}\right) = -2\log(L_0) + 2\log(L_1) = D(\mu_0) - D(\mu)$ converges to a chi-squared distribution with one degree

of freedom. A $100(1-\alpha)\%$ confidence region can then be defined as

$$\left\{\mu_0 : D(\mu_0) - D(\mu) < z^2_{1-\alpha/2}\right\}.$$

For a given dataset, $D(\mu) = D(\widehat{\mu})$, and then determining the confidence regions requires some form of a grid-search algorithm as noted in Williams (1986).

Existence of confidence intervals

Special case 1: partial responses As previously noted, **DELTA** and **FIELLER** require the existence of the maximum likelihood estimates in order to define their confidence intervals on μ. Unfortunately, maximum likelihood estimates do not exist if there are zero or one partial responses. To define a partial response, let s_i be the number of successes out of n_i for a given dose x_i, $i = 1, \ldots, k$. A partial response occurs if $0 < s_i < n_i$ for some $i = 1, \ldots, k$. A single partial response occurs if $0 < s_j < n_j$ and $s_i = 0$ or n_i for $i = 1, \ldots, j-1, j+1, \ldots, k$. This issue tends to occur for small samples when the slope of the dose-response curve (the inverse logit curve where the doses are allowed to vary) is very steep. This will be referred to as *special case 1*.

Special case 2: Wald test If a standard Wald test of the hypothesis $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ is not rejected at a level α for a given dataset, Fieller's confidence regions end up as either disjoint intervals or the entire real line. This will be called *special case 2*.

Special case 3: LR test Similar to special case 2, if the likelihood ratio test is not able to reject the hypothesis $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$, **LR** does not provide a confidence region that is an interval. This will be referred to as *special case 3*.

Williams (1986) advocated the use of the likelihood ratio method for confidence intervals on the median lethal dose. To illustrate the issues with relying on Fieller's Theorem to define confidence intervals, Williams (1986) picked six data set examples as listed in Table 3.1. Each data set has 5 observations within each dose level of $(-2, -1, 0, 1, 2)$. Example A has no special cases and so the intervals are displayed for **DELTA**, **FIELLER**, **LR** and **FID**.

Special case 2 happens in both example B and C resulting in the whole real line and disjoint intervals, respectively, for **FIELLER**. Example D has special case 2 and 3 resulting in disjoint intervals for **FIELLER** and **LR**. Finally, examples E and F have special cases 1 and 2 and so no intervals are produced for **DELTA** and **FIELLER**.

Intervals for **FID** are displayed in Table 3.1 using 2,500, 5,000, 7,500, 10,000, and 20,000 particles, and does not appear to be affected by the special cases, and produces finite intervals. We note this as a benefit of **FID** over **DELTA**, **FIELLER**, and **LR**. From Table 3.1, we see that in examples A, E, and F, even 1,000 particles seems to be enough to produce a converged confidence interval as evidenced by the small discrepancy between the 1,000-particle intervals and the corresponding 20,000-particle intervals. Example B the intervals defined using 2,500 to 20,000 particles are reasonably similar; however the interval constructed with 1,000 particles is clearly distinct suggesting that the algorithm had not yet converged. Example C seems to fluctuate a bit among the confidence intervals defined using varying number of particles, but the fluctuations are not too drastic. In Example D, like with Example B, 1,000 particles is not sufficient, but the intervals are reasonably stable for the larger particle sample sizes. The overall message seems to be that, in practice, one should try at least two different particle sample sizes to gauge convergence until a better methodology is developed for determining an appropriate particle sample size.

Simulation study design

A simulation study has been run in order to compare the frequentist performance of 95% two-sided confidence intervals on μ defined using the proposed method, **FID**, with the performance of the competing methods **DELTA**, **FIELLER**, and **LR**. Performance is based on the empirical coverage of $(1-\alpha)100\%$ two-sided confidence intervals and median interval length for the median lethal dose, μ. We define a two-sided equal-tailed confidence interval on μ as the interval $[L_{\alpha/2}, U_{\alpha/2}]$ such that $P(L_{\alpha/2} \leq \theta \leq U_{\alpha/2}) = 1-\alpha$.

For each model and parameter design combination, 1,000 independent data sets were generated, and 5,000 particles were simulated for the proposed method. Based on the normal approximation to the binomial distribution, we will consider empirical coverage between 93.6%

Ex.	Deaths	$\widehat{\mu}$	**DELTA**	**FIELLER**	**LR**	**FID**	N
A	1, 3, 2, 4, 5	-0.61	(-1.66, 0.44)	(-3.36, 0.75)	(-2.63, 0.49)	(-2.94, 0.68)	1,000
						(-2.95, 0.71)	2,500
						(-2.87, 0.74)	5,000
						(-2.84, 0.75)	7,500
						(-2.95, 0.77)	10,000
						(-3.02, 0.70)	20,000
B	2, 2, 4, 3, 5	-1.02	(-2.49, 0.45)	(-∞, 0.59)$\cup$ (62.76, ∞)	(-12.34, 0.33)	(-0.23, 1.77)	1,000
						(-8.13, 1.77)	2,500
						(-6.94, 1.64)	5,000
						(-7.49, 1.57)	7,500
						(-7.04, 2.00)	10,000
						(-7.06, 1.53)	20,000
C	1, 3, 2, 4, 4	-0.46	(-1.86, .95)	($-\infty$, ∞)	(-11.59, 1.65)	(-4.51, 2.78)	1,000
						(-4.56, 2.41)	2,500
						(-6.54, 3.56)	5,000
						(-4.96, 4.08)	7,500
						(-4.93, 3.20)	10,000
						(-5.54, 2.93)	20,000
D	3, 2, 3, 4, 5	-1.45	(-3.33, 0.44)	(∞, 0.16)$\cup$ (6.42, ∞)	(-∞, 0.01)$\cup$ (24.80, ∞)	(-21.65, 10.58)	1,000
						(-12.74, 8.56)	2,500
						(-14.21, 8.91)	5,000
						(-13.41, 9.04)	7,500
						(-11.92, 7.75)	10,000
						(-14.20, 8.36)	20,000
E	0, 0, 4, 5, 5	NA	Indetermin.	Indetermin.	(-0.70, 0.11)	(-0.91, 0.26)	1,000
						(-0.89, 0.30)	2,500
						(-0.90, 0.26)	5,000
						(-0.88, 0.28)	7,500
						(-0.91, 0.24)	10,000
						(-0.90, 0.24)	20,000
F	0, 0, 5, 5, 5	NA	Indetermin.	Indetermin.	(-1.00, 0.00)	(-0.99, -0.04)	1,000
						(-0.98, -0.01)	2,500
						(-1.00, -0.01)	5,000
						(-1.00, -0.01)	7,500
						(-1.00, -0.02)	10,000
						(-1.01, -0.02)	20,000

Table 3.1: The 95% two-sided confidence intervals based on designs from Williams (1986). Each example uses doses $(-2, -1, 0, 1, 2)$ with 5 observations per dose level. The intervals for **FID** are displayed with varying number of particles to illustrate the variation in intervals as the number of particles increases. The values listed under $\widehat{\mu}$ are the maximum likelihood estimates for the corresponding example data set.

and 96.4% appropriate for 95% two-sided confidence intervals.

The model designs selected for this study are displayed in Table 3.2. Designs 1 and 2 were first defined in Williams (1986), and also used in Harris et al. (1999) and Huang (2001). The dose levels of Designs 3, 4, and 5 were selected from Huang et al. (2002). The dose levels used in Design 6 were first presented in Abdelbasit and Plackett (1983) for experiments on the toxicity of oil to *Sitophilus Granarius*. The same dose levels were subsequently used in Sitter and Wu (1993), Harris et al. (1999), Huang (2001), Huang et al. (2002).

Design	β_1	μ	x_i
1	2.0	3.0	1, 2, 3, 4, 5
2	1.0	4.0	1, 2, 3, 4, 5
3	2.0	5.1	2.056, 3.233, 4.411
			5.589, 6.767, 7.944
4	1.0	4.9	2.056, 3.233, 4.411
			5.589, 6.767, 7.944
5	1.0	2.0	0.000, 0.463, 3.045
			3.296, 3.584, 3.932
			4.394, 5.142
6	7.0	0.1	-0.3098, -0.2147, -0.1487
			-0.0809, -0.0362, 0.0864
			0.1523, 0.2304, 0.2810

Table 3.2: Model designs used in the simulation study for the LD50 as displayed in Equation (3.9).

Simulation study results

The number of occurrences of the three special cases are listed in Table 3.3. Designs 1, 2, and 3 have the greatest number of occurrences of special cases when the cluster size is low. The empirical coverage of **DELTA**, **FIELLER**. and **LR** tend to have very liberal coverage for the smaller cluster sizes when the special cases are included as displayed in Figures 3.1 and 3.2. In those scenarios, **FID** is within the stated coverage to slightly conservative. When the cluster size is 20 and the special cases are included, **DELTA**, **FIELLER**. and **LR** better maintain the stated coverage with **DELTA** becoming liberal in two settings as displayed in Figure 3.3 while **FID** is within the stated coverage.

When the special cases are all excluded from the summary (i.e. the independent data sets that produced one or more of the three special cases were removed from the analysis for all methods), the performance of the methods are analyzed based on empirical coverage of the 95% two-sided confidence intervals and on the median lengths of those intervals. The results are displayed in Figures 3.4, 3.5, and 3.6 for cluster sizes 6, 10, and 20, respectively. **DELTA** still tends to be the most liberal of all the methods in this study for the smaller cluster sizes, but its empirical coverage improves as the cluster size increases. Not surprisingly, **DELTA** also has the shortest median interval lengths. **FIELLER** tends to be conservative for the smaller cluster sizes, but is within the stated coverage for the largest cluster size. **FIELLER** tends to have the longest median intervals for a cluster size of 6 and 10, but then is comparable in length to **FID** when the cluster size is 20. **LR** has acceptable empirical coverage for all cluster sizes with a couple instances of liberal coverage. **LR**'s median interval lengths are, in general, shorter than **FIELLER** and **FID**. **FID** is slightly conservative when the cluster size is 6, but then tends to maintain correct coverage when the cluster size is 10 or 20. The median lengths for **FID** are generally longer than **DELTA** and **LR**, but shorter than **FIELLER** except when the cluster size is 20 where the median interval lengths of **FID** and **FIELLER** are comparable.

Throughout this simulation study, the competing methods do not consistently provide acceptable coverage for 95% two-sided confidence intervals except when the sample size is large. Additionally, a confidence interval may not even be calculable or finite. Though **FID** does not have the overall shortest median interval lengths, it does consistently provide correct, or mildly conservative, coverage along with a solution for all the designs.

3.4.2 Mixed effects design

In this section, a simulation study using two different designs of Model (3.3) are considered. Both designs have two fixed effects and one normally distributed random effect.

The performance of the proposed method, **FID**, is analyzed on two levels. First, we compare the empirical coverage of 95% lower, upper, and two-sided confidence intervals and median interval lengths with a Maximum likelihood (ML) approach (discussed below). A

Design	Cluster size	Case 1	Case 2	Case 3
1	6	179	173	0
	10	41	41	0
	20	1	1	1
2	6	16	147	92
	10	0	22	16
	20	0	0	0
3	6	289	281	0
	10	71	70	0
	20	4	4	0
4	6	8	9	2
	10	0	0	0
	20	0	0	0
5	6	5	17	0
	10	0	1	0
	20	0	0	0
6	6	0	11	7
	10	0	0	0
	20	0	0	0

Table 3.3: Count of special cases per model design and cluster size of Table 3.2 for the LD50 using 1000 independent data sets.

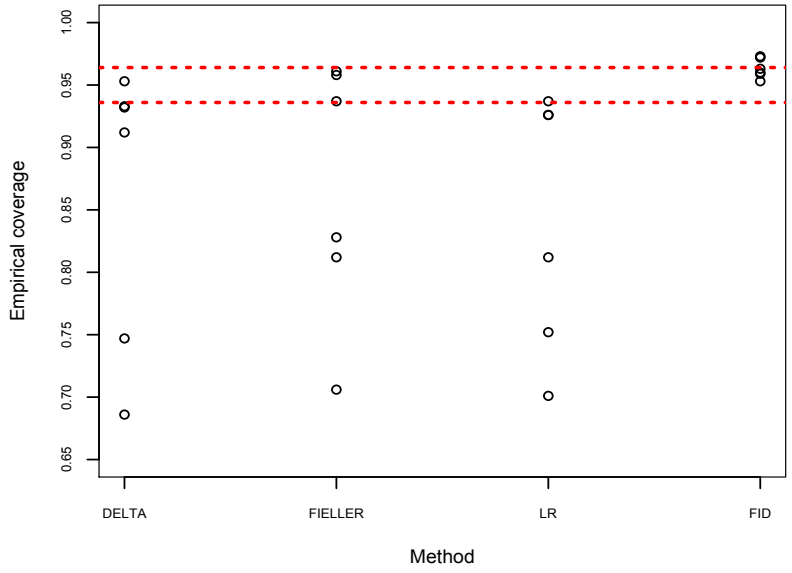


Figure 3.1: Simulation results summary for Designs 1 - 6 of Table 3.2 with a cluster size of 6. This summary includes the special cases The plot displays the empirical coverage of 95% two-sided confidence intervals on the median lethal dose, μ.

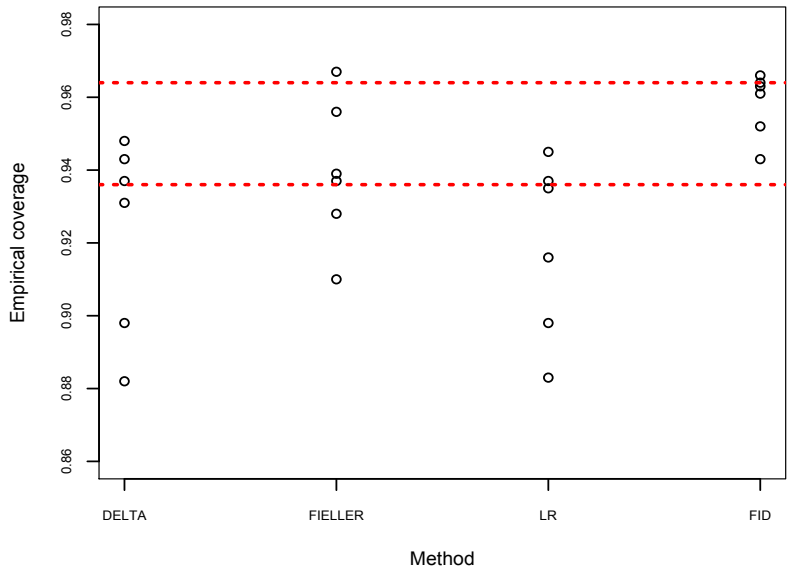


Figure 3.2: Simulation results summary for Designs 1 - 6 of Table 3.2 with a cluster size of 10. This summary includes the special cases The plot displays the empirical coverage of 95% two-sided confidence intervals on the median lethal dose, μ.

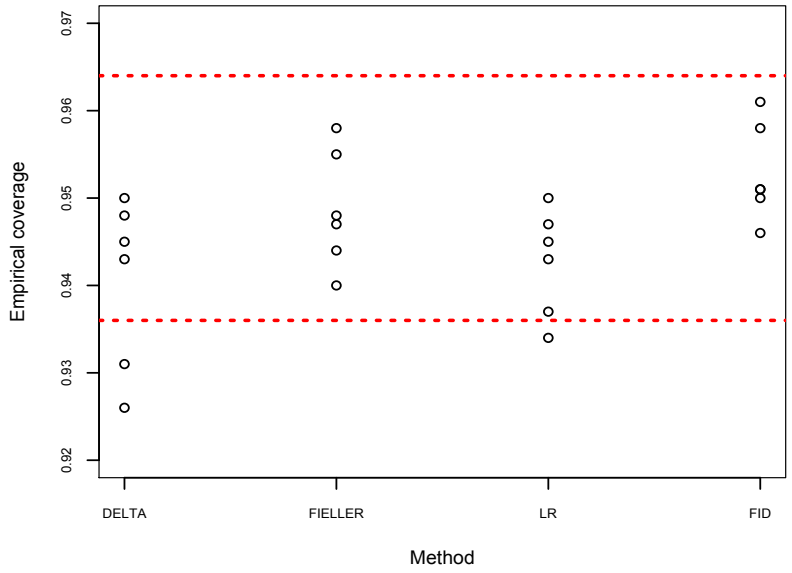


Figure 3.3: Simulation results summary for Designs 1 - 6 of Table 3.2 with a cluster size of 20. This summary includes the special cases The plot displays the empirical coverage of 95% two-sided confidence intervals on the median lethal dose, μ.

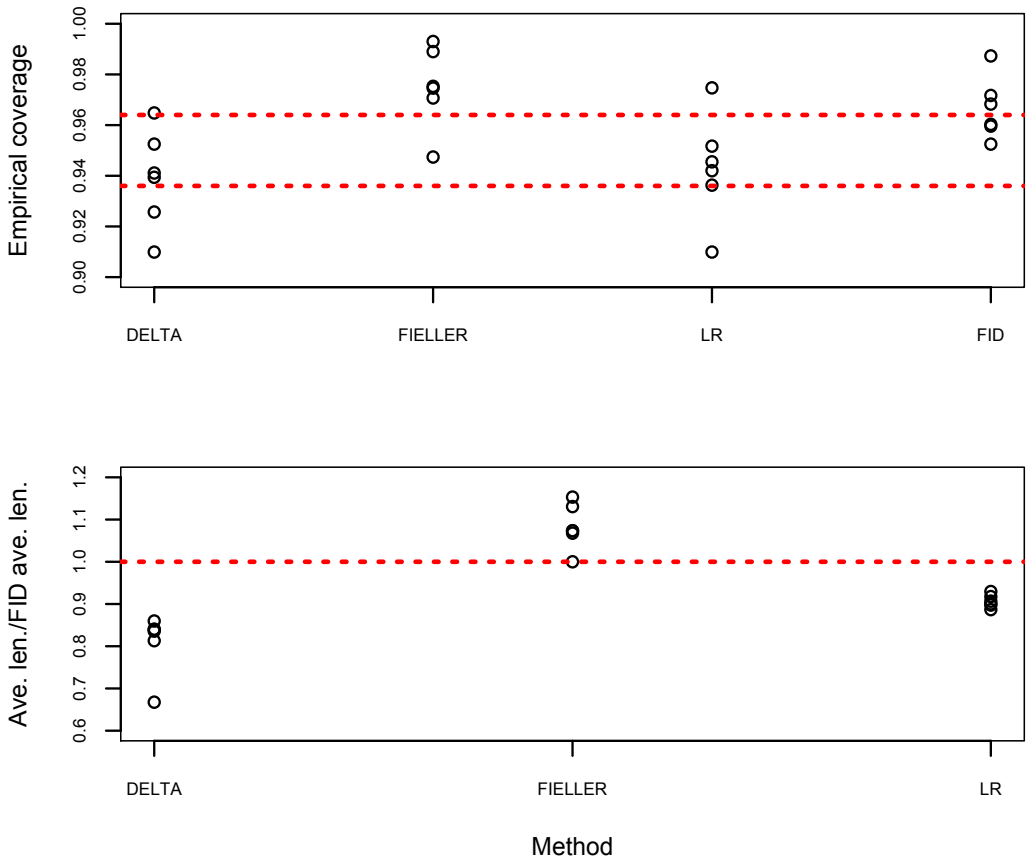


Figure 3.4: Simulation results summary for Designs 1 - 6 of Table 3.2 with a cluster size of 6. This summary excludes the special cases The top plot is the empirical coverage of 95% two-sided confidence intervals on the median lethal dose, μ. The bottom plot is the corresponding median confidence interval length divided by the median confidence interval length of **FID**.

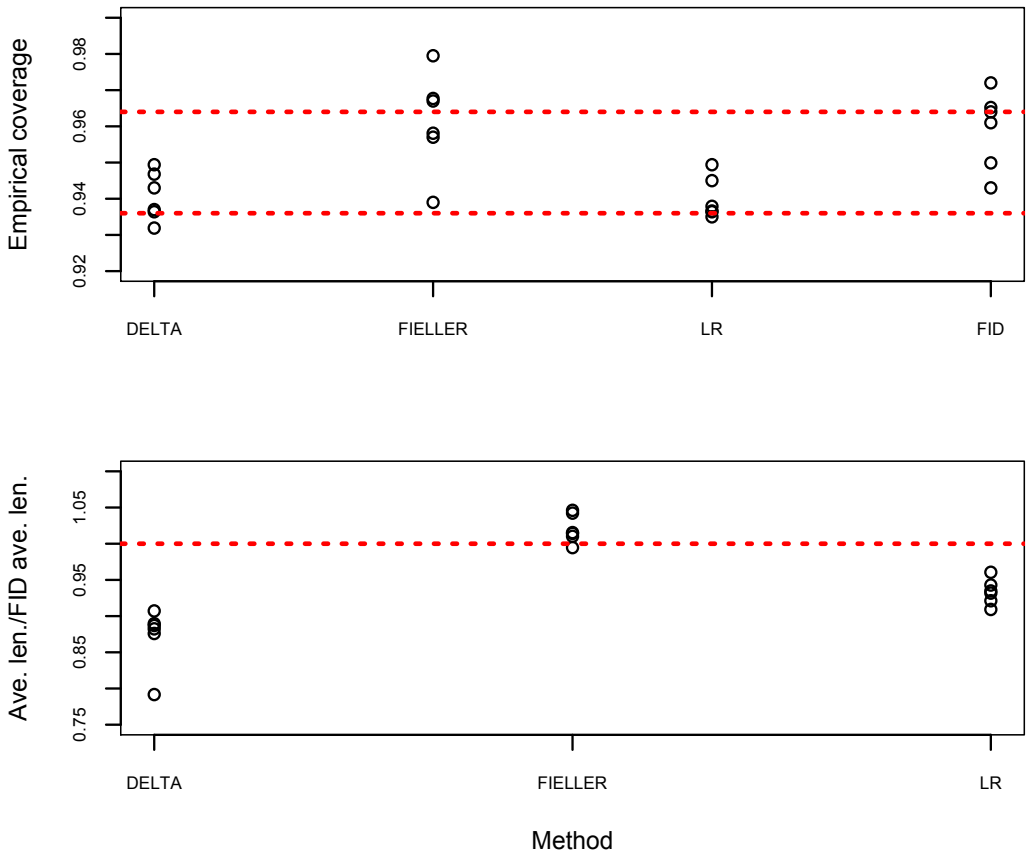


Figure 3.5: Simulation results summary for Designs 1 - 6 of Table 3.2 with a cluster size of 10. This summary excludes the special cases The top plot is the empirical coverage of 95% two-sided confidence intervals on the median lethal dose, μ. The bottom plot is the corresponding median confidence interval length divided by the median confidence interval length of **FID**.

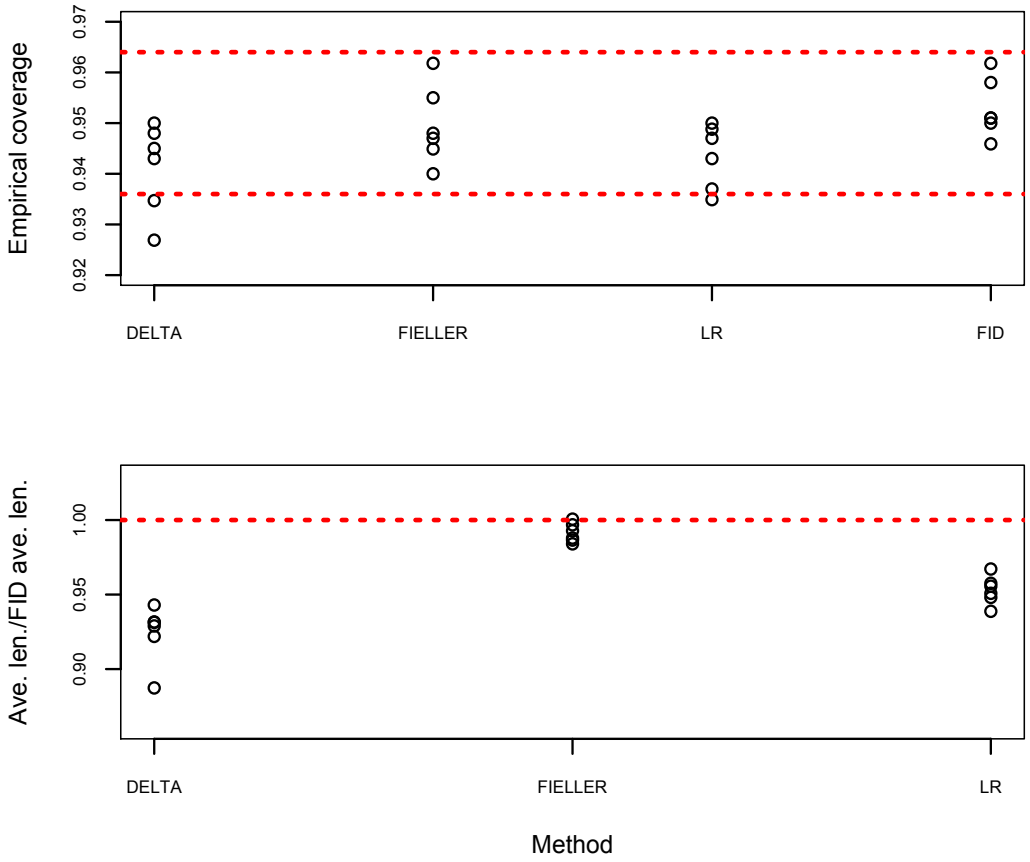


Figure 3.6: Simulation results summary for Designs 1 - 6 of Table 3.2 with a cluster size of 20. This summary excludes the special cases The top plot is the empirical coverage of 95% two-sided confidence intervals on the median lethal dose, μ. The bottom plot is the corresponding median confidence interval length divided by the median confidence interval length of **FID**.

lower-tailed $(1-\alpha)100\%$ confidence interval on θ is the interval $(-\infty, U_\alpha]$ such that $P(-\infty < \theta \leq U_\alpha) = 1-\alpha$, an upper-tailed $(1-\alpha)100\%$ confidence interval on θ as the interval $[L_\alpha, \infty)$ such that $P(L_\alpha \leq \theta < \infty) = 1-\alpha$, and a two-sided equal-tailed confidence interval on θ as the interval $[L_{\alpha/2}, U_{\alpha/2}]$ such that $P(L_{\alpha/2} \leq \theta \leq U_{\alpha/2}) = 1-\alpha$. The second level of the study is to compare estimates to a ML-based approach introduced below. ML methods tend to have good optimality properties when it comes to point estimation, and so box plots of point estimates are produced to compare performance of the **FID** point estimates (determined as the median of the weighted particles) with the ML approach.

To implement the ML approach, the statistical package **hglm** for R (R Development Core Team, 2011) was used. The function *hglm()* is used for fitting hierarchical generalized linear models, and, hence, can be used to fit generalized linear mixed models; more information on this package can be found in Rönnegard et al. (2010). Background on the methodology can be found in Lee et al. (2006), which also relies on the hierarchical likelihood, or h-likelihood approach of Lee and Nelder (1996). The h-likelihood is particular way to extend the likelihood of the data to include the distribution(s) of the random effect(s). This approach will be denoted as **HGLM**.

Design 1

Using the notation of the logistic regression with mixed effects model as displayed in (3.3), the first model design has $\beta_1 = 0$, $\beta_2 = 5$, $\sigma^2 = 2$, $x_{ij} = i/15$, $l_1 = 10$, $n_j = 15$, $n = 150$, $i = 1, \ldots, n_j$ and $j = 1, \ldots, l_1$. This design was found in McCulloch (1997).

Design 2

Design 2 first appeared in McGilchrist and Aisbett (1991), and had subsequent appearances in McGilchrist (1995) and Kuk (1995). The parameter values were set as $\beta_1 = 0.2$, $\beta_2 = 0.1$, and $\sigma = 1$, $l_1 = 15$, $n_j = 12$, $n = 180$, $i = 1, \ldots, n_j$ and $j = 1, \ldots, l_1$. The fixed effects design matrix has a column of ones, followed by $x_{i1} = 2i-16$ and $x_{i2} = 2i-15$ for $i = 1, \ldots, 15$. That is, for each level i of the random effect, there are six observations of x_{i1} and six observations of x_{i2}. In McGilchrist (1995), the simulation study compared the BLUP, ML, and REML

estimates with a focus on the variance component. It was found that all the methods produced negatively biased estimates of σ^2 with the BLUP estimator having the largest bias and the REML estimate having the smallest; however, BLUP had the lowest standard error and REML had the highest standard error.

Results

In this study, 100 independent data sets were generated. For Design 1, 5,000, 10,000, and 15,000 particles were generated for **FID**; Design 2 used 5,000, 8,000, 10,000, and 20,000 particles for **FID**. Based on the normal approximation to the binomial distribution, we will consider empirical coverage between 90.7% and 99.3% appropriate for 95% confidence intervals.

Table 3.4 contains the empirical coverage of 95% lower, upper, and two-sided confidence intervals under Design 1 for **FID** and **HGLM** along with their median interval lengths. One of the datasets for **HGLM** did not converge and was removed from the summary in Table 3.4 for **HGLM** only. The lower confidence intervals on β_1 are on the liberal end of acceptable coverage for **FID** and **HGLM**, but improves for **FID** as the number of particles increases. The lower intervals on β_2 are on the conservative end of acceptable coverage for both **FID** and **HGLM**, but not does change with increasing particle sample size. **FID** and **HGLM** have very conservative coverage for the lower confidence intervals on σ^2. The upper confidence intervals are acceptable for both methods and all parameters except **HGLM**'s coverage of σ^2 is mildly liberal. The two-sided intervals are within the stated coverage for **FID** and **HGLM**. **HGLM** has shorter median interval lengths than **FID**.

Figure 3.7 displays box plots of the estimated parameters using **FID** and **HGLM**. The **FID** estimates were determined by the medians of the weighted particle systems. Both **FID** and **HGLM** have median fixed effects estimates very close to the true parameter value (indicated by the horizontal dashed line). The estimates of the σ^2 are positively biased under **FID**. For **HGLM**, one of the 100 independent data sets did not converge, and so that data set was removed from the results displayed in Figure 3.7 for both methods.

Design	Method	N	Parameter	Lower	Upper	Two-sided	Median length
1	**FID**	5,000	β_1	0.91	0.99	0.96	2.41
			β_2	0.98	0.93	0.96	5.60
			σ^2	1.00	0.96	0.97	3.95
		10,000	β_1	0.92	0.99	0.97	2.46
			β_2	1.00	0.93	0.96	5.82
			σ^2	1.00	0.95	0.97	4.86
		15,000	β_1	0.94	0.99	0.98	2.49
			β_2	0.98	0.93	0.97	5.59
			σ^2	1.00	0.95	0.97	5.03
	HGLM		β_1	0.90	0.99	0.98	0.98
			β_2	0.97	0.97	0.97	5.09
			σ^2	1.00	0.89	0.93	1.99

Table 3.4: The performance of 95% confidence intervals for **FID** based on Design 1 described in Section 3.4.2. One hundred independent data sets were generated, and based on the normal approximation to the binomial distribution, empirical coverage between 90.7% and 99.3% appropriate for 95% confidence intervals.

Table 3.5 contains the empirical coverage of 95% lower, upper, and two-sided confidence intervals under Design 2 for **FID** and **HGLM** along with their median interval lengths. The empirical coverage for the intervals on β_1 are on the low end of correct coverage for **FID** and **HGLM** with slight improvement for **FID** as the particle sample size increases. The intervals on β_2 tends to be slightly liberal for **FID** except the lower confidence interval coverage tends to improve with increasing particle sample size while **HGLM** tends to have empirical coverage within the acceptable bounds. Interestingly, the lower intervals on σ^2 for **FID** are slightly liberal with improvement at 20,000 particles, and within the stated coverage for the upper and two-sided intervals (but on the low end of correct coverage for the two-sided intervals). **HGLM**, on the other hand, is rather liberal for the lower and two-sided intervals, and within the stated coverage for the upper intervals.

Figure 3.8 displays box plots of the estimated parameters using **FID** and **HGLM**. As before, the **FID** estimates were determined by the medians of the weighted particle systems.

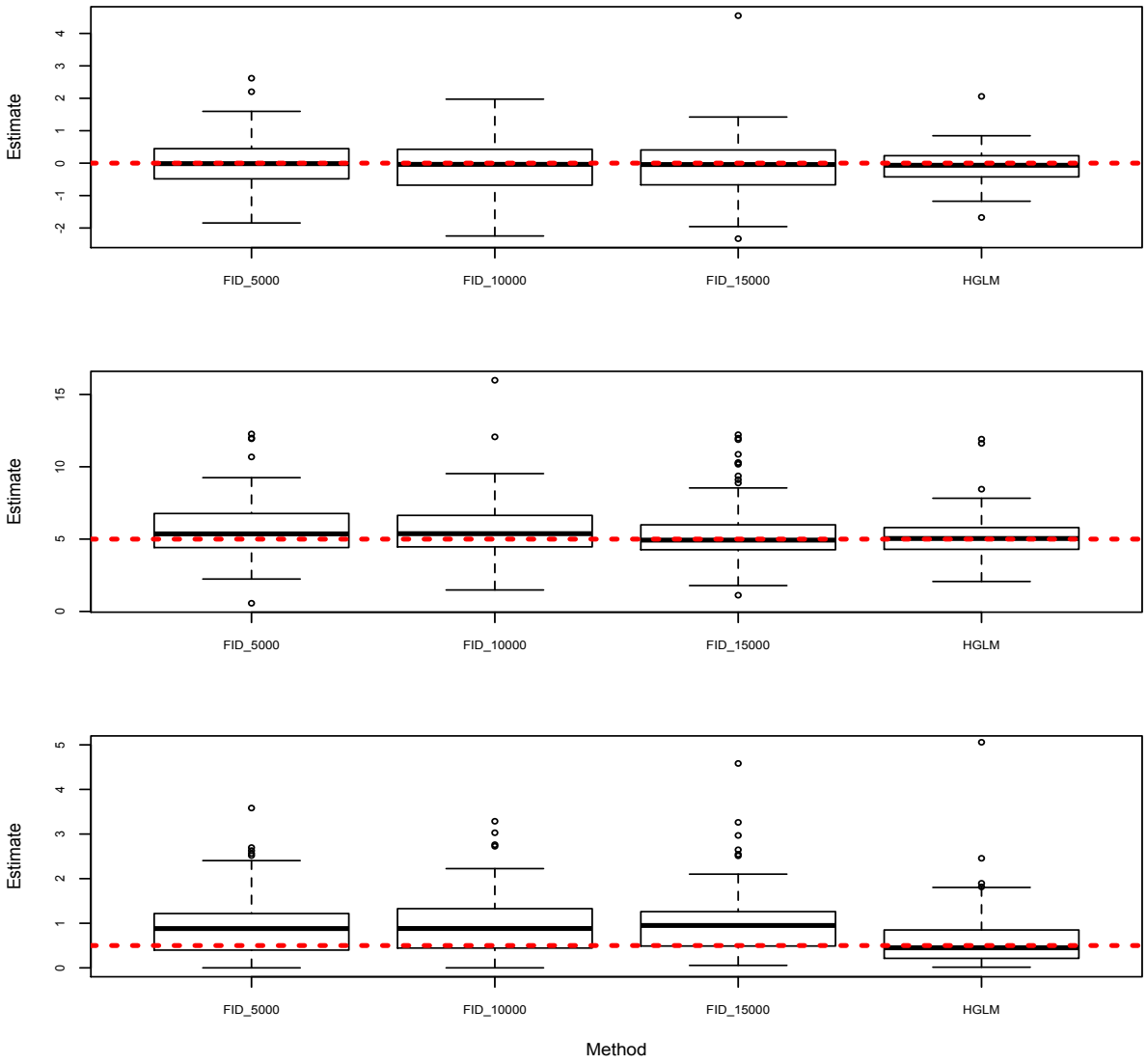


Figure 3.7: Simulation results for Design 1. The top box plot is of the estimates of β_1; the middle plot is of the estimates of β_2; the bottom plot is of the estimates of σ^2. The red dashed line indicates the location of the true parameter value used in the simulations study.

Both **FID** and **HGLM** have median fixed effects estimates very close to the true parameter value. The estimates of the σ^2 are slightly negatively biased under **FID** and **HGLM** except for **FID** with 20,000 particles; this negative bias was also observed in McGilchrist (1995).

Design	Method	N	Parameter	Lower	Upper	Two-sided	Median length
2	**FID**	5,000	β_1	0.92	0.93	0.90	1.28
			β_2	0.91	0.90	0.86	0.12
			σ^2	0.88	0.95	0.90	2.84
		8,000	β_1	0.92	0.95	0.92	1.29
			β_2	0.93	0.85	0.87	0.13
			σ^2	0.88	0.95	0.90	2.90
		10,000	β_1	0.92	0.93	0.92	1.32
			β_2	0.91	0.88	0.87	0.13
			σ^2	0.88	0.96	0.91	3.04
		20,000	β_1	0.95	0.93	0.93	1.33
			β_2	0.95	0.87	0.89	0.14
			σ^2	0.90	0.94	0.90	3.20
	HGLM		β_1	0.92	0.92	0.91	1.16
			β_2	0.94	0.97	0.95	0.13
			σ^2	0.83	0.97	0.86	1.89

Table 3.5: The performance of 95% confidence intervals for **FID** based on Design 2 described in Section 3.4.2. One hundred independent data sets were generated, and based on the normal approximation to the binomial distribution, empirical coverage between 90.7% and 99.3% appropriate for 95% confidence intervals.

3.5 Conclusion and future research

In this chapter, we introduced methodology for inference on the parameters of a logistic regression model with both fixed and random effects. The model is specified in Equation (3.3). The methodology is based on the ideas of generalized fiducial inference, and the implementation required the development of an SMC algorithm. Inference is based on the generalized fiducial distribution approximated by the generated particle system, and can be used, for example, for point estimation and confidence intervals.

Extensions to the proposed method in this chapter are similar to the extensions listed for the normal linear mixed model (e.g. spatial dependence, non-Gaussian random effects, etc.). There are a number of other topics that would be interesting for future research such as considering other link functions, or applying the methodology to more complicated real-data examples such as the famous salamander mating data set (Shun, 1997; Karim and Zeger,

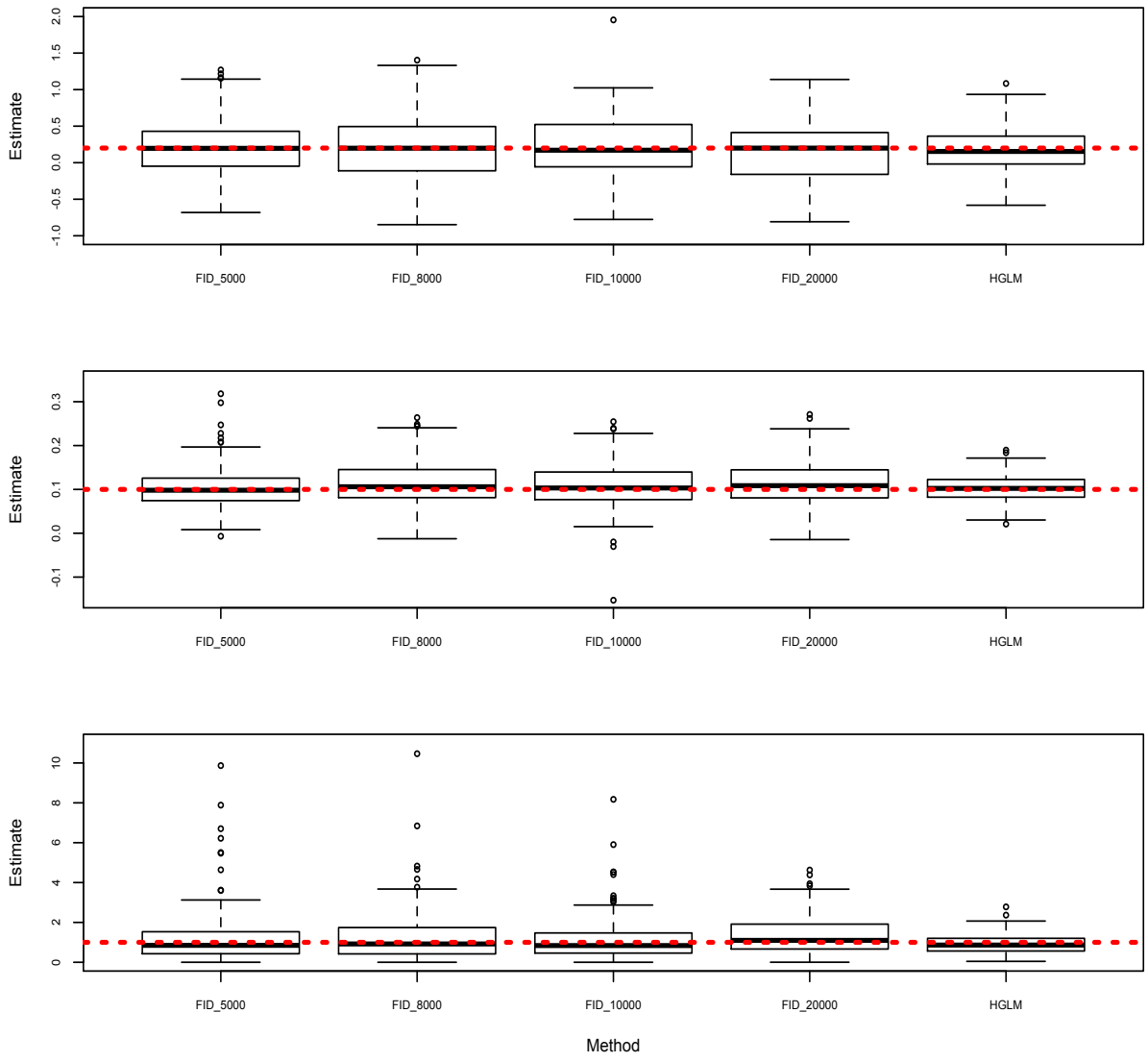


Figure 3.8: Simulation results for Design 2. The top box plot is of the estimates of β_1; the middle plot is of the estimates of β_2; the bottom plot is of the estimates of σ^2. The red dashed line indicates the location of the true parameter value used in the simulations study.

1992). Including other methods in the simulation study, such as a Bayesian method, would be interesting. Furthermore, we would like to develop a methodology for applying generalized fiducial inference to any generalized linear mixed model. The natural next step is to consider the log-linear model where the response is count data.

Chapter 4

Classification of suspect powders using spectral data

4.1 Summary

Classification of suspect powders, using Laser Induced Breakdown Spectroscopy (LIBS) spectra, to determine if they could contain *Bacillus anthracis* spores is difficult due to variability both in the composition of the spores and the LIBS analysis. We developed a method that builds a support vector machine classification model for such spectra relying on the known elemental composition of the *Bacillus* spores. A wavelet transformation was incorporated in this method to allow for possible thresholding or standardization, and then a linear model technique using the known elemental structure of the spores was incorporated for dimension reduction, and a support vector machine approach was employed for the final classification of the substance. The method was applied to real-data produced from a LIBS device. Several methods used to test the predictive performance of the classification model revealed promising results. This project was completed in collaboration with Emily Snyder and Lukas Oudejans from the US Environmental Protection Agency's Office of Research and Development.

4.2 Introduction

When a large building, complex or area has been contaminated with a powder substance that may contain *Bacillus* spores (causative agent for anthrax), it is crucial to determine if

the substance is potentially harmful quickly and efficiently. These powder substances could be non-hazardous hoaxes (e.g. dust, chalk, or sugar), or they could be or contain *Bacillus anthracis* (*B. anthracis*). Laser Induced Breakdown Spectroscopy (LIBS) devices have the capability of generating sample spectra that can aid in determining if a substance is or contains a spore material like *B. anthracis*. In LIBS, a laser is focused onto a sample producing a-plasma. This plasma atomizes, ionizes and subsequently excites the interrogated sample. The light emitted from the plasma is collected generating a characteristic wavelength spectrum. LIBS is an attractive technique for field analysis of suspect powders because it does not require preparation of samples, yields spectra in real time and is easily made man-portable.

In this chapter, a new statistical technique is presented for distinguishing *B. anthracis* surrogate spore powders from other innocuous suspect powders using LIBS spectra. The proposed method exploits a known property of the *B. anthracis* and its surrogate spores to aid in classification. Specifically, there are eight elements typically detected in *B. anthracis* and their surrogate spores (Gibb-Snyder et al., 2006). The principal idea of the proposed technique is, after preprocessing the data using wavelets, to determine the combination pattern of the LIBS spectra for those eight elements that form the LIBS spectra of *B. anthracis* spores and other innocuous powders using linear regression. The combination pattern is then used to build a classification model using a support vector machine approach to classify the substance as harmful (i.e. *B. anthracis* spores), or not.

A discussion of the LIBS systems and materials used in this project is included in Appendix B.

The idea after preprocessing the data was to reduce the dimension of the data using known structural information about *B. anthracis* spores and linear models, and build a support vector machine classification model using the dimension-reduced data. The statistical analysis begins with preprocessing of the LIBS spectra produced for the substances described in Section 4.3.1. The first step in preprocessing was to remove outlying spectra followed by a logarithmic transformation and a wavelet transformation. Wavelet transformations are described below along with an introduction to support vector machines, which were used for the classification model. A description of the proposed method follows these introductions, which is then

followed by an analysis and discussion of the performance of the model. The statistical analysis was completed using the statistical software R (R Development Core Team, 2011) with packages kernlab (Karatzoglou et al., 2004) and e1071 (Dimitriadou et al., 2011). A comparison of these packages can be found in Karatzoglou et al. (2006).

4.2.1 Wavelets

The wavelet transformation is a methodology useful in modeling data characterized by sharp peaks, or spikes, and other local features using a set of wavelet basis functions (Nason, 2008). There are several wavelet families, and among the most popular are the Daubechies wavelets, which form an orthonormal basis in the space of square-integrable functions. In modeling, the mother wavelet ψ is dilated and translated (i.e. stretched/squeezed and shifted) to represent some function f where, in this analysis, f is a LIBS spectrum. The general form of the representation of using a wavelet basis is

$$f(x) = c_{00}\phi_0(x) + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} d_{jk}\psi_{ji}(x) \tag{4.1}$$

where ϕ_0 is referred to as the father wavelet, or the scaling function, with coefficients c_{00}, and $\psi_{jk}(x) = 2^{\frac{j}{2}}\phi(2^j x - k)$ with integers j and k indexing the dilations and translations, respectively, of the mother wavelet. The dilation can be thought of as the window width of the wavelet. The wavelet coefficient (WC) at level j and location k, is defined as

$$d_{jk} = \int f(x)\psi_{jk}(x)dx. \tag{4.2}$$

In practice, f is observed at discretized points so Equation (4.2) is replaced by an approximation, and the range of the indices of the summations in Equation (4.1) are truncated based on values computed by the available data. Specifically, the sum indexed by j is truncated to $\log_2(n) - 1$, where n is the number of observations of the function f, and $k = 0, 1, \ldots, 2^j - 1$.

In the proposed methodology, a wavelet transformation was taken of each spectrum, and only the estimated WC, called the discrete wavelet coefficients (DWC), were retained for the classification model. Including a wavelet transformation in the framework provides flexibility

when more preprocessing of the data are required (e.g. for noisy data). More details about the wavelet transformation used in the proposed methodology can be found in Section 4.3.1.

4.2.2 Support vector machines

Support vector machines have many uses in statistics, in particular for classification. An overview of the methods can be found in Shawe-Taylor and Cristianini (2000) or Hastie et al. (2001). The main idea for Support Vector Machine (SVM) classification is to find the hyperplane that best separates the data into two classes by maximizing the margin between the closest points in each class.

Consider a data set $\{\mathbf{x}_i, \mathbf{y}_i\}$ where $i = 1, \ldots, n$, $\mathbf{y}_i = \{-1, 1\}$, and $\mathbf{x}_i \in \mathcal{R}^d$ where d is the dimension of the data set. For simplicity, suppose $d = 2$. Then there is a two-dimensional vector of n data points, and each point is assigned into class -1 or class 1. The goal is to find the hyperplane, which for $d = 2$ is the line, that best separates the two classes. Specifically, one needs to find variable b and vector $\mathbf{w}$ that defines the hyperplane

$$(\mathbf{x}_i \cdot \mathbf{w} + b)y_i \geq 1 \text{ for } i = 1, \ldots, n \tag{4.3}$$

where the hyperplane is such that it is as far as possible from the closest data points of each class. That is, the goal is to maximize the margin between the two classes. The distance between the hyperplane of Equation (4.3) and the support vectors of each class is equal to $||\mathbf{w}||^{-1}$, where the $\mathbf{w}$ is orthogonal to the hyperplane. Hence the distance between the two classes (i.e. the margin) is equal to $2||\mathbf{w}||^{-1}$. In order to maximize the margin, one can minimize $||\mathbf{w}||$ subject to the constraints defined in Equation (4.3). To make this computationally easier, $\frac{1}{2}||\mathbf{w}||^2$ is minimized with the same constraints. The variable b is the offset from the origin of the hyperplane.

Because most data will not be perfectly linearly separable, Equation (4.3) is modified to allow for misclassifications (data points on the wrong side of the separating margin) as follows:

$$(\mathbf{x}_i \cdot \mathbf{w} + b)y_i \geq 1 - \xi_i,\ \xi_i \geq 0 \text{ for } i = 1, \ldots, n. \tag{4.4}$$

This now allows for some values to be misclassified, and the objective function is modified to

$$\min\left(\frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i\right) \tag{4.5}$$

subject to the constraints of Equation (4.4). The parameter C is chosen to reflect the degree to which misclassifications are penalized and is referred to as the cost parameter, and the variables ξ_i are measures of the degree of the misclassification of $\mathbf{x}_i$. A non-zero ξ_i suggests that the observation $\mathbf{x}_i$ is on the "wrong" side of the hyperplane resulting in the application of a penalty.

The SVM classification model is completely defined by the vector $\mathbf{w}$ and b where a new point $\mathbf{x}^*$ is classified by the sign of $\mathbf{w}\cdot\mathbf{x}^* + b$. If there is asymmetry in the number of observations falling into each class, there are procedures for reducing the impact of the imbalance (Tang et al., 2004).

The above derivation assumes that the data are linearly separable, but this is not always the case (e.g. one of the classes could be circumscribed by the other class). However, the data can be transformed using a kernel function (determined by the data) onto feature space in order to improve linear separability (Shawe-Taylor and Cristianini, 2000).

For the proposed methodology, the $\mathbf{x}$ are a measure of the presence of the eight elements typically detected in *B. anthracis* spores (described in detail below), and the $\mathbf{y}$ indicate if the substance is a spore powder or not.

4.3 Methodology

The overall idea of the model is first to regularize the data via wavelets, then reduce the dimension of the samples using a linear model, and finally build a classification model that will categorize an unknown substance as a *B. anthracis* spore, or not, using support vector machines.

4.3.1 Data pre-processing

This section addresses the steps for pre-processing the data followed by the SVM classification model. The data used in the analysis were the spectra (intensity at different wavelengths) collected from the LIBS system and the substances described in Appendix B, and listed in Table 4.1. The spectra used to build the model had 13,701 data points (one for each wavelength at which the system recorded a measurement). A sample spectrum of *B. stearothermophilus* (ATCC 12979) and Dipel 150 Dust are displayed in Figure 4.1.

Substance	**Type**	**Number of observations**
B. stearothermophilus (ATCC 12979)	spore	21
B. thuringiensis (ATCC 51912)	spore	21
B. cereus (ATCC 14603)	spore	23
B. atrophaeus (ECBC)	spore	29
Blank steel plate	confusant	18
Flour	confusant	25
Arm & Hammer detergent	confusant	26
Baking powder	confusant	25
Baking soda	confusant	24
BC powder	confusant	25
Crayola chalk	confusant	25
Dipel	confusant	25
Equal	confusant	25
Gain laundry detergent (LD)	confusant	25
Ibuprofen	confusant	25
Johnson's baby powder (BP)	confusant	25
Powdered sugar	confusant	25
Sugar	confusant	25
Sweet 'n low	confusant	25
Tide LD	confusant	25
Tylenol gel cap	confusant	26

Table 4.1: The spores and confusant substances used in the analysis.

In addition to the LIBS spectra for *Bacillus* spore powders and other innocuous powders described in Appendix B, LIBS spectra for the eight elements typically detected in *B. anthracis* spores - sodium (Na), potassium (K), magnesium (Mg), manganese (Mn), silicon (Si), carbon (C), calcium (Ca), and iron (Fe) were considered. Each element has a characteristic spectrum with peaks at known wavelengths, and the characteristic elemental spectra were used in the analysis to aid in dimension reduction of the other LIBS spectra along with providing a way to

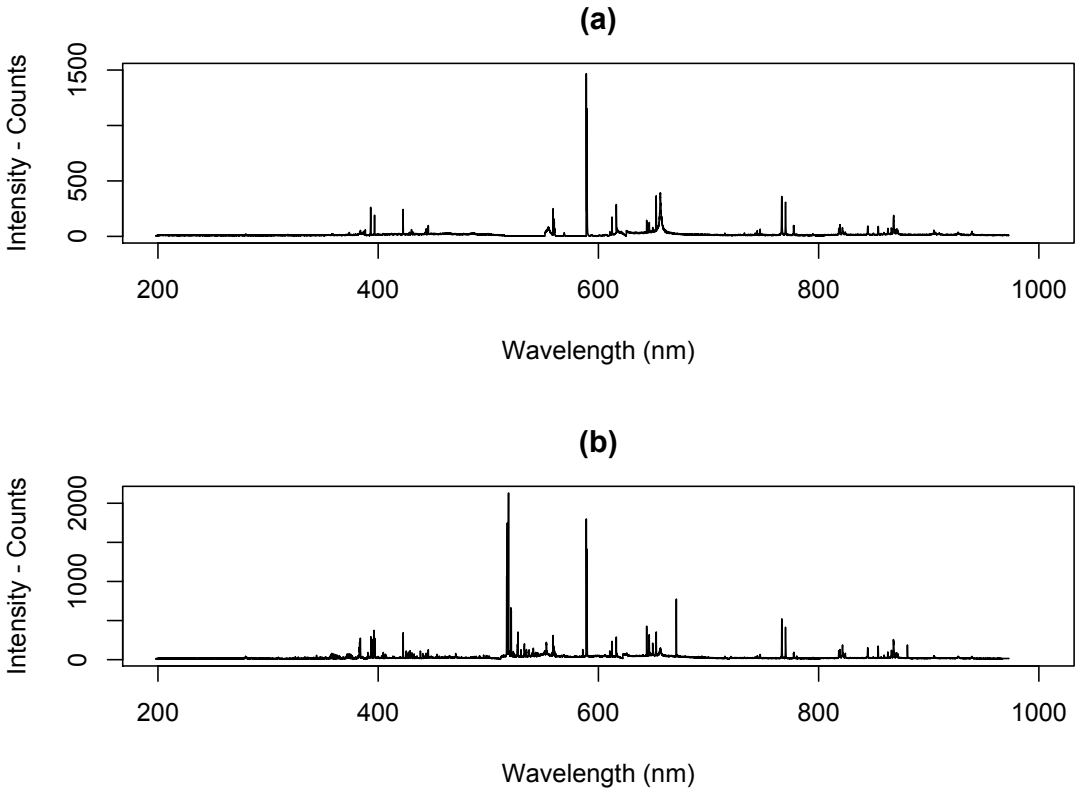


Figure 4.1: A spectrum generated by a LIBS system of (a) a sample of *B. stearothermophilus* (ATCC 12979), and (b) a sample of Dipel 150 Dust.

focus on the wavelengths of the spectra where peaks are expected when *B. anthracis* surrogate spores are present.

The first step in the analysis was to remove obvious outliers in the data. To locate suspected outliers, each set of spectra for each substance was analyzed in the following manner. For each substance (e.g. Sweet 'n Low spectra), a mean spectrum and median spectrum were defined. The sum of the absolute difference between each sample and the mean spectrum and between each sample and the medium spectrum were calculated along with noting the samples minimum and maximum intensities. Four plots were generated to compare these four values between the samples. If a samples point for any of the plots did not follow a pattern similar to the majority of the other points, its spectrum was visually compared to the mean and median spectra. In the few cases where the spectra were clear outliers (e.g. no peaks were appearing at any wavelength or for large portions of wavelength ranges), the associated samples were removed from the analysis (less than 0.6% of the sample spectra were removed as outliers).

After removing outlying spectra, a logarithmic transformation was taken of all the remaining data. There are a few points to note about the sample spectra displayed in Figure 4.1. First, the spectra have slightly irregular patterns tracing their bases (the irregularities are more pronounced after the logarithmic transformation); second, the spectra have a number of sharp peaks; lastly, the top spectrum in Figure 4.1 has peaks in locations and at heights different from the spectrum in the bottom of Figure 4.1. For these reasons, along with potential for thresholding and other forms of regularization, a wavelet transformation was employed rather than another functional representation of the data. While a wavelet transformation of the LIBS spectra was not required for use of the proposed classification model, the irregular base pattern was removed by only retaining the DWCs, i.e. the estimated d_{jk} from Equation (4.2), and discarding the information related to the scaling function, while preserving information about the peaks. The type of wavelet filter selected for the analysis was the Daubechies 4 filter, which has two vanishing moments. Symmetric boundary conditions were selected because it was reasonable to assume the unobserved signal to the right of the domain would be better represented as a continuation of the right part of the spectrum (rather than

the left part of the spectrum), and vice versa for the region to the left of the domain. Only 9 of the 13 possible levels, i.e. dilations of the mother wavelet, of coefficients were needed to capture the required detail of the spectra. The DWCs for each spectrum were vectorized and replaced the raw spectra as the data.

Next, a linear regression model was used to determine the combination pattern of the characteristic elemental spectra for the *Bacillus* spores' and the other substances' LIBS spectra. The components of the linear model are described below next.

For $e = 1, \ldots, 9$,

$$E^e = \text{ vector of DWCs for chemical element } e \tag{4.6}$$

where the ninth element included is boron due to its role in obtaining the elemental LIBS spectra. Including boron in the determination of the combination pattern, but not including it in the classification model, kept the presence of the boron in the elemental spectra from influencing classification. In addition to the E^e's, indicators for the seven CCDs (see Appendix B), defined by wavelength ranges, were carried through the wavelet transformation. The seven indicator vectors initially contained 1 for wavelengths within the range of the corresponding CCD and 0 everywhere else. The wavelength ranges are disjoint between the 7 CCDs and therefore for every wavelength there is only one column with the corresponding entry 1. These indictors were then subjected to the same wavelet transformation as the sample spectra, and vectorized as above, then defined for measuring device $i = 1, \ldots, 7$ as

$$I^i = \text{ vector of DWCs for measuring device indicator } i. \tag{4.7}$$

The I^is account for any overall inconsistencies between the ranges of each measuring device.

Every sample spectrum was transformed onto the wavelet domain using the same wavelet transformation with the DWCs retained. Each collection of DWCs was vectorized and defined as Y_{sj} = vector of DWCs, for substance $s = 1, \ldots, S$ and sample $j = 1, \ldots, J_s$ (each substance had 18 to 25 sample spectra).

The combination pattern of the elemental spectra for each substance is called its loadings. To get the loadings for every sample of each substance, the following linear model was fit:

$$Y_{sj} = \mathbf{X}\beta_{sj} + \epsilon. \tag{4.8}$$

$\mathbf{X}$ is a matrix composed of a column of ones, followed by $\mathbf{I}^i$, $i = 1, \ldots, 6$ defined in Equation (4.7) (only six of the seven CCDs were used because of the multicollinearity that would result, and thus non-identifiability of parameters; this is a mathematical issue and does not impact the results), and E_e, $e = 1, \ldots, 9$ as defined in Equation (4.6). Furthermore, β is the unknown 16 dimensional parameter vector for sample j of substance s, and ϵ is the unknown error in the model (note that normality of the data is not assumed). The loadings vector is defined as part of the least-squares estimates, $\widehat{\beta}_{sj}$, where $\widehat{\beta}_{sj} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y_{sj}$. The loading vector is denoted as $\widetilde{\beta}_{sj}$ for sample j of substance s, where this only includes the values from $\widehat{\beta}_{sj}$, corresponding to the eight elements (i.e. excluding boron).

In summary, each sample of each substance had an eight dimensional loading vector associated with it. The dimension of data was reduced from 13,701 to 8 by focusing on the locations of the spectra where the signature elements of *Bacillus* spores were found to have peaks. The loadings for the *Bacillus* spores (the particular combination of the elemental spectra for each sample) are displayed in Figures 4.2. The behavior of the loadings for each of the surrogates appears to be similar suggesting that these elemental spectra do combine in a similar way to form the *Bacillus* spores spectra, while the loadings for two examples of innocuous substances, baking soda and baking powder (see Figure 4.2), clearly follow a different pattern. These 8-dimensional loading vectors were used in the classification model described in the next section.

4.3.2 Classification model

Support vector machines were used to develop a classification model using the loadings associated with the data described in Appendix B. The two classes were *Bacillus* spores and other non-biological confusant substances. As noted in the Section 4.2.2, the goal of SVM

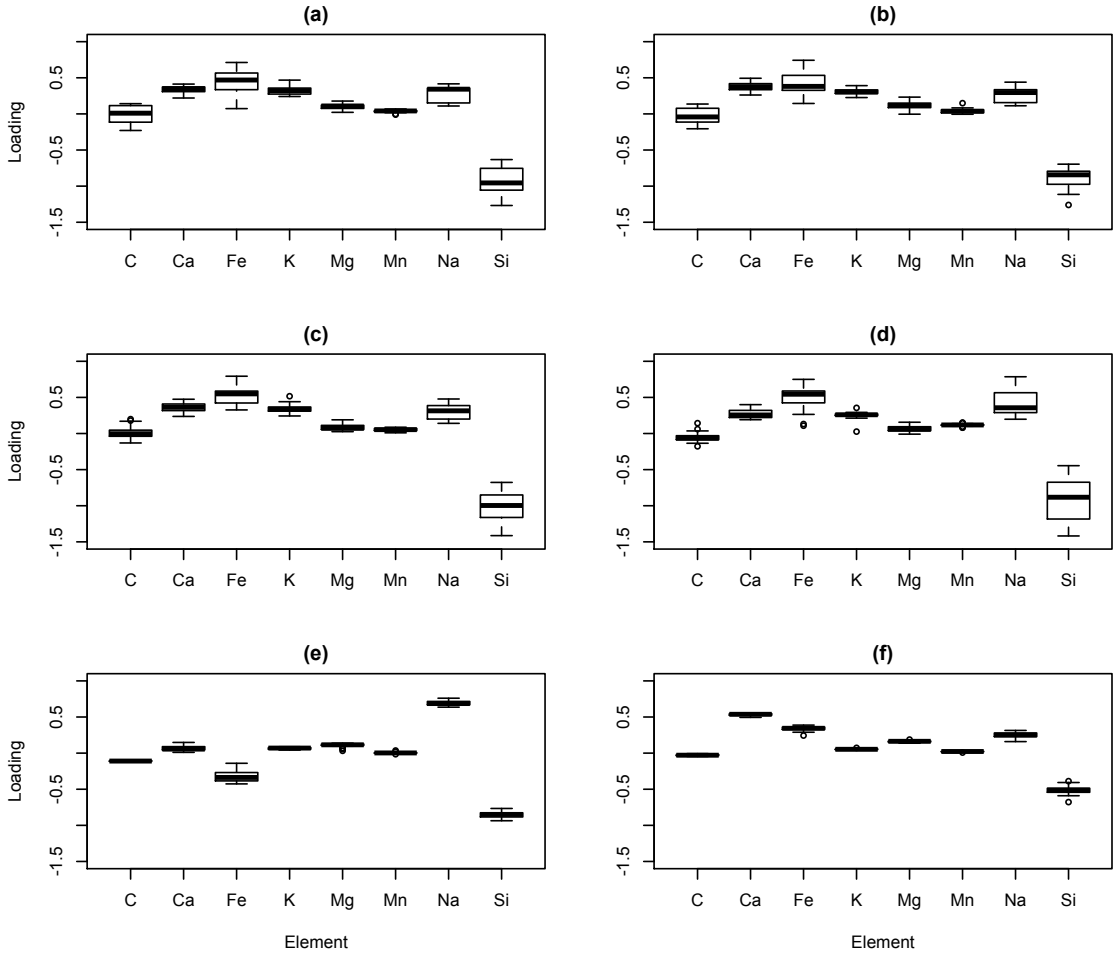


Figure 4.2: Box plots of the loadings for *Bacillus* spores (a) *B. stearothermophilus* (ATCC 12979), (b) *B. thuringiensis* (ATCC 51912), (c) *B. atrophaeus* (ECBC) and (d) *B. cereus* (ATCC 14603), and confusant substances (e) baking soda and (f) baking powder. The horizontal axis displays the eight elements typically detected in *Bacillus* spores, and the values along the vertical axis are the corresponding loading values for each sample summarized as a box plot. There were 21 - 29 samples used for each plot.

was to minimize Equation (4.5) subject to the constraints of Equation (4.4). In this analysis, $y_i = 1$ for spores, $y_i = -1$ for other substances and the corresponding vectors $\mathbf{x}$ were the 8-dimensional loading vectors defined above. In addition to our confusant powders, we included loadings for the spectra of the eight elements in the non-*Bacillus* group ($y_i = -1$).

Including these elemental loadings helped to guard against an unknown confusant substance being misclassified as *Bacillus* spores simply because it was made up of only one or several of these elements. For support vector machines, disproportionate class sizes, if not accounted for, can lead to incorrect classification (Shin and Cho, 2003; Wu and Chang, 2005). Because of the imbalance between the number of spores samples and other samples (94 sample spore spectra were used compared to 419 confusant substances), the substances were adjusted to alleviate the disproportion (to diminish the effect of asymmetric class sizes) using the *class.weights* option in the R packages mentioned previously. This option allows the user to mitigate the class size imbalance between the two classes by assigning a higher weight to the class with fewer samples and a lower weight to the samples with more samples.

The model was built using 70 percent of the data (randomly selected for each substance), and then verified using the remaining 30 percent. Because the data appeared to be linearly separable, the linear kernel was sufficient for this model, i.e. a hyperplane, or flat surface, separated the two classes of data and no additional transformation was needed. Figures 4.3 and 4.4 provide some validation of the claim to linear separability. In Figure 4.3, the median loading values for each substance are displayed. We see that the spore loadings tend to group together for each element, and are at an extreme for Si, K, Fe, and C. This has an aroma of linear separability in the eight-dimensional space that the loading vectors live in, which is further corroborated when considering the two-dimensional projections displayed in Figure 4.4.

During the construction of the classification model, the cost parameter C was set using a grid search over a specified range of values. The final C was chosen based on 10-fold cross validation error over the grid. The performance based on model error and 10-fold cross validation error was low for all C's presented in the grid. While all of the values considered for C performed well, the grid-search algorithm used for tuning parameters of SVM selected $C = 2$

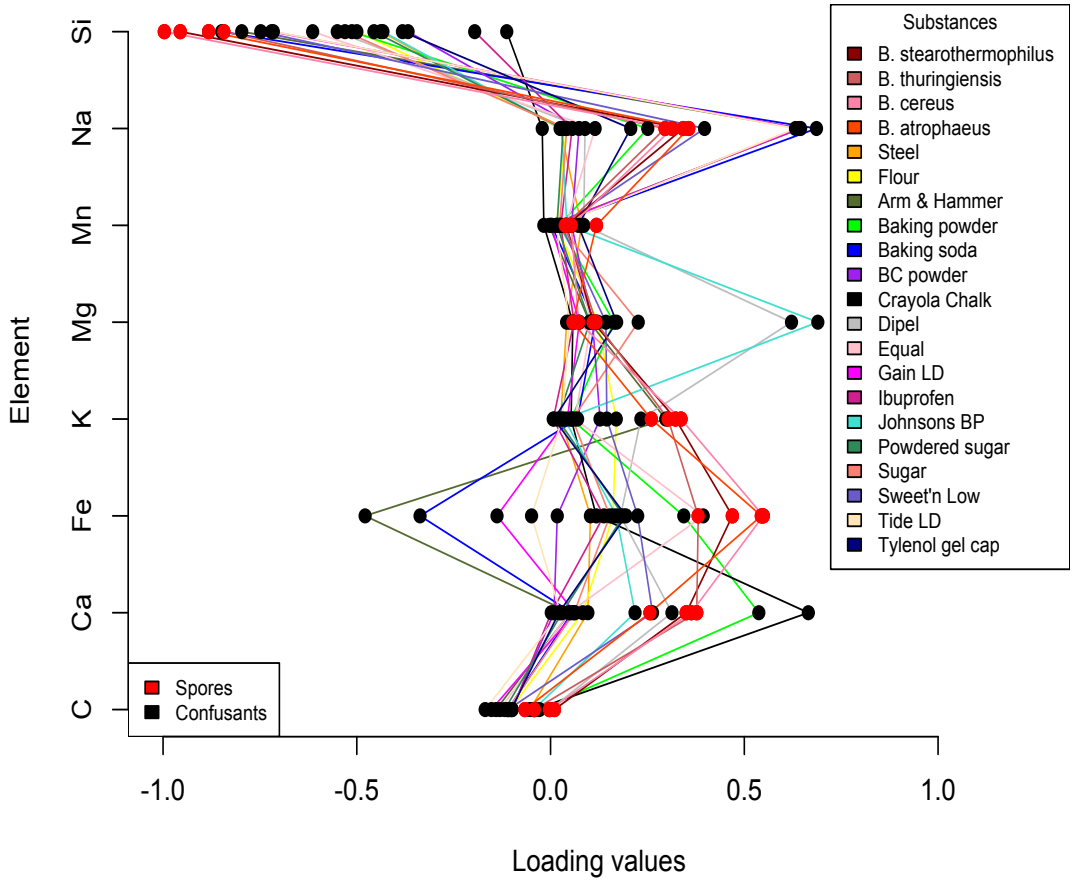


Figure 4.3: The red and black dots represent the median loading values for the spores and the confusant substances, respectively, displayed for the eight elements for each substance. The lines connecting the dots are colored to specify the substance. First, we note that the four spores tend to be grouped. Though we cannot visualize in eight dimensions, given that the spores are grouped at the extremes for some elements (Si, K, Fe, C) suggesting that the data are linearly separable and, hence, a transformation of the data is not necessary.

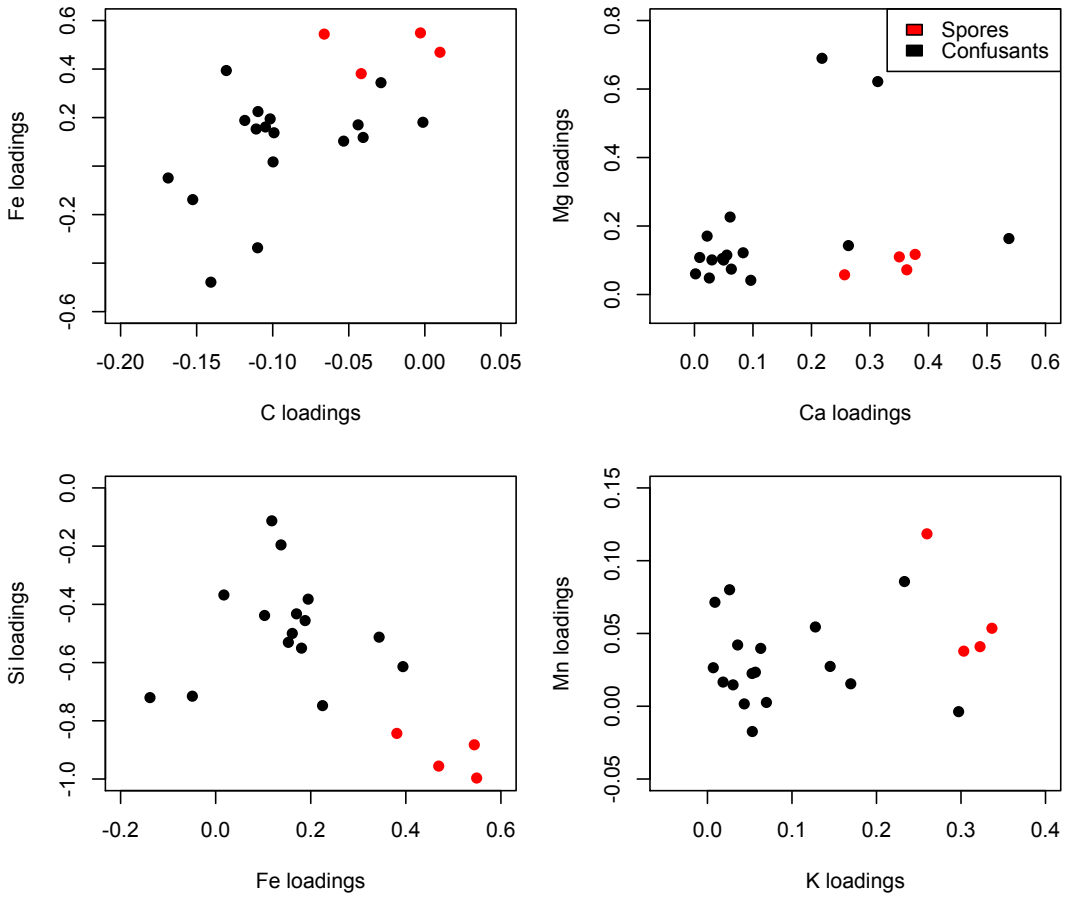


Figure 4.4: The red and black dots represent the median loading values for the spores and the confusant substances, respectively. Several two-dimensional projections are displayed, and were selected to illustrate the linear separability of the data. The particular substances are not specified because we are only concerned with the two classes (spore or not), but all the substances are represented in the plots.

as the best value based on 10-fold cross validation. Once the model was built, the remaining 30% of the data was put into the classification model. Misclassifications, or prediction error, occurred if the outcome of the model did not accurately predict the class of the substance, i.e. if a surrogate powder was classified as a confusant powder, or vice versa. Table 4.2 displays the results of the prediction error of the fitted model using the 30 percent of the data not included in the model selection and parameter tuning.

Confusant powders	***Bacillus* spore powders**
0.6%	3.3%

Table 4.2: Prediction error using cost parameter $C = 2$. The prediction error was determined by the frequency of misclassifications of the 30 percent of the data values not used in model selection.

In order to further test the model, sample spectra of two different *Bacillus* spores not used to build the model were classified using the model: *B. thuringiensis* (ATCC51912) (8 spectra) and *B. stearothermophilus* (ATCC12979) (5 spectra). All 13 spectra were correctly classified.

To further verify the performance, and due to the limited data available, a leave-one-out method was employed. The model was built using spectra from only three of the four *Bacillus* spores samples, (i.e. one *Bacillus* spores sample was left out of the model building stage). Seventy percent of the three remaining *Bacillus* spores samples and confusant powder samples were randomly selected to build the model. The models predictive power was tested on the *Bacillus* spores sample left out and the remaining 30% of the data. This was repeated for each *Bacillus* spore sample, and the results of this test are displayed in Table 4.3. The prediction error was between 0.0% and 3.4%.

A similar procedure was done by randomly selecting five non-*Bacillus* spore samples to be left out of building the model. The four confusant substances selected were baking soda, Gain Laundry Detergent (Gain LD), sugar, and Tide Laundry Detergent (Tide LD); the blank stainless steel spectra were also randomly selected. These five substances were grouped together in the determination of the prediction error. The results are also listed as the last row in Table 4.3. The prediction error was 0.9% and 3.3% for confusant powders and *Bacillus*

spores, respectively.

Testing substance left out	**Confusant powders**	***Bacillus* spore powders**
B. stearothermophilus (ATCC 12979)	0.0%	0.0%
B. thuringiensis (ATCC 51912)	1.3%	0.0%
B. cereus (ATCC 14603)	0.6%	0.0%
B. atrophaeus (ECBC)	0.0%	3.4%
Confusant	0.9%	3.3%

Table 4.3: Prediction error of model built with all samples except the samples listed under "Testing substance left out". *B. stearothermophilus* (ATCC 12979), *B. thuringiensis* (ATCC 51912), *B. cereus* (ATCC 14603), and *B. atrophaeus* (ECBC) are spores. The category "Confusant" includes stainless steel, baking soda, GainLD, sugar, and TideLD.

4.4 Conclusion and future research

The proposed methodology provides a way to classify suspect powders, like *Bacillus* spores, from other substances using LIBS spectra generated using the same LIBS system. Several statistical techniques were brought together to produce the classification model. A wavelet transformation was used to reduce irregularities in the LIBS spectra and focus the classification analysis on the peaks. Regressing the DWCs of the spores and other substances on the DWC of the eight elements helped to both reduce the dimension of the data, and focus on the regions of a spectrum where peaks were expected if spores are present in the substance. Finally, the output loadings vectors were then used to build the classification model using a SVM approach.

The overall classification model performed well for the data and setting presented. More complex classification goals would require an increased number of samples of *Bacillus* spores, and investigation of a non-linear relationship between spores spectra and the elemental spectra. In order to generalize this method to other LIBS systems, random effects could be incorporated to account for the variation differences between LIBS systems. For example, random effect could be included at the LIBS system level and at the operator level. The usefulness of incorporating random effects would be assessed by comparing the results of the method outlined in this chapter with the results of the same method proposed above except using a mixed linear model to determine the loading vectors rather than the linear model

used in Equation (4.8).

The analysis was completed using pure spores and pure confusants. An interesting and very relevant next step would be to see how the method performs when the spores are mixed with confusant substances at varying concentrations. In practice, this would be the more likely situation, and could have important national defense implications. More analysis could also be done on selecting which elements should be used in the classification model. We used all eight elements that are typically detected in *B. anthracis* surrogate spore powders, but perhaps fewer are necessary for good classification results. This would be an interesting issue to investigate using mixed, impure powders.

Finally, though the proposed methodology was implemented in the specific setting of *B. anthracis* surrogate spore classification, it could be used in other cases where one of the classes has some known elemental structure.

Appendix A

Normal linear mixed model results by model design

A.1 Simulation results: five-number summaries

In this section, the five-number summaries corresponding to the box plots in Chapter 2 are displayed.

A.1.1 Two-fold nested model of Equation (2.34)

σ^2_α	Method	Min	Q1	Q2	Q3	Max
		Two-sided CI Empirical coverage				
	USS	0.9385	0.9465	0.9510	0.9550	0.9700
	TYPEI	0.9245	0.9465	0.9515	0.9555	0.9650
	BAY1_3	0.9665	0.9780	0.9880	0.9975	1.0000
	BAY2_3	0.9660	0.9790	0.9895	0.9970	1.0000
	BAY$1_{1.5}$	0.9665	0.9765	0.9875	0.9965	1.0000
	BAY$2_{1.5}$	0.9695	0.9805	0.9885	0.9975	1.0000
	FID	0.9830	0.9880	0.9915	0.9985	1.0000
		Average length/ average length of **FID**				
	USS	0.7717	0.9415	0.9863	1.0138	1.1239
	TYPEI	0.6795	0.8678	0.9507	0.9863	1.0509
	BAY1_3	1.5641	1.7756	2.0990	2.6156	3.7342
	BAY2_3	1.6367	2.1150	2.3545	2.9533	6.7196
	BAY$1_{1.5}$	0.4829	0.6083	0.9450	1.6855	1.9474
	BAY$2_{1.5}$	1.0649	1.4059	1.5177	1.6983	2.4726
	FID	1.0000	1.0000	1.0000	1.0000	1.0000
		Upper CI Empirical coverage				
	USS	0.9420	0.9505	0.9535	0.9580	0.9870
	TYPEI	0.9290	0.9480	0.9545	0.9655	0.9845
	BAY1_3	0.9275	0.9420	0.9645	0.9850	1.0000
	BAY2_3	0.9295	0.9500	0.9670	0.9845	1.0000
	BAY$1_{1.5}$	0.9345	0.9505	0.9675	0.9835	0.9995
	BAY$2_{1.5}$	0.9370	0.9530	0.9690	0.9890	1.0000
	FID	0.9620	0.9745	0.9815	0.9950	1.0000
		Lower CI Empirical coverage				
	USS	0.9385	0.9455	0.9480	0.9515	0.9575
	TYPEI	0.9270	0.9385	0.9430	0.9455	0.9575
	BAY1_3	0.9980	1.0000	1.0000	1.0000	1.0000
	BAY2_3	0.9970	1.0000	1.0000	1.0000	1.0000
	BAY$1_{1.5}$	0.9960	1.0000	1.0000	1.0000	1.0000
	BAY$2_{1.5}$	0.9965	1.0000	1.0000	1.0000	1.0000
	FID	0.9625	0.9950	0.9995	1.0000	1.0000

Table A.1: Two-fold nested model: five-number summaries for 95% confidence intervals on σ^2_α of Figures 2.3 and 2.4.

σ^2_β	Method	Min	Q1	Q2	Q3	Max
		Two-sided CI Empirical coverage				
	USS	0.9385	0.9455	0.9475	0.9515	0.9535
	TYPEI	0.9235	0.9425	0.9480	0.9515	0.9580
	BAY1_3	0.9445	0.9800	0.9855	0.9880	0.9915
	BAY2_3	0.9405	0.9795	0.9845	0.9880	0.9930
	BAY$1_{1.5}$	0.9425	0.9770	0.9840	0.9870	0.9895
	BAY$2_{1.5}$	0.9400	0.9815	0.9855	0.9890	0.9935
	FID	0.9420	0.9800	0.9870	0.9910	0.9945
		Average length/ average length of **FID**				
	USS	0.9950	1.1581	1.5359	1.7280	131.1739
	TYPEI	0.8201	1.1314	1.5359	1.6826	131.1739
	BAY1_3	1.2282	1.3299	1.3955	1.9938	2.1843
	BAY2_3	1.2184	1.2982	1.3204	1.7752	2.0739
	BAY$1_{1.5}$	1.2258	1.2969	1.3183	1.4948	1.8172
	BAY$2_{1.5}$	1.1356	1.2008	1.2421	1.3638	1.5267
	FID	1.0000	1.0000	1.0000	1.0000	1.0000
		Upper CI Empirical coverage				
	USS	0.9335	0.9470	0.9515	0.9525	0.9570
	TYPEI	0.9315	0.9465	0.9510	0.9525	0.9560
	BAY1_3	0.9260	0.9600	0.9695	0.9735	0.9775
	BAY2_3	0.9285	0.9625	0.9675	0.9740	0.9785
	BAY$1_{1.5}$	0.9285	0.9595	0.9645	0.9730	0.9775
	BAY$2_{1.5}$	0.9290	0.9605	0.9715	0.9760	0.9790
	FID	0.9585	0.9705	0.9780	0.9820	0.9890
		Lower CI Empirical coverage				
	USS	0.9340	0.9445	0.9480	0.9515	0.9540
	TYPEI	0.9345	0.9445	0.9470	0.9505	0.9540
	BAY1_3	0.9595	0.9890	1.0000	1.0000	1.0000
	BAY2_3	0.9560	0.9885	1.0000	1.0000	1.0000
	BAY$1_{1.5}$	0.9570	0.9860	0.9995	1.0000	1.0000
	BAY$2_{1.5}$	0.9565	0.9855	1.0000	1.0000	1.0000
	FID	0.9160	0.9645	0.9955	1.0000	1.0000

Table A.2: Two-fold nested model: five-number summaries on 95% confidence intervals for σ^2_β of Figures 2.5 and 2.6.

μ	Method	Min	Q1	Q2	Q3	Max
		Two-sided CI Empirical coverage				
	BAY1_3	0.9770	0.9925	0.9995	1.0000	1.0000
	BAY2_3	0.9780	0.9915	0.9990	0.9995	1.0000
	BAY$1_{1.5}$	0.9770	0.9895	0.9930	0.9965	1.0000
	BAY$2_{1.5}$	0.9755	0.9885	0.9940	0.9965	0.9995
	FID	0.9650	0.9805	0.9920	0.9970	0.9990
		Average length/ average length of **FID**				
	BAY1_3	1.1498	1.3229	1.4824	1.5965	1.9567
	BAY2_3	1.1396	1.2751	1.5533	1.6491	2.4609
	BAY$1_{1.5}$	0.8569	1.0596	1.1698	1.2419	1.3922
	BAY$2_{1.5}$	0.8921	1.1367	1.1790	1.2157	1.5327
	FID	1.0000	1.0000	1.0000	1.0000	1.0000
		Upper CI Empirical coverage				
	BAY1_3	0.9745	0.9880	0.9970	0.9980	1.0000
	BAY2_3	0.9685	0.9865	0.9960	0.9980	1.0000
	BAY$1_{1.5}$	0.9705	0.9820	0.9905	0.9925	1.0000
	BAY$2_{1.5}$	0.9655	0.9855	0.9880	0.9925	0.9980
	FID	0.9580	0.9750	0.9795	0.9895	0.9975
		Lower CI Empirical coverage				
	BAY1_3	0.9720	0.9885	0.9965	0.9985	0.9995
	BAY2_3	0.9690	0.9875	0.9960	0.9980	0.9995
	BAY$1_{1.5}$	0.9690	0.9855	0.9900	0.9935	0.9995
	BAY$2_{1.5}$	0.9705	0.9830	0.9890	0.9940	0.9990
	FID	0.9555	0.9735	0.9815	0.9915	0.9980

Table A.3: Two-fold nested model: five-number summaries on 95% confidence intervals for μ of Figures 2.7 and 2.8.

σ^2_ϵ	Method	Min	Q1	Q2	Q3	Max
		Two-sided CI Empirical coverage				
	BAY1_3	0.9360	0.9460	0.9565	0.9595	0.9675
	BAY2_3	0.9355	0.9490	0.9555	0.9585	0.9690
	BAY$1_{1.5}$	0.9345	0.9480	0.9550	0.9585	0.9655
	BAY$2_{1.5}$	0.9330	0.9470	0.9565	0.9605	0.9675
	FID	0.8955	0.9505	0.9530	0.9575	0.9615
		Average length/ average length of **FID**				
	BAY1_3	0.8886	1.0697	1.0862	1.0999	1.2970
	BAY2_3	0.8438	1.0588	1.0822	1.0899	1.2959
	BAY$1_{1.5}$	0.8867	1.0686	1.0851	1.0981	1.3046
	BAY$2_{1.5}$	0.8909	1.0602	1.0831	1.0938	1.3109
	FID	1.0000	1.0000	1.0000	1.0000	1.0000
		Upper CI Empirical coverage				
	BAY1_3	0.9090	0.9225	0.9370	0.9520	0.9665
	BAY2_3	0.9085	0.9235	0.9375	0.9525	0.9665
	BAY$1_{1.5}$	0.9105	0.9250	0.9370	0.9510	0.9635
	BAY$2_{1.5}$	0.9095	0.9230	0.9365	0.9525	0.9665
	FID	0.8605	0.9415	0.9565	0.9655	0.9760
		Lower CI Empirical coverage				
	BAY1_3	0.9440	0.9640	0.9705	0.9735	0.9820
	BAY2_3	0.9455	0.9630	0.9705	0.9740	0.9820
	BAY$1_{1.5}$	0.9475	0.9640	0.9690	0.9745	0.9820
	BAY$2_{1.5}$	0.9470	0.9635	0.9695	0.9735	0.9830
	FID	0.9295	0.9450	0.9515	0.9660	0.9730

Table A.4: Two-fold nested model: five-number summaries on 95% confidence intervals for σ^2_ϵ of Figures 2.9 and 2.9.

A.1.2 The two-factor crossed with interaction of Equation (2.35)

σ^2_α	Method	Min	Q1	Q2	Q3	Max
		Two-sided CI Empirical coverage				
	HB	0.9405	0.9490	0.9523	0.9550	0.9645
	BAY1_3	0.9685	0.9785	0.9850	0.9900	0.9930
	BAY2_3	0.9675	0.9790	0.9855	0.9915	0.9940
	BAY$1_{1.5}$	0.9680	0.9775	0.9848	0.9885	0.9930
	BAY$2_{1.5}$	0.9710	0.9810	0.9853	0.9910	0.9960
	FID	0.9815	0.9890	0.9943	0.9965	0.9990
		Average length/ average length of **FID**				
	HB	0.2714	0.9030	0.9865	1.0237	1.0365
	BAY1_3	0.6182	3.0159	3.6835	3.9085	4.4357
	BAY2_3	0.6059	3.2548	3.6505	4.6360	5.7224
	BAY$1_{1.5}$	0.5877	1.3033	1.6039	2.2952	2.7000
	BAY$2_{1.5}$	0.5086	2.0343	2.2739	2.4848	2.7559
	FID	1.0000	1.0000	1.0000	1.0000	1.0000
		Upper CI Empirical coverage				
	HB	0.9415	0.9510	0.9530	0.9565	0.9600
	BAY1_3	0.9330	0.9480	0.9575	0.9670	0.9785
	BAY2_3	0.9350	0.9515	0.9593	0.9685	0.9790
	BAY$1_{1.5}$	0.9355	0.9485	0.9588	0.9665	0.9800
	BAY$2_{1.5}$	0.9405	0.9530	0.9628	0.9700	0.9795
	FID	0.9610	0.9750	0.9838	0.9875	0.9940
		Lower CI Empirical coverage				
	HB	0.9360	0.9445	0.9495	0.9515	0.9605
	BAY1_3	1.0000	1.0000	1.0000	1.0000	1.0000
	BAY2_3	0.9995	1.0000	1.0000	1.0000	1.0000
	BAY$1_{1.5}$	0.9995	1.0000	1.0000	1.0000	1.0000
	BAY$2_{1.5}$	0.9995	1.0000	1.0000	1.0000	1.0000
	FID	0.9790	1.0000	1.0000	1.0000	1.0000

Table A.5: Two-way crossed with interaction model: five-number summaries on 95% confidence intervals for σ^2_α of Figures 2.12 and 2.13.

σ^2_β	Method	Min	Q1	Q2	Q3	Max
		Two-sided CI Empirical coverage				
	HB	0.9450	0.9490	0.9520	0.9535	0.9635
	BAY1_3	0.9220	0.9605	0.9743	0.9835	0.9925
	BAY2_3	0.9280	0.9630	0.9758	0.9850	0.9940
	BAY$1_{1.5}$	0.9270	0.9625	0.9773	0.9830	0.9905
	BAY$2_{1.5}$	0.9325	0.9645	0.9783	0.9865	0.9925
	FID	0.9710	0.9850	0.9898	0.9960	0.9985
		Average length/ average length of **FID**				
	HB	0.8897	0.9758	0.9994	1.0215	1.0414
	BAY1_3	2.0614	2.8717	3.4176	4.0489	6.0724
	BAY2_3	2.5869	3.2937	4.3123	4.8400	6.6194
	BAY$1_{1.5}$	0.6731	0.9853	1.3007	1.7788	2.4014
	BAY$2_{1.5}$	1.5087	1.9015	2.1036	2.4251	3.5400
	FID	1.0000	1.0000	1.0000	1.0000	1.0000
		Upper CI Empirical coverage				
	HB	0.9390	0.9510	0.9538	0.9575	0.9635
	BAY1_3	0.8650	0.9230	0.9353	0.9520	0.9710
	BAY2_3	0.8725	0.9270	0.9388	0.9525	0.9715
	BAY$1_{1.5}$	0.8715	0.9275	0.9385	0.9535	0.9725
	BAY$2_{1.5}$	0.8760	0.9320	0.9460	0.9650	0.9730
	FID	0.9480	0.9650	0.9758	0.9840	0.9950
		Lower CI Empirical coverage				
	HB	0.9350	0.9445	0.9465	0.9535	0.9580
	BAY1_3	1.0000	1.0000	1.0000	1.0000	1.0000
	BAY2_3	0.9995	1.0000	1.0000	1.0000	1.0000
	BAY$1_{1.5}$	1.0000	1.0000	1.0000	1.0000	1.0000
	BAY$2_{1.5}$	0.9990	1.0000	1.0000	1.0000	1.0000
	FID	0.9615	0.9960	1.0000	1.0000	1.0000

Table A.6: Two-way crossed with interaction model: five-number summaries on 95% confidence intervals for σ^2_β of Figures 2.14 and 2.15.

$\sigma^2_{\alpha\beta}$	Method	Min	Q1	Q2	Q3	Max
		Two-sided CI Empirical coverage				
	HB	0.9435	0.9485	0.9508	0.9535	0.9620
	BAY1_3	0.9600	0.9840	0.9878	0.9900	0.9955
	BAY2_3	0.9595	0.9850	0.9888	0.9905	0.9955
	BAY$1_{1.5}$	0.9570	0.9810	0.9863	0.9890	0.9925
	BAY$2_{1.5}$	0.9600	0.9845	0.9883	0.9900	0.9960
	FID	0.9445	0.9855	0.9910	0.9940	0.9975
		Average length/ average length of **FID**				
	HB	1.1395	1.3342	1.4998	1.6228	1.9391
	BAY1_3	1.1172	1.1654	1.2204	1.3624	1.4333
	BAY2_3	1.1026	1.1472	1.2025	1.2917	1.3745
	BAY$1_{1.5}$	1.1105	1.1479	1.2072	1.3118	1.3849
	BAY$2_{1.5}$	0.9536	1.1248	1.1629	1.1839	1.2811
	FID	1.0000	1.0000	1.0000	1.0000	1.0000
		Upper CI Empirical coverage				
	HB	0.9440	0.9470	0.9513	0.9560	0.9620
	BAY1_3	0.9565	0.9705	0.9738	0.9800	0.9870
	BAY2_3	0.9595	0.9725	0.9743	0.9800	0.9880
	BAY$1_{1.5}$	0.9585	0.9680	0.9725	0.9780	0.9860
	BAY$2_{1.5}$	0.9570	0.9720	0.9750	0.9795	0.9895
	FID	0.9595	0.9780	0.9823	0.9875	0.9945
		Lower CI Empirical coverage				
	HB	0.9395	0.9455	0.9490	0.9530	0.9570
	BAY1_3	0.9490	0.9905	0.9990	1.0000	1.0000
	BAY2_3	0.9485	0.9900	0.9988	1.0000	1.0000
	BAY$1_{1.5}$	0.9480	0.9860	0.9975	1.0000	1.0000
	BAY$2_{1.5}$	0.9450	0.9865	0.9985	1.0000	1.0000
	FID	0.9235	0.9700	0.9950	1.0000	1.0000

Table A.7: Two-way crossed with interaction model: five-number summaries on 95% confidence intervals for $\sigma^2_{\alpha\beta}$ of Figures 2.16 and 2.17.

μ	Method	Min	Q1	Q2	Q3	Max
		Two-sided CI Empirical coverage				
	BAY1_3	0.9975	0.9995	1.0000	1.0000	1.0000
	BAY2_3	0.9970	0.9995	1.0000	1.0000	1.0000
	BAY$1_{1.5}$	0.9895	0.9975	0.9988	1.0000	1.0000
	BAY$2_{1.5}$	0.9890	0.9975	0.9988	0.9995	1.0000
	FID	0.9805	0.9945	0.9975	0.9995	1.0000
		Average length/ average length of **FID**				
	BAY1_3	1.7155	1.7631	1.8151	1.9477	2.1148
	BAY2_3	1.7002	1.7864	1.8298	1.9473	2.0953
	BAY$1_{1.5}$	1.1219	1.2070	1.2793	1.3721	1.5267
	BAY$2_{1.5}$	1.1959	1.2828	1.3330	1.4153	1.5603
	FID	1.0000	1.0000	1.0000	1.0000	1.0000
		Upper CI Empirical coverage				
	BAY1_3	0.9940	0.9990	0.9995	1.0000	1.0000
	BAY2_3	0.9900	0.9985	0.9995	1.0000	1.0000
	BAY$1_{1.5}$	0.9840	0.9950	0.9980	0.9995	1.0000
	BAY$2_{1.5}$	0.9805	0.9950	0.9970	0.9990	1.0000
	FID	0.9695	0.9895	0.9935	0.9975	0.9995
		Lower CI Empirical coverage				
	BAY1_3	0.9935	0.9985	0.9995	1.0000	1.0000
	BAY2_3	0.9925	0.9985	0.9995	1.0000	1.0000
	BAY$1_{1.5}$	0.9855	0.9955	0.9965	0.9990	1.0000
	BAY$2_{1.5}$	0.9845	0.9940	0.9975	0.9985	1.0000
	FID	0.9720	0.9880	0.9933	0.9975	1.0000

Table A.8: Two-way crossed with interaction model: five-number summaries on 95% confidence intervals for μ of Figures 2.18 and 2.19.

σ^2_ϵ	Method	Min	Q1	Q2	Q3	Max
		Two-sided CI Empirical coverage				
	BAY1_3	0.9450	0.9510	0.9553	0.9575	0.9640
	BAY2_3	0.9445	0.9510	0.9553	0.9590	0.9625
	BAY$1_{1.5}$	0.9460	0.9510	0.9555	0.9565	0.9620
	BAY$2_{1.5}$	0.9455	0.9515	0.9553	0.9570	0.9620
	FID	0.9415	0.9515	0.9565	0.9615	0.9645
		Average length/ average length of **FID**				
	BAY1_3	1.0466	1.0537	1.0750	1.0808	1.1096
	BAY2_3	1.0441	1.0536	1.0746	1.0810	1.1079
	BAY$1_{1.5}$	1.0428	1.0493	1.0722	1.0783	1.1084
	BAY$2_{1.5}$	1.0002	1.0499	1.0726	1.0788	1.1047
	FID	1.0000	1.0000	1.0000	1.0000	1.0000
		Upper CI Empirical coverage				
	BAY1_3	0.9170	0.9340	0.9390	0.9470	0.9665
	BAY2_3	0.9160	0.9325	0.9398	0.9470	0.9670
	BAY$1_{1.5}$	0.9185	0.9325	0.9395	0.9480	0.9655
	BAY$2_{1.5}$	0.9170	0.9320	0.9398	0.9490	0.9650
	FID	0.9330	0.9480	0.9568	0.9635	0.9730
		Lower CI Empirical coverage				
	BAY1_3	0.9510	0.9605	0.9655	0.9745	0.9810
	BAY2_3	0.9505	0.9600	0.9660	0.9745	0.9815
	BAY$1_{1.5}$	0.9505	0.9610	0.9660	0.9740	0.9800
	BAY$2_{1.5}$	0.9505	0.9600	0.9655	0.9750	0.9810
	FID	0.9315	0.9460	0.9535	0.9670	0.9715

Table A.9: Two-way crossed with interaction model: five-number summaries on 95% confidence intervals for σ^2_ϵ of Figures 2.20 and 2.21.

A.2 Simulation results: summarized by model design

In this section, the simulation results from Chapter 2 are summarized by model design. The results include empirical coverage of the 95% two-sided confidence intervals and average interval length for the non-error variance components.

A.2.1 Two-fold nested model of Equation (2.34)

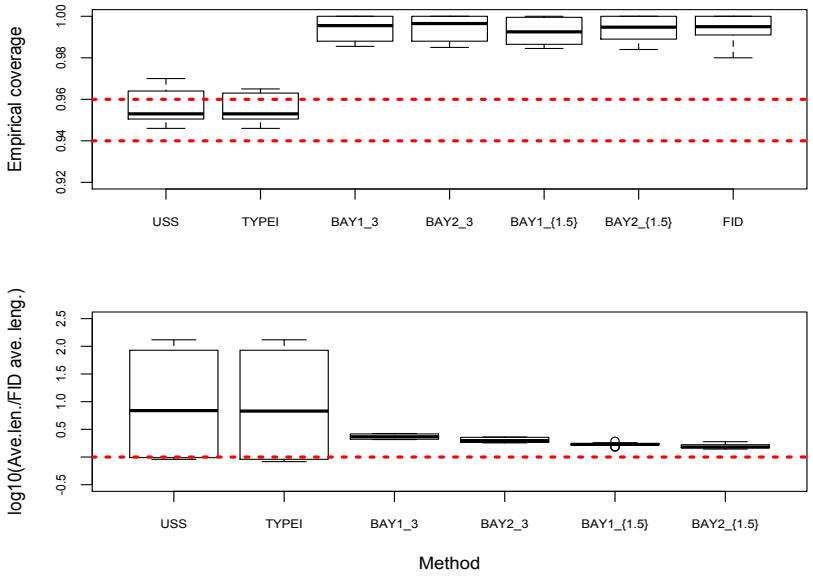


Figure A.1: Combined simulation results for 95% two-sided confidence intervals on σ^2_α and σ^2_β for the two-fold nested model of Equation (2.34) using model design MI - 1 of Table 2.3. The top plot is of the empirical coverage probabilities of the intervals, and the bottom plot is of the base 10 logarithm of average interval lengths divided by the average interval lengths of **FID**.

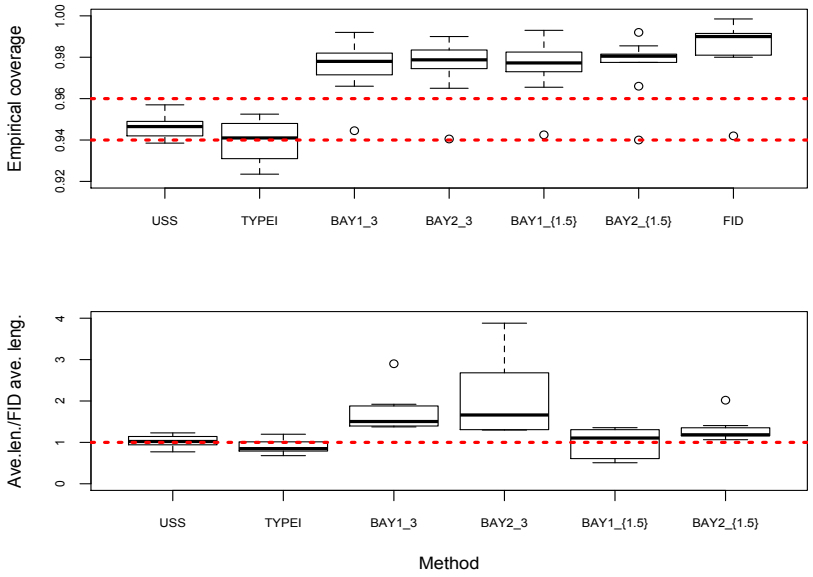


Figure A.2: Combined simulation results for 95% two-sided confidence intervals on σ_α^2 and σ_β^2 for the two-fold nested model of Equation (2.34) using model design MI - 2 of Table 2.3. The top plot is of the empirical coverage probabilities of the intervals, and the bottom plot is of the average interval lengths divided by the average interval lengths of **FID**.

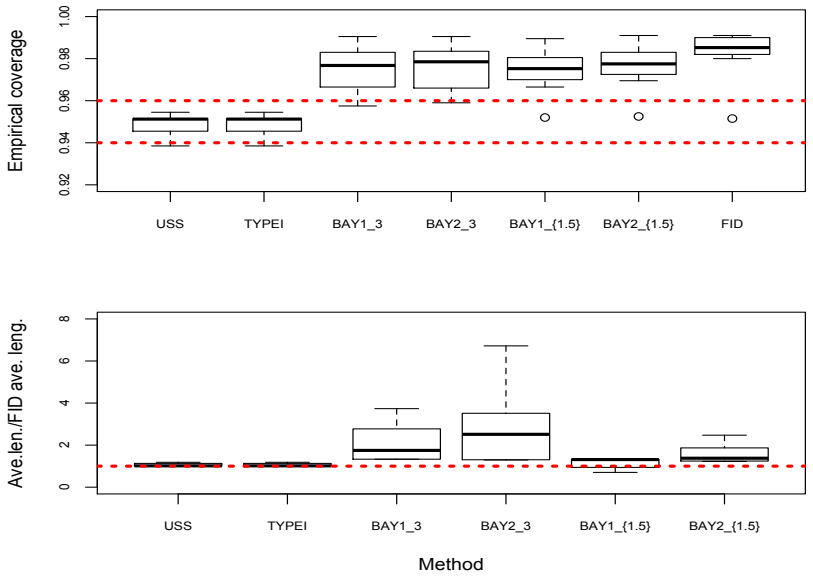


Figure A.3: Combined simulation results for 95% two-sided confidence intervals on σ_α^2 and σ_β^2 for the two-fold nested model of Equation (2.34) using model design MI - 3 of Table 2.3. The top plot is of the empirical coverage probabilities of the intervals, and the bottom plot is of the average interval lengths divided by the average interval lengths of **FID**.

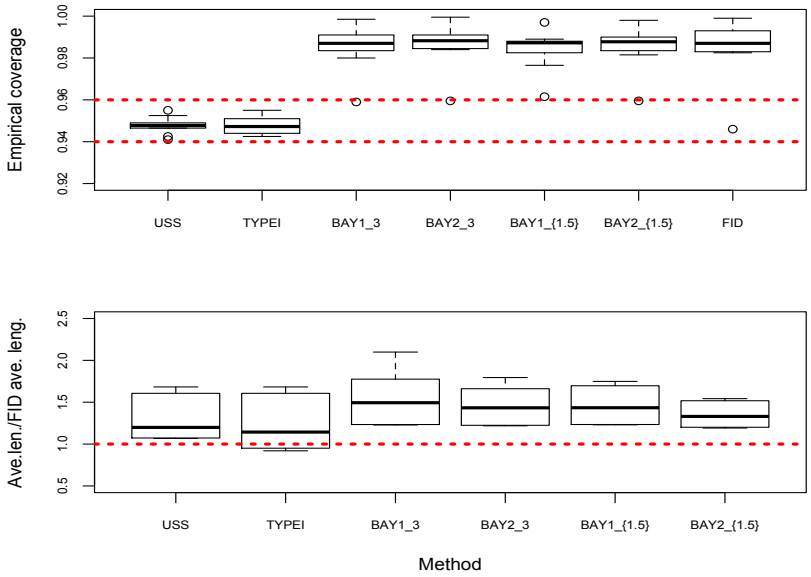


Figure A.4: Combined simulation results for 95% two-sided confidence intervals on σ_α^2 and σ_β^2 for the two-fold nested model of Equation (2.34) using model design MI - 4 of Table 2.3. The top plot is of the empirical coverage probabilities of the intervals, and the bottom plot is of the average interval lengths divided by the average interval lengths of **FID**.

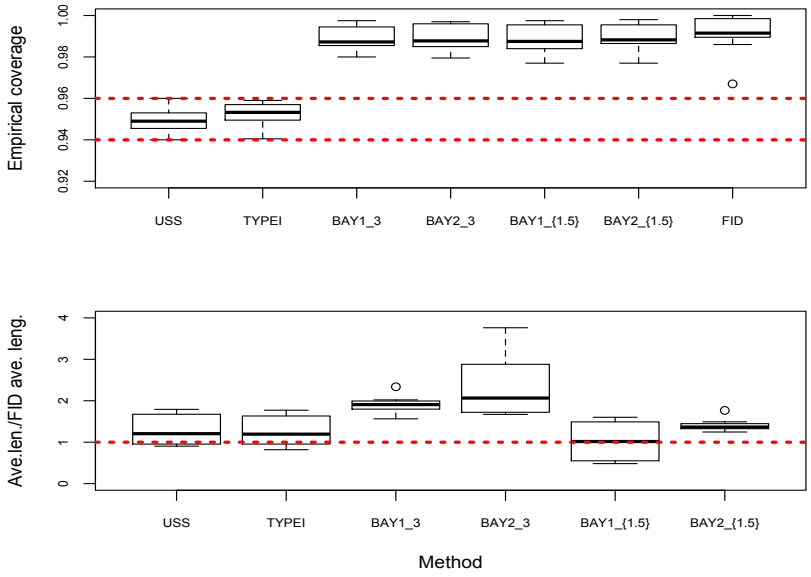


Figure A.5: Combined simulation results for 95% two-sided confidence intervals on σ_α^2 and σ_β^2 for the two-fold nested model of Equation (2.34) using model design MI - 5 of Table 2.3. The top plot is of the empirical coverage probabilities of the intervals, and the bottom plot is of the average interval lengths divided by the average interval lengths of **FID**.

A.2.2 Two-factor crossed with interaction of Equation (2.35)

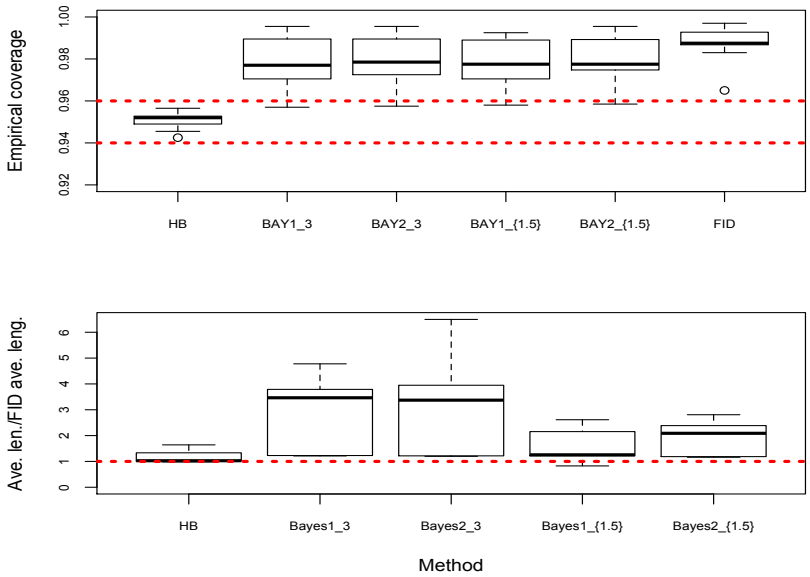


Figure A.6: Combined simulation results for 95% two-sided confidence intervals on σ^2_α, σ^2_β, and $\sigma^2_{\alpha\beta}$ for the two-factor crossed design with interactionl of Equation (2.35) using model design MII - 1 of Table 2.6. The top plot is of the empirical coverage probabilities of the intervals, and the bottom plot is of the average interval lengths divided by the average interval lengths of **FID**.

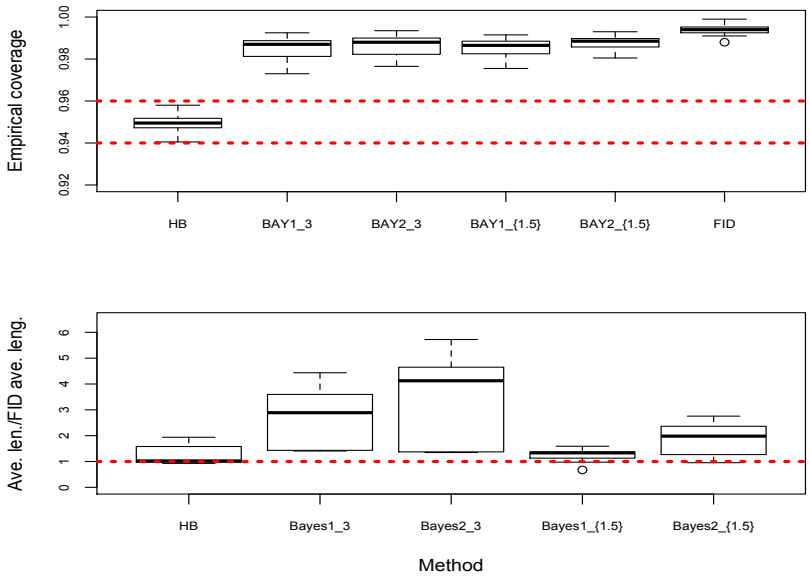


Figure A.7: Combined simulation results for 95% two-sided confidence intervals on σ^2_α, σ^2_β, and $\sigma^2_{\alpha\beta}$ for the two-factor crossed design with interactionl of Equation (2.35) using model design MII - 2 of Table 2.6. The top plot is of the empirical coverage probabilities of the intervals, and the bottom plot is of the average interval lengths divided by the average interval lengths of **FID**.

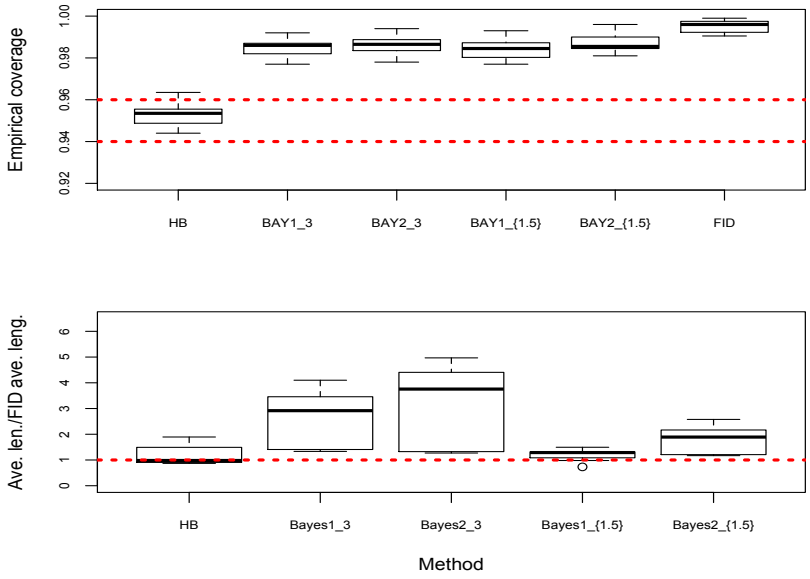


Figure A.8: Combined simulation results for 95% two-sided confidence intervals on σ^2_α, σ^2_β, and $\sigma^2_{\alpha\beta}$ for the two-factor crossed design with interactionl of Equation (2.35) using model design MII - 3 of Table 2.6. The top plot is of the empirical coverage probabilities of the intervals, and the bottom plot is of the average interval lengths divided by the average interval lengths of **FID**.

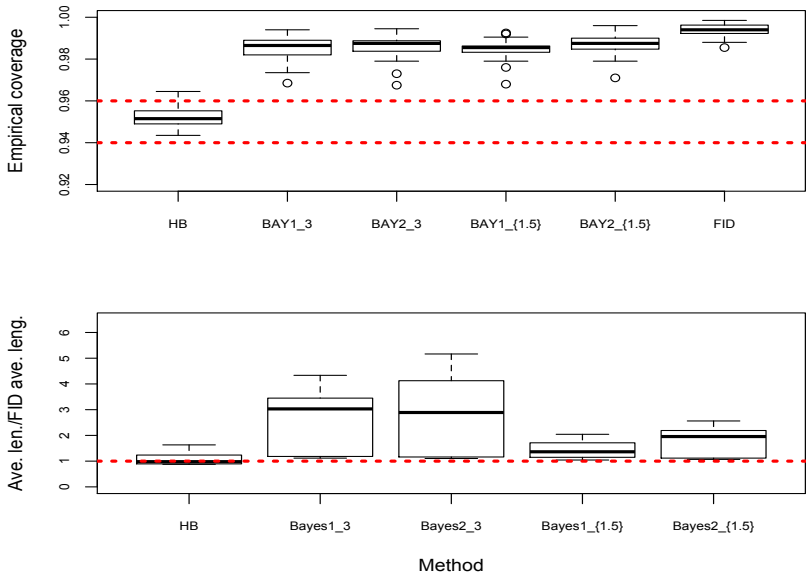


Figure A.9: Combined simulation results for 95% two-sided confidence intervals on σ^2_α, σ^2_β, and $\sigma^2_{\alpha\beta}$ for the two-factor crossed design with interactionl of Equation (2.35) using model design MII - 4 of Table 2.6. The top plot is of the empirical coverage probabilities of the intervals, and the bottom plot is of the average interval lengths divided by the average interval lengths of **FID**.

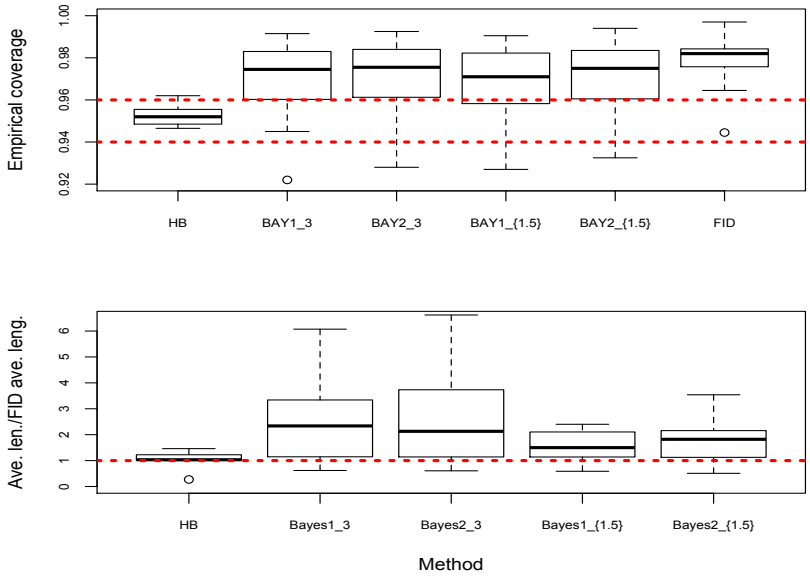


Figure A.10: Combined simulation results for 95% two-sided confidence intervals on σ^2_α, σ^2_β, and $\sigma^2_{\alpha\beta}$ for the two-factor crossed design with interactionl of Equation (2.35) using model design MII - 5 of Table 2.6. The top plot is of the empirical coverage probabilities of the intervals, and the bottom plot is of the average interval lengths divided by the average interval lengths of **FID**.

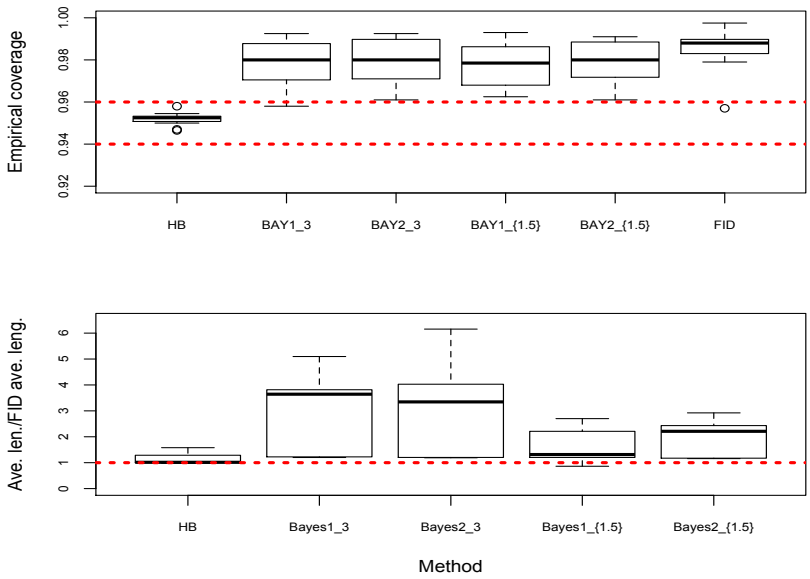


Figure A.11: Combined simulation results for 95% two-sided confidence intervals on σ^2_α, σ^2_β, and $\sigma^2_{\alpha\beta}$ for the two-factor crossed design with interactionl of Equation (2.35) using model design MII - 6 of Table 2.6. The top plot is the of empirical coverage probabilities of the intervals, and the bottom plot is of the average interval lengths divided by the average interval lengths of **FID**.

Appendix B

Supplementary material: classification of suspect powders

This information in this appendix was provided by Emily Snyder and Lukas Oudejans, and is included for completeness. They explain other procedures that have been developed for classification of LIBS data, the LIBS system and the materials used in the statistical analysis.

B.1 Background

Differentiation of *B. anthracis* spore powder LIBS spectra from LIBS spectra of other innocuous powders via classification methods can be difficult due to the inhomogeneity of the spore powder itself and the variability typically associated with LIBS spectra. This variability is seen even when employing the same LIBS system and is due to a range of factors including the following: pulse-to-pulse variations in the laser energy and profile, sample topography (directly affects the distance of the plasma to the collection lens which subsequently impacts the distance from the plasma to the collection fiber), creation of sampling craters (can be avoided by moving to a fresh spot for each laser shot), physical and chemical characteristics of the sample (surface adsorption, reflection and thermal conductivity which are determined by the composition, roughness, color and moisture content of the sample) and matrix effects (Tognoni et al., 2006; Rai and Thakur, 2007; Wisbrun et al., 1994; Tognoni et al., 2007; Lednev et al., 2010). Normalization methods, such as use of other emission lines from elements in the surrounding gas or reference elements in the matrix, and use of excitation temperatures and/or electron temperatures are often applied to correct for these matrix effects (Tognoni et al., 2006). However, these corrections frequently are not an option with heterogeneous

samples and/or when ungated (non-intensified) charge coupled device (CCD) detectors are used. These ungated CCDs are found on less expensive and portable LIBS systems.

Statistical methods have been implemented to overcome this issue, particularly in the area of analysis of biological agent powders. Using Partial Least Squares Discriminant Analysis (PLS-DA), Gottfried and coworkers were able to differentiate stand-off LIBS spectra intensity ratios of pure spore powders and other powders such as talcum powder, sugar, dust and flour on aluminum and glass substrates (Gottfried et al., 2008). Previously, biological agent surrogate spectra were classified using linear and rank correlation (Gibb-Snyder et al., 2006). These statistical techniques had difficulty distinguishing spore spectra, specifically *B. atrophaeus* spectra, in mixtures of potentially interfering compounds such as urban particulate matter. Munson and coworkers explored the use of soft independent modeling of class analogies (SIMCA) for classification of three *Bacillus* species, molds, Arizona road dust, and pollens as well as a mixture of Arizona road dust and *B. globigii*. They found that SIMCA models could be used to distinguish between spores in mixtures of the road dust and the road dust itself (Munson et al., 2005). PLS-DA was also able to distinguish between spores in mixtures of the road dust and the road dust itself (Gottfried et al., 2008). Employing these statistical methods still did not sufficiently resolve the issues of false positives for some materials (fertilizer and outdoor air particulate matter).

Other statistical methods have also been employed for classification of LIBS spectra of heterogeneous samples. Rehse et al. (2009) used discriminant function analysis (DFA) to discriminate LIBS spectra of one genus of bacteria, applied as a thin smear on an agar plate, from another and obtained greater than 90% accuracy regardless of the nutrient medium in which the bacteria were cultured (Rehse et al., 2009). Hierarchical cluster analysis, artificial neural networks (ANN) and PLS-DA were used to classify LIBS spectra of chicken tissue samples (kidney, lung, liver, brain, muscle and spleen) (Singh et al., 2009). ANNs were also employed to classify rocks and soils, and average classification accuracy of 78% was observed when spectra that were not used to train the original model were classified (including some spectra of unknown rock and soil materials) (Koujelev et al., 2010).

B.2 LIBS System

Data were collected on a bench-top LIBS system that consisted of a CFR400 Nd-YAG laser (Big Sky, Bozeman, Montana) operating at the fundamental wavelength of 1064 nm, a pulse duration of 8 ns, and maximum pulse energy of 400 mJ, a series of focusing and collection optics, and a LIBS 2000 broadband (200 to 980 nm with 0.1 nm resolution) spectrometer (Ocean Optics, Dunedin, Florida). During operation, a single laser pulse (≈65 mJ/pulse) from the laser is triggered by the LIBS software. This beam passes through a pierced parabolic mirror and is focused onto the sample surface with a 5 cm lens, producing the LIBS plasma. The resulting plasma emission is reflected by the pierced mirror to a 10 cm focal length lens that focuses the plasma emission onto a fiber optic bundle consisting of seven fibers. The fiber bundle delivers light to a broadband spectrometer that contains seven charge coupled devices (CCDs). Throughout operation of the system, the laser and the spectrometer are controlled by the Ocean Optics, Incorporated LIBS (OOILIBS) software. All spectra were taken at a delay time (time after plasma initiation) of 1.5 μs. Collection of plasma emission at this delay time optimizes the ratio of elemental emission lines to background plasma continuum emission.

B.3 Materials

Pellets (2.54 cm diameter, 2-4 mm thickness - depending on substance) were made using a pellet press (XPRESS 3630, SPEX Sample Prep Metuchen, NJ) that applied 20 tons of pressure for 30 seconds. The non-biological powder pellets analyzed via the bench-top LIBS system were: Food Lion Brand flour, Arm & Hammer Detergent, Rumford Baking Powder, Arm & Hammer Baking Soda, BC Powder, Crayola Chalk, DiPel 150 Dust, Equal artificial sweetener, Gain Laundry Detergent, Advil Ibuprofen tablets, Johnsons Baby Powder, Food Lion Brand Powdered Sugar, Food Lion Brand Sugar, SweetN Low artificial sweetener, Tide Laundry Detergent and Tylenol Acetaminophen Capsules. The 5% elemental standard powder pellets were made from the following powders: magnesium sulfate (99%, Sigma Aldrich yields magnesium spectral lines), sodium chloride (99.999%, Sigma Aldrich yields sodium

spectral lines), potassium iodide (≥99.0%, Sigma Aldrich yields potassium spectral lines) ferric sulfate hydrate (97%, Sigma Aldrich yields iron spectral lines), manganese(II) sulfate monohydrate (≥98.0%, Sigma Aldrich - yields manganese spectral lines), sand (white quartz 50 + 70 mesh, Sigma Aldrich - yields silicon spectral lines), graphite (99.99% 100 mesh powder, Sigma Aldrich -yields carbon spectral lines), calcium chloride (≥99%, Sigma Aldrich- yields calcium spectral lines)), and boron oxide (99.999% Alfa Aesar has minimal spectral features). Boron oxide was used as the diluent for the elemental standards because of its inertness and low spectral background. Spectral lines from the other component of the elemental standard (sulfur, iodine, chlorine) were not observed but hydrogen, nitrogen and oxygen spectral lines (due to the ambient air surrounding the sample) were observed in all samples except the graphite (which absorbs some emitted light from the plasma due to its color). The analyzed anthrax spore surrogate powders were *B. atrophaeus* (U.S. Army Dugway Proving Ground, Dugway, Utah), B. cereus (ATCC 14603), *B. thuringiensis* (ATCC 51912), *B. stearothermophilus* (ATCC 12979). All ATCC spores were used as received from ATCC. The B. atrophaeus was prepared as an 80:20 mixture of dry spores to fumed silica particles by mass (Brown et al., 2007). Spectra were also taken of a stainless steel coupon blank, used as the pellet backing during analysis. Note that DiPel 150 Dust has traces (less that 0.065%) of *B. thuringiensis*; however, because the LIBS system would not detect this low concentration of *B. thuringiensis*, we include DiPel 150 Dust as a confusant sample rather than a spore surrogate powder.

Bibliography

Abdelbasit, K. M. and Plackett, R. L. (1983), "Experimental design for binary data," *Journal of American Statistical Association*, 78, 90 – 98.

Barnard, G. A. (1995), "Pivotal Models and the Fiducial Argument," *International Statistical Reviews*, 63, 309–323.

Brown, G. S., Betty, R. G., Brockmann, J. E., Lucero, D. A., Souza, C. A., Walsh, K. S., Boucher, R. M., Tezak, M., Wilson, M. C., and Rudolph, T. (2007), "Evaluation of a wipe surface sample method for collection of Bacillus spores from nonporous surfaces," *Appl Environ Microbiol*, 73, 706 – 710.

Burch, B. D. (2011), "Assessing the performance of normal-based and REML-based confidence intervals for the intraclass correlation coefficient," *Computational Statistics and Data Analysis*, 55, 1018 – 1028.

Burch, B. D. and Iyer, H. K. (1997), "Exact Confidence Intervals for a Variance Ratio (or Heritability) in a Mixed Linear Model," *Biometrics*, 53, 1318 – 1333.

Burdick, R. K. and Graybill, F. A. (1992), *Confidence intervals on variance components*, New York: Marcel Dekker, Inc.

Casella, G. and Berger, R. L. (2002), *Statistical Inference*, Pacific Grove, CA: Wadsworth and Brooks/Cole Advanced Books and Software, 2nd ed.

Chopin, N. (2002), "A sequential particle filter method for static models," *Biometrika*, 89, 539–551.

— (2004), "Central Limit Theorem for Sequential Monte Carlo Methods and its Application to Bayesian Inference," *Ann. Statist.*, 32, 2385–2411.

Dalal, S. R., Fowlkes, E. B., and Hoadley, B. (1989), "Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure," *Journal of American Statistical Association*, 84, 945 – 957.

Dawid, A. P. and Stone, M. (1982), "The functional-model basis of fiducial inference," *Ann. Statist.*, 10, 1054–1074, with discussions by G. A. Barnard and by D. A. S. Fraser, and a reply by the authors.

Del Moral, P., Doucet, A., and Jasra, A. (2006), "Sequential Monte Carlo samplers," *J. Roy. Statist. Soc. Ser. B*, 68, 411–436.

Dempster, A. P. (1968), "A generalization of Bayesian inference. (With discussion)," *Journal Royal Statistical Society Series B*, 30, 205–247.

Diaconis, P. and Freedman, D. (1986a), "On inconsistent Bayes estimates of location," *Ann. Statist.*, 14, 68–87.

— (1986b), "On the consistency of Bayes estimates," *The Annals of Statistics*, 14, 1–67, with a discussion and a rejoinder by the authors.

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., and Weingessel, A. (2011), "e1071," .

Douc, R., Cappé, O., and Moulines, E. (2005), "Comparison of resampling schemes for particle filtering," in *In 4th International Symposium on Image and Signal Processing and Analysis*, pp. 64–69.

Douc, R. and Moulines, E. (2008), "Limit Theorems for Weighted Samples with Applications to Sequential Monte Carlo Methods," *The Annals of Statistics*, 36, 2344–2376.

Doucet, A., De Freitas, N., and Gordon, N. (2001), *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, New York: Springer-Verlag.

E, L., Hannig, J., and Iyer, H. K. (2008), "Fiducial Intervals for Variance Components in an Unbalanced Two-Component Normal Mixed Linear Model," *Journal of American Statistical Association*, 103, 854–865.

Elster, C. (2000), "Evaluation of measurement uncertainty in the presence of combined random and analogue-to-digital conversion errors," *Measurement Science and Technology*, 11, 1359 – 1363.

Faraggi, D., Izikson, P., and Reiser, B. (2003), "Confidence intervals for the 50 per cent response dose," *Statistics in Medicine*, 22, 1977 – 1988.

Fieller, E. (1954), "Some problems in interval estimation," *Journal Royal Statistical Society Series B*, 16, 175 – 185.

Fisher, R. A. (1930), "Inverse Probability," *Proceedings of the Cambridge Philosophical Society*, xxvi, 528–535.

— (1933), "The concepts of inverse probability and fiducial probability referring to unknown parameters," *Proceedings of the Royal Society of London series A*, 139, 343–348.

— (1935), "The Fiducial Argument in Statistical Inference," *Annals of Eugenics*, VI, 91 – 98.

Fraser, D. A. S. (1961a), "The fiducial method and invariance," *Biometrika*, 48, 261–280.

— (1961b), "On fiducial inference," *Ann. Math. Statist.*, 32, 661–676.

— (1966), "Structural probability and a generalization," *Biometrika*, 53, 1–9.

— (1968), *The Structure of Inference*, New York-London-Sydney: John Wiley & Sons Inc.

Frenkel, R. B. and Kirkup, L. (2005), "MonteCarlo-based estimation of uncertainty owing to limited resolution of digital instruments," *Metrologia*, 42, L27 – L30.

Gelman, A. (2006), "Prior distributions for variance parameters in hierarchical models," *Bayesian Analysis*, 1, 515–533.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis*, Texts in Statistical Science, Chapman and Hall, 2nd ed.

Geweke, J. (1989), "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, 57, 1317 – 1339.

Ghosal, S., Ghosh, J. K., and van der Vaart, A. (2000), "Convergence Rates of Posterior Distributions," *The Annals of Statistics*, 28, 500 – 531.

Gibb-Snyder, E., Gullett, B., Ryan, S., Oudejans, L., and Touati, A. (2006), "Development of size-selective sampling of Bacillus anthracis surrogate spores from simulated building air 12 intake mixtures for analysis via laser-induced breakdown spectroscopy," *Appl Spectrosc*, 60, 860 – 870.

Gottfried, J. L., De Lucia, F. C., Munson, C. A., and Miziolek, A. W. (2008), "Standoff detection of chemical and biological threats using laser-induced breakdown spectroscopy," *Appl Spectrosc*, 62, 353 – 363.

Grundy, P. (1956), "Fiducial distributions and prior distributions: an example in which the former cannot be associated with the latter," *J. Roy. Statist. Soc. Ser. B*, 18, 217 – 221.

GUM (1995), *Guide to the Expression of Uncertainty in Measurement*, International Organization for Standardization (ISO), Geneva, Switzerland.

Hannig, J. (2009a), "On Asymptotic Properties of Generalized Fiducial Inference for Discretized Data," Tech. Rep. UNC/STOR/09/02, Department of Statistics and Operations Research, The University of North Carolina.

— (2009b), "On Generalized Fiducial Inference," *Statistica Sinica*, 19, 491–544.

Hannig, J., Iyer, H. K., and Patterson, P. (2006), "Fiducial Generalized Confidence Intervals," *Journal of American Statistical Association*, 101, 254 – 269.

Hannig, J., Iyer, H. K., and Wang, J. C.-M. (2007), "Fiducial approach to uncertainty assessment: accounting for error due to instrument resolution," *Metrologia*, 44, 476–483.

Hannig, J. and Lee, T. C. M. (2009), "Generalized Fiducial Inference for Wavelet Regression," *Biometrika*, 96, 847 – 860.

Harris, P., Hann, M., Kirby, S. P. J., and Dearden, J. C. (1999), "Interval estimation of the median effective dose for a logistic dose - response curve," *Journal of Applied Statistics*, 26, 715 – 722.

Hartley, H. O. and Rao, J. (1967), "Maximum-Likelihood Estimation for the Mixed Analysis of Variance Model," *Biometrika*, 54, 93–108.

Hastie, T., Tibshirani, R., and Friedman, J. (2001), *The Element of Statistical Learning: data mining, inference, and prediction*, New York: Springer.

Hernandez, R. P. and Burdick, R. K. (1993), "Confidence Intervals and Tests of Hypotheses on Variance Components in an Unbalanced Two-factor Crossed Design with Interactions," *Journal of Statistical Computation and Simulation*, 47, 67–77.

Hernandez, R. P., Burdick, R. K., and Birch, N. J. (1992), "Confidence Intervals and Tests of Hypotheses on Variance Components in an Unbalanced Two-fold Nested Design," *Biometrical Journal*, 34, 387–402.

Huang, Y. (2001), "Various Methods of Interval EStimation of the Median Effective Dose," *Communications in Statistics - Simulation and Computation*, 30, 99 – 112.

Huang, Y., Harris, P., Kirby, S. P. J., and Dearden, J. C. (2002), "A modified Fieller interval for the interval estimation of effective doses for a logistic dose-response curve," *Computational Statistics and Data Analysis*, 40, 59 – 74.

Jasra, A., Stephens, D., and Holmes, C. (2007), "On population-based simulation for static inference," *Statistics and Computing*, 17, 263–279.

Jeyaratnam, S. and Graybill, F. A. (1980), "Confidence Intervals on Variance Components in Three-Factor Cross-Classification Models," *Technometrics*, 22, 375–380.

Jiang, J. (2007), *Linear and Generalized Linear Mixed Models and Their Applications*, New York: Springer.

Karatzoglou, A., Meyer, D., and Hornik, K. (2006), "Support Vector Machines in R," *Journal of Statistical Software*, 15.

Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004), "kernlab - An S4 Package for Kernel Methods," *R Journal of Statistical Software*, 11, 1 – 20.

Karim, M. R. and Zeger, S. L. (1992), "Generalized linear models with random effects; Salamander mating revisited," *Biometrics*, 48, 631 – 644.

Khuri, A. (1987), "Measures of Imbalance for Unbalanced Models," *Biometrical Journal*, 29, 383 – 396.

Khuri, A. and Littell, R. (1987), "Exact tests for the main effects variance components in an unbalanced random two-way model," *Biometrics*, 43, 545–560.

Khuri, A. and Sahai, H. (1985), "Variance Components Analysis: A Selective Literature Survey," *International Statistical Reviews*, 53, 279 – 300.

Kitagawa, G. (1996), "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *Journal of Computational and Graphical Statistics*, 5, 1 – 25.

Kong, A., Liu, J., and Wong, W. (1994), "Sequential Imputations and Bayesian Missing Data Problems," *J. Amer. Statist. Assoc.*, 89, 278–288.

Koujelev, A., Sabsabi, M., Motto-Ros, V., Laville, S., and Lui, S. L. (2010), "Laser-induced breakdown spectroscopy with artificial neural network processing for material identification," *Planet Space Sci*, 58, 682 – 690.

Kreger, M. D. (1992), *The LD50 (median lethal dose) and LC50 (median lethal concentration) toxicity tests January 1980 - March 1992*, Animal Welfare Information Center, Beltsville, Maryland.

Kuk, A. Y. C. (1995), "Asymptotically unbiased estimation in generalized linear models with random effects," *Journal of the Royal Statistical Society Series BRoyal Statistical Society Series B*, 57, 395 – 407.

Laird, N. M. and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963–974.

Le Cam, L. (1986), *Asymptotic Methods in Statistical Decision Theory*, Statistics, New York: Springer-Verlag.

Lednev, V., Pershin, S. M., and Bunkin, A. F. (2010), "Laser beam profile influence on LIBS analytical capabilities: single vs. multimode beam," *J Anal Atom Spectrom*, 25, 1745 – 1757.

Lee, Y. and Nelder, J. A. (1996), "Hierarchical generalized linear models," *Journal Royal Statistical Society Series B*, 58, 619 – 678.

Lee, Y., Nelder, J. A., and Pawitan, Y. (2006), *Generalized linear models with random effects: unified analysis via H-likelihood*, no. 106 in Monographs on Statistics and Applied Probability, Boca Raton, FL: Chapman and Hall/CRC.

Lindley, D. V. (1958), "Fiducial distributions and Bayes' theorem," *J. Roy. Statist. Soc. Ser. B*, 20, 102–107.

Lira, I. H. and Woger, W. (1997), "The evaluation of standard uncertainty in the presence of limited resolution of indicating devices," *Measurement Science and Technology*, 8, 441 – 443.

Liu, J. and Chen, R. (1995), "Blind Deconvolution via Sequential Imputations," *J. Amer. Statist. Assoc.*, 90, 567–576.

— (1998), "Sequential Monte Carlo Methods for Dynamic Systems," *J. Amer. Statist. Assoc.*, 93, 1032–1044.

Liu, J. and West, M. (2001), *Combined parameter and state estimation in simulation-based filtering*, New York: Springer-Verlag, chap. 10, Statistics for Engineering and Information Science, pp. 197 – 217.

McCulloch, C. E. (1997), "Maximum likelihood algorithms for generalized linear mixed models," *Journal of the American Statistical Association*, 92, 162 – 170.

McGilchrist, C. A. (1995), "Estimation in generalized linear models," *Journal of the Royal Statistical Society Series B*, 56, 61 – 69.

McGilchrist, C. A. and Aisbett, C. W. (1991), "Restricted BLUP for mixed linear models," *Biometrical Journal*, 2.

Mossman, D. and Berger, J. O. (2001), "Intervals for Posttest Probabilities: A Comparison of 5 Methods," *Medical Decision Making*, 21, 498 – 507.

Munson, C. A., De Lucia, F. C., Piehler, T., McNesby, K. L., and Miziolek, A. W. (2005), "Investigation of statistics strategies for improving the discriminating power of laser-induced breakdown spectroscopy for chemical and biological warfare agent simulants," *Spectrochim Acta B*, 60, 1217 – 1224.

Nason, C. (2008), *Wavelet Methods in Statistics with R*, New York: Springer.

Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized Linear Models," *J. Roy. Statist. Soc. Ser. A*, 135, 370 – 384.

O'Connell, A. A. and McCoach, D. B. (eds.) (2008), *Multilevel Modeling of Educational Data*, Information Age Publishing, Inc.

R Development Core Team (2011), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rai, V. N. and Thakur, S. N. (2007), *Laser-Induced Breakdown Spectroscopy*, Oxford, U.K.: Elsevier, chap. Physics of Plasma in LIBS, pp. 83 – 108.

Rehse, S. J., Jeyasingham, N., Diedrich, J., and Palchaudhuri, S. (2009), "A membrane basis for bacterial identification and discrimination using laser-induced breakdown spectroscopy," *J Appl Phys*, 105.

Rönnegard, L., Shen, X., and Alam, M. (2010), "hglm: a package for fitting hierarchical generalized linear models," *R Journal of Statistical Software*, 2, 20 – 28.

Satterthwaite, F. (1941), "Synthesis of variance," *Psychometrika*, 6, 309–316.

— (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*, 2, 110–114.

Schall, R. (1991), "Estimation in generalized linear models with random effects," *Biometrika*, 78, 719 – 27.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, New York: John Wiley & Sons Inc.

Shawe-Taylor, J. and Cristianini, N. (2000), *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, New York: Cambridge University Press.

Shin, H. and Cho, S. (eds.) (2003), *How to Deal with Large Datasets and Binary Output in SVM Based Response Models*, Korean Data Mining Society Conference.

Shun, Z. (1997), "Another look at the salamander mating data: a modified Laplace approximation approach," *Journal of American Statistical Association*, 92, 341 – 349.

Singh, J. P., Yueh, F. Y., Zheng, H. B., and Burgess, S. (2009), "Preliminary evaluation of laser- induced breakdown spectroscopy for tissue classification," *Spectrochim Acta B*, 64, 1059 – 1067.

Sitter, R. and Wu, C. (1993), "On the accuracy of Fieller intervals for binary response data," *Journal of American Statistical Association*, 88, 1021 – 1025.

Tang, Y., Zhang, Y.-Q., and Chawla, N. V. (2004), "Editorial: Special Issue on Learning from Imbalanced Data Sets," *SIGKDD Explorations*, 6, 1 – 6.

Taraldsen, G. (2006), "Instrument resolution and measurement accuracy," *Metrologia*, 43, 539 – 544.

Ting, N., Burdick, R. K., Graybill, F. A., Jeyaratnam, S., and Lu, T.-F. C. (1990), "Confidence Intervals on Linear Combinations of Variance Components that are Unrestricted in Sign," *Journal of Statistical Computation and Simulation*, 35, 135 – 143.

Tognoni, E., Cristoforetti, G., Legnaloli, S., Palleschi, V.; Salvetti, A., Mueller, M., Panne, U., and Gomushkin, I. (2007), "A numerical study of expected accuracy and precision in Calibration-Free Laser-Induced Breakdown Spectroscopy in the assumption of ideal analytical plasma," *Spectrochim Acta B*, 62, 1287 – 1302.

Tognoni, E.; Palleschi, V., Corsi, M., Cristoforetti, G., Omenetto, N., Gornushkin, I., Smith, B. W., and Winefordner, J. D. (2006), *From Sample to Signal in LIBS: a complex route to quantitative analysis*, Cambridge University Press, chap. From Sample to Signal in LIBS: a complex route to quantitative analysis, pp. 122 – 167.

Trevan, J. W. (1927), "The error of determination of toxicity," *Proceedings of the Royal Society of London series B: biological sciences*, 101, 483 – 514.

— (1951), "The design of toxicity tests," *British Medical Bulletin*, 7, 328 – 329.

Tsui, K.-W. and Weerahandi, S. (1989), "Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters," *J. Amer. Statist. Assoc.*, 84, 602–607.

van der Vaart, A. (2007), *Asympotitc Statistics*, no. 8 in Statististical and Probabilistic Mathematics, New York: Cambridge University Press.

Wandler, D. V. and Hannig, J. (In press), "Generalized Fiducial Confidence Intervals for Extremes," *Extremes*.

Weerahandi, S. (1993), "Generalized confidence intervals," *J. Amer. Statist. Assoc.*, 88, 899–905.

Williams, D. A. (1986), "Interval Estimation of the Median Lethal Dose," *Biometrics*, 42, 641 – 645.

Willink, R. (2007), "On the uncertainty of the mean of digitized measurements," *Metrologia*, 44, 73 – 81.

Wisbrun, R., Schechter, I., Niessner, R., Schroder, H., and Kompa, K. L. (1994), "Detector for Trace Elemental Analysis of Solid Environmental-Samples by Laser-Plasma Spectroscopy," *Analytical Chemistry*, 66, 2964 – 2975.

Wolfinger, R. D. and Kass, R. E. (2000), "Nonconjugate Bayesian Analysis of Variance Component Models," *Biometrics*, 56, 768 – 774.

Wu, G. and Chang, E. (2005), "KBA: kernel boundary alignment considering imbalanced data distribution," *IEEE Transactions*, 17, 786 – 795.

Zabell, S. L. (1992), "R. A. Fisher and the fiducial argument," *Statist. Sci.*, 7, 369–387.