1-1-2012

Disease Diagnosis from Immunoassays with Plate to Plate Variability

Oliver Entine
University of Pennsylvania, entine4@gmail.com

Disease Diagnosis from Immunoassays with Plate to Plate Variability

Abstract
The standard methods of diagnosing disease based on antibody microtiter plates are quite crude. Few methods create a rigorous underlying model for the antibody levels of populations consisting of a mixture of positive and negative subjects, and fewer make full use of the entirety of the available data for diagnoses. In this paper, we propose a Bayesian hierarchical model that provides a systematic way of pooling data across different plates, and accounts for the subtle sources of variations that occur in the optical densities of typical microtiter data. In addition to our Bayesian method having good frequentist properties, we find that our method outperforms one of the standard crude approaches (the "3SD Rule") under reasonable assumptions, and provides more accurate disease diagnoses in terms of both sensitivity and specificity.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Statistics

First Advisor
Dylan Small

Keywords
Chagas Disease, Disease diagnosis, Hierarchical Bayesian modeling, Immunoassay

Subject Categories
Biostatistics | Epidemiology | Statistics and Probability

DISEASE DIAGNOSIS FROM IMMUNOASSAYS WITH PLATE TO PLATE VARIABILITY

Oliver Entine

A DISSERTATION

in

Statistics

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2012

Supervisor of Dissertation

Dylan Small, Associate Professor, Statistics

Graduate Group Chairperson

Eric Bradlow, Professor of Marketing, Statistics, and Education

Dissertation Committee

Michael Levy, Assistant Professor, Epidemiology

Shane Jensen, Associate Professor, Statistics

DISEASE DIAGNOSIS FROM IMMUNOASSAYS WITH PLATE TO PLATE VARIABILITY

Acknowledgment

First and foremost, I would like to thank my advisor, Dr. Dylan Small, for his dedicated mentorship and support during my years at Wharton. Graduate school was a bit of a struggle for me, but he helped keep me motivated and encouraged to move forward in my research, even when I doubted my own progress. I would not have been able to succeed here without him.

I also would like to express my gratitude to the faculty for all their support. Dr. Shane Jensen and Dr. Michael Levy not only served on my dissertation committee, but also provided invaluable feedback on my work on this thesis and helped co-author one of my first submitted papers. Dr. Andreas Buja also served as mentor to me during this time, and during our Friday afternoon conversations, he taught me a whole new way of analyzing and visualizing data. Thanks also to Dr. Ed George, for bringing me into the statistics doctoral program, and for his continued faith in me.

I'm very grateful to my peers here at Wharton, especially my fellow cohort members, Mike Baiocchi, Hui Nie, and Sivan Aldor-Noiman, older mentors such as Blake McShane, Abhishek Gupta, and Shaun Lysen, and younger inspirations including James Piette, Sathya Anand, Emil Pitkin, and Justin Bleich. Whether collaborating on problem sets and research, griping about typical graduate student frustrations, or just hanging out in Philadelphia, you guys all helped make graduate school an

enjoyable and worthwhile experience.

Also supporting me during my time here at Penn were my old friends from Boston, Paul Messenger and Becky Hock, as well as my buddies from my undergraduate days at Penn, including Oren Isacoff and my fellow brothers of the Theta chapter of Sigma Alpha Mu. Fast and Firm to a stellar brotherhood that to this day continues to prosper.

Last but not least, I would like to thank Dad, Mom, and my sister Victoria for their ongoing support. I continue to be eternally greatful.

ABSTRACT

DISEASE DIAGNOSIS FROM IMMUNOASSAYS WITH PLATE TO PLATE VARIABILITY

Oliver Entine

Dylan Small

The standard methods of diagnosing disease based on antibody microtiter plates are quite crude. Few methods create a rigorous underlying model for the antibody levels of populations consisting of a mixture of positive and negative subjects, and fewer make full use of the entirety of the available data for diagnoses. In this paper, we propose a Bayesian hierarchical model that provides a systematic way of pooling data across different plates, and accounts for the subtle sources of variations that occur in the optical densities of typical microtiter data. In addition to our Bayesian method having good frequentist properties, we find that our method outperforms one of the standard crude approaches (the "3SD Rule") under reasonable assumptions, and provides more accurate disease diagnoses in terms of both sensitivity and specificity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Chapter 1

Introduction and Motivation

In the past two decades, Chagas disease has become an epidemic in rural communities in South America. Caused by the protozoan parasite *Trypanosoma cruzi*, Chagas has become the deadliest parasitic disease in the continent (Levy et al., 2011). More recently, the disease has also become a major urban problem in the city of Arequipa Peru (Levy et al., 2011).

Spread primarily by the insect vector *Triatoma infestans*, transmission of *T. cruzi* occurs when contaminated feces of the infected insect enters a person's bloodstream through the location of the insect bite, or from mucous membranes (Kirchhoff et al., 2004; Tustin et al., To appear). The parasite can also be transmitted through food contamination, blood transfusions, organ transplants, and in 26% of new cases, directly from a mother to her child during pregnancy (Bern et al., 2009).

Since the disease has only recently been introduced to this city, the debilitating symptoms of the disease elsewhere in South America are rarely seen in the hospitals of Arequipa (Levy et al., 2011). Most infected people here have the indeterminate form of

the disease and may not show symptoms for many years. Thus, various serologic and immunoassay tests are used to test potentially infected individuals. However, there are no "gold standards" reference tests for identifying the parasite (Tarleton et al., 2007). It is vital though that the disease be diagnosed during this asymptomatic stage, when patients will likely have a good prognosis in response to treatment. Once an individual reaches the stage where he starts showing physical symptoms, the treatment prognosis becomes quite poor. Therefore, the accuracy and effectiveness of Chagas diagnosis methods are of paramount importance.

The current immunoassay methodology for testing Chagas disease (as well as for various other infections) is often based on analyzing the optical density of patients' antibody levels, using a series of microtiter plates. However, the classification procedures currently derived from these microtiter plates are surprisingly crude, and do not make use of a comprehensive model that utilizes the plethora of available microtiter plate data. In this dissertation, we propose a hierarchical Bayesian model, that under reasonable assumptions, performs more accurate and reliable classifications of both negative and positive cases of diseases than the so-called 3-SD rule, one of the most common methods currently in use.

The structure of this dissertation is as follows. Chapter 2 provides background for the various Bayesian simulation techniques utilized throughout the thesis. In particular, we focus on the Metropolis-Hastings algorithm and the Gibbs sampler.

Chapter 3 is where we construct the Bayesian model that allows us to borrow strength across the different microtiter plates in creating a disease classification scheme. In addition to detailing the model's implementation, we use extensive simulations to demonstrate the improved accuracy of the full hierarchical framework versus the 3-

SD rule when our proposed underlying lognormal mixture model is valid. Finally, we compare the diagnostic results of these two methods on actual microtiter data from a lab in Peru, where they are currently using the 3-SD rule to test for Chagas disease.

In Chapter 4, we investigate the robustness of our procedure by looking at our model's Type I and Type II error rates when specific model assumptions are violated. In Chapter 5, we test the validity of the lognormal mixture model on this specific Chagas disease dataset, using both Bayesian posterior predictive checks, and also classical parametric bootstrap techniques. Finally, the Appendix provides documentation for an R function that can be used to run many of the MCMC simulations performed in the dissertation.

Chapter 2

Background on Markov Chain Monte Carlo Methods

One of the biggest challenges in Bayesian modeling is that one often needs to simulate random variables from very complex high-dimensional distributions, whose mathematical forms are non standard. These typically involve computing multivariate integrals that cannot be solved analytically. Markov Chain Monte Carlo (MCMC) methods are a class of algorithms in which one creates a Markov chain whose long-run stationary distribution is the complex distribution of interest. By carefully sampling iteratively from this chain, the eventual long-run samples will converge in distribution to this target density. The tricky part of performing an MCMC is the construction of an appropriate Markov Chain. The Bayesian simulation performed in this thesis uses two related methods of MCMC; The Metropolis Hastings Algorithm, and the Gibbs Sampler.

2.1. The Metropolis-Hastings Algorithm

The idea behind the Metropolis Hastings (MH) Algorithm is that by sampling from a simple, yet carefully chosen proposal density q , one can arrive at our complicated target distribution, denoted by π, by accepting or rejecting the draw with a certain probability (Hastings, 1970; Chib and Greenberg, 1996). The only assumptions we need for π is that the density of $\pi(x)$ be easy to evaluate up to a normalization constant for any vector x.

1. Select a reasonable starting value for the parameter vector $x^{(0)}$ that is within the support of π.

2. For $t = 1, 2, \ldots$, sample $x^* \sim q(x^*|x^{(t-1)})$.

3. Define the acceptance probability $r = min\{\frac{\pi(x^*)q(x^{(t-1)}|x^*)}{\pi(x^{(t-1)})q(x^*|x^{(t-1)})}, 1\}$

4. Sample $u \sim U[0, 1]$.

5. If $u < r$, then set $x^{(t)} = x^*$. Otherwise, set $x^{(t)} = x^{(t-1)}$.

An example of a common jumping proposal is a Normal Kernel, in which $x^* \sim N(x^{(t-1)}, \eta^2)$, where η^2 is a tuning parameter used to determine the size of the iteration "jumps" (Geweke, 1989). All of the proposal densities in the MH steps used in our analysis are of this form. This is an example of a symmetric proposal, as $q(x^*|x^{(t-1)}) = q(x^{(t-1)}|x^*)$. The expression for the acceptance probability then simplifies to $r = min\{\frac{\pi(x^*)}{\pi(x^{(t-1)})}, 1\}$.

One needs to choose the value of η^2 carefully. If η^2 is too large, the algorithm will reject too many draws, and the MCMC chain will tend to get stuck on certain values

for a long period of time. If η^2 is too small, one will accept a large fraction of the draws, but the jump sizes will be so small that the sampler will be exploring the parameter space very slowly, and may take an unreasonable amount of time to converge to the target distribution. Tierney (1994) and Roberts, Gelman, and Gilks (1996) describe more sophisticated methods for optimizing the step sizes for various jumping proposal kernels (Tierney, 1994; Gelman et al., 1996b).

2.2. Detailed Balance and the Intuition behind Metropolis-Hastings

Let $\pi(y)$ denote the density function of the target distribution that one wish to simulate from, and let $\pi(x)$ be the density function of a proposal density for a distribution from which is easy to simulate. Let $p(x, y)$ denote a transition kernel that assigns a probability of drawing a certain value y when currently at value x (and $p(x, y)$ defined likewise). We say that the function $p(x, y)$ satisfies *detailed balance* if $\pi(x)p(x, y) = \pi(y)p(y, x)$. Detailed balance essentially means that the probability that one observes a jump from x to y is the same as the probability that one observes a jump from y to x, assuming the Markov Chain has reached its stationary state. When detailed balance holds, then this Markov Chain procedure will ultimately lead to selecting values from the target density $\pi(y)$ (Chib and Greenberg, 1995).

If detailed balance does not hold, such as if $\pi(x)q(x, y) > \pi(y)q(y, x)$ for some x and y, then the process moves from x to y too often, and from y to x too rarely. To correct this, we need to reduce the chance of moving from x to y by introducing a probability $\alpha(x, y) < 1$ that the move is made in the first place. If the move is not made, the process returns x again as a value from the target distribution. Thus, transitions from x to y ($x \neq y$) are made according to $p_{MH}(x, y) \equiv q(x, y)\alpha(x, y)$, where $\alpha(x, y)$ is to be determined. It is now the transition kernel $p_{MH}(x, y)$ for which we want to

satisfy detailed balance, which is:

$$\pi(x)q(x,y)\alpha(x,y) = \pi(y)q(y,x)\alpha(y,x)$$

$$= \pi(y)q(y,x)$$

This implies that $\alpha(x,y) = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}$ provided that this value is less than 1. With this expression for $\alpha(x,y)$, using a transition kernel $p_{MH}(x,y)$ will allow a chain to converge to the proper target density π. The general logic of the MH algorithm, though a bit subtle, makes intuitive sense. One is allowed to stay put, and not make a transition a certain fraction of the time, so that the "net flow" between two states is in equilibrium.

2.3. The Gibbs Sampler

The Gibbs sampler is another technique for generating random numbers from non-standard complex joint-posterior distributions (Geman, 1988; Metropolis et al., 1953; Casella and George, 1992). The underlying idea is that if one can simulate values from the full conditional distributions of the individual parameters, then one can use these conditional distributions to construct a chain that satisfies detailed balance, and hence converges to the target density. The Gibbs sampler does not require one to be able to compute the joint density at a given point, in contrast to the standard Metropolis-Hastings algorithm, where one needed to evaluate $\pi(x)$ up to a normalization constant.

Let $x^{(t)} = (x_1^{(t)}, x_2^{(t)}, \ldots, x_n^{(t)})$ and let p_i be the full conditional distribution of x_i (that is conditioned on all parameter values except x_i). The algorithm is as follows.

1. Select a reasonable starting value for the parameter vector $x^{(0)}$ that is within

the support of π.

2. Given a previous iteration $x(t-1)$, generate the following:

   - $x_1^{(t)} \sim p_1(x_1|x_2^{(t-1)}, x_3^{(t-1)}, \ldots, x_n^{(t-1)})$.
   - $x_2^{(t)} \sim p_2(x_2|x_1^{(t)}, x_3^{(t-1)}, \ldots, x_n^{(t-1)})$.
   - $\vdots$
   - $x_n^{(t)} \sim p_n(x_n|x_1^{(t)}, x_2^{(t)}, \ldots, x_{n-1}^{(t)})$.

3. Repeat for $t = 1, 2, \ldots N$ where N is the number of desired iterations(before thinning and burning).

2.4. Intuition behind the Gibbs Sampler

The Gibbs sampler can actually be thought of a special case of the Metropolis Hastings algorithm (Gelman, 1992). Suppose that we begin with the vector

$$x^{(t-1)} = (x^{(t-1)}, x_2^{(t-1)}, x_3^{(t-1)}, \ldots, x_n^{(t-1)})$$

, and we want to now sample $x_1^{(t)}$ by considering the "jump" to a proposed vector

$$x^* = (x_1^*, x_2^{(t-1)}, x_3^{(t-1)}, \ldots, x_n^{(t-1)})$$

so that x^* only differs from $x^{(t-1)}$ in the first component. If we use the conditional density of x_1 from the Gibbs procedure as our proposal density, we then have

$$q(x^*|x^{(t-1)}) = p_1(x_1|x^{(t-1)}) = p_1(x_1^*|x_{-1}^{(t-1)})$$

To evaluate the density for the reverse jump move, $q(x^{(t-1)}|x^*)$, note that x^* and $x^{(t-1)}$ only differ in that first component. Therefore we have

$$q(x^{(t-1)}|x^*) = p_1(x_1^{(t-1)}|x_{-1}^{(t-1)})$$

Plugging these into the acceptance/rejection probability r for the MH given above, and rearranging a bit, we get

$$r = min\{\frac{\frac{\pi(x^*)}{p_1(x_1^*|x_{-1}^{(t-1)})}}{\frac{\pi(x^{(t-1)})}{p_1(x_1^{(t-1)}|x_{-1}^{(t-1)})}}, 1\}$$

Noting that $x = (x_1, x_{-1})$ and $x_{-1}^* = x_{-1}^{(t-1)}$, the laws of conditional probability give us

$$\frac{\pi(x^*)}{p_1(x_1^*|x_{-1}^{(t-1)})} = p(x_{-1}^{(t-1)})$$

$$\frac{\pi(x^{(t-1)})}{p_1(x_1^{(t-1)}|x_{-1}^{(t-1)})} = p(x_{-1}^{(t-1)})$$

Therefore, the equation for the ratio r reduces to $r = min(1, 1)$, and hence, the Gibbs sampler is a special case of the Metropolis-Hastings algorithm where we accept every draw with probability 1.

Chapter 3

A Hierarchical Bayesian Approach for Disease Diagnosis from Immunoassays with Plate to Plate Variability

The standard methods of diagnosing disease based on antibody microtiter plates are quite crude. Few methods create a rigorous underlying model for the antibody levels of populations consisting of a mixture of positive and negative subjects, and fewer make full use of the entirety of the available data for diagnoses. In this paper, we propose a Bayesian hierarchical model that provides a systematic way of pooling data across different plates, and accounts for the subtle sources of variations that occur in the optical densities of typical microtiter data. In addition to our Bayesian method having good frequentist properties, we find that our method outperforms one of the standard

crude approaches (the "3SD Rule") under reasonable assumptions, and provides more accurate disease diagnoses in terms of both sensitivity and specificity.

3.1. Introduction

The diagnosis of a large number of infectious diseases relies on the detection of antibodies in the sera or saliva of patients. The presence of a high amount of antibodies specific to a disease agent is a strong indicator of the presence of the disease, as these antibodies are evidence that the body is trying to fight off an infection, or has fought off an infection in the recent past. Though these antibody levels are hard to measure directly, one can measure the optical density of a sample through enzyme linked immunosorbent assays (ELISA), comparing these optical densities with those calibrated to antibody concentrations of known samples (i.e. controls), allows one to estimate the antibody level of those unknown samples.

ELISAs are typically run on a series of microtiter plates, each containing 96 wells (See Figure 1) . On each plate, the majority of wells contain samples from the population of interest, some of whom may be infected with the disease and others who are not. A portion of the wells on each plate are filled with negative controls (samples from patients known to be negative), while another portion of the wells contains samples from positive controls.

When analyzing data from microtiter plates, the standard procedures used to classify people as positive or negative are surprisingly crude. One such example is what we informally call "The 3SD Rule." For each microtiter plate, the negative control samples are used to compute an empirical average and standard deviation that is taken to represent the population of uninfected individuals. For each plate, a cutoff value is assigned as three SDs above the mean of these negative controls. All of the

Figure 1: Example of a 96-Well microtiter plate used for ELISA test.

unknown samples on that plate are classified as positive if their measurements exceed this cutoff, and are diagnosed as negative otherwise (Irion et al., 2002). This is an example of an *unpooled procedure*, as each plate has its own cutoff, determined only by measurements that are on that specific plate, and makes no use of the data from any of the other plates in the dataset. Even more sophisticated methods, which utilize serial dilution curves to better capture nonlinear relationships between optical densities and antibody concentrations, are unpooled procedures as each plate's dilution curve is derived only from measurements on that plate (Higgins et al., 1998; Gelman et al., 2004).

A more sophisticated method for diagnosis is by directly modeling the antibody levels of a population through a two-component mixture model, where each component corresponds to the negative and positive sub-populations, with an unknown fraction of people belonging to each group. Two-component mixture models have been used frequently in bioassay analysis (Moulton et al., 2002).

If all of the plates used for diagnostics were identical, one could fit a single two-component mixture model across all wells and plates. However, even under highly

controlled conditions, the plates are likely to have small, but noticeable differences that systematically skew the antibody levels of positive and negative subjects upward or downward. A second option would be to fit a separate mixture model for each plate. Such analysis would have limited power due to a relatively small sample size on each plate. Rather, we propose a compromise between these two extremes. Namely, we develop a hierarchical Bayesian approach, in which we consider the effects of the plates to come from a common distribution. This allows us to have distinct models for each plate, but also lets us "share" information across the plates regarding expected variation of within-plate antibody levels.

3.2. An Example of Microtiter Plate Data

To motivate the incorporation of plate effects into our model, we first look at a series of microtiter plates from a lab in Peru. Sera tested on these plates were previously found to react relatively weakly to a commercially-available ELISA test for Chagas disease (Delgado et al.). Researchers are therefore developing an ELISA test which derives antigens from a local strain of *T. cruzi*. In our exploratory data analysis, we denote the optical densities of the Chagas data by Y_{ij}, where i indexes plates, and j indexes wells within plates. We focus for now on differences in the logarithm of Y_{ij} across plates i. In the first set of boxplots (Figure 2, top), we plot $log(Y_{ij})$ for all of our non-control subjects. There are substantial differences in the average response across the 18 plates. However, this is not necessarily due to plate effects because some plates will likely have more positive subjects than others (even though we do not have definitive positive/negative diagnostic results for these subjects), and the difference between density levels of these two groups is expected to be large (even on the log scale). However, when we look at the plot of the optical levels for the same 7 negative controls on each plate, the disparity in the readings from plate-to-plate is also stark.

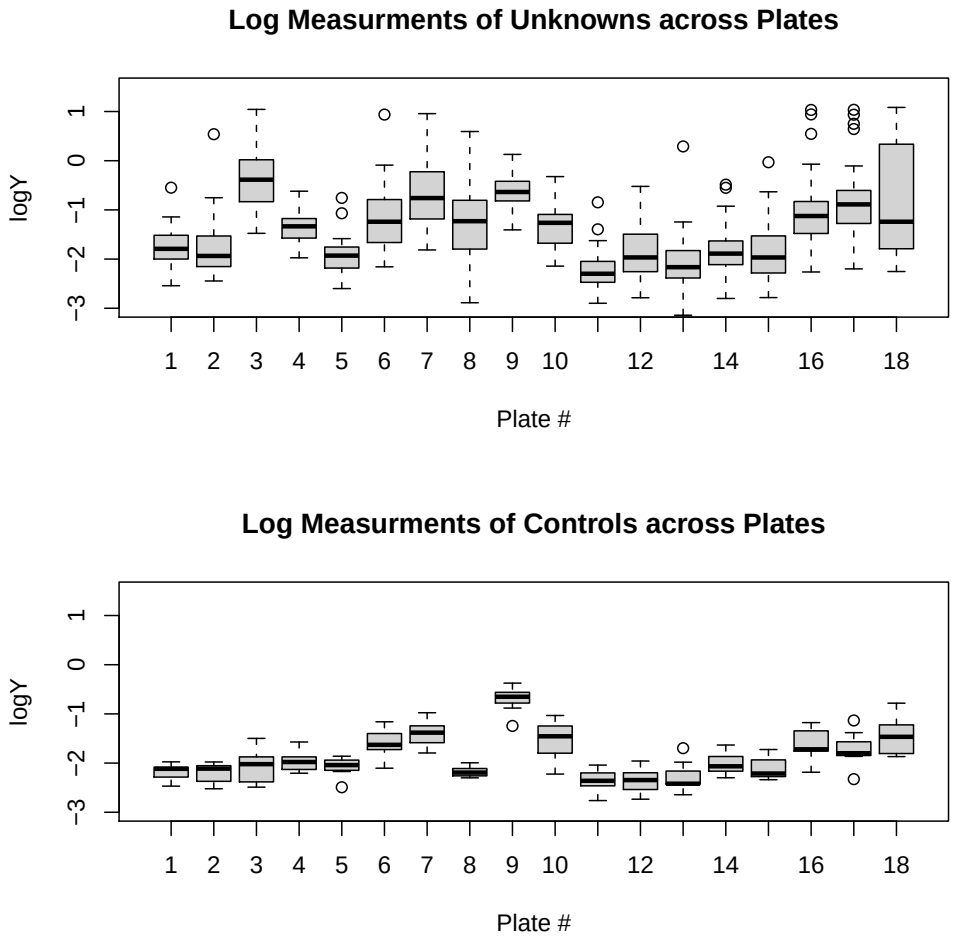

Figure 2: Boxplots of Logarithms of Optical Densities for Patients (above) and Negative Controls (below).

The bottom plot of Figure 2 suggests that the plate effects are substantial.

3.3. The Model

We consider a model in which the optical densities of the population come from a mixture of log-normal distributions (one component for people infected with the disease, one component for those not infected), but scaled by a factor corresponding to the contamination of a particular plate. Again, as introduced before, let Y_{ij} be the

optical density for an observation in the i-th plate and j-th well. β_i is the multiplicative effect from plate i upon the optical densities in its corresponding wells. α is the mixture plate parameter that indicates the fraction of subjects in the population who are infected. We augment our data with the latent variables I_{ij} where $I_{ij} = 1$ if Y_{ij} comes from an infected subject, and 0 otherwise.

Our model for the unknown data is $Y_{ij} = \beta_i (X_{ij}^{neg})^{1-I_{ij}} (X_{ij}^{pos})^{I_{ij}}$, where

$$log(X_{ij}^{neg}) \sim N(\mu_{neg}, \sigma_{neg}^2)$$

$$log(X_{ij}^{pos}) \sim N(\mu_{pos}, \sigma_{pos}^2)$$

If in conjunction to these unlabeled observations, we include some labeled data, namely the optical densities of a predetermined number of negative control wells, one can in theory obtain much better estimates for the parameters of this mixture model. In the particular dataset we will examine in Section 5, the negative control subjects are simply a group of 7 people who are known to be free of Chagas disease. Each plate has seven wells, each corresponding to one of these negative controls. Thus we observe variability both between controls, as indicated by the variation across these seven optical densities on a given plate, and also variability within the control subjects, measured by the variation of readings of each individual across the plates. To fully reflect the nature of these replications, our model should be able to capture both of these forms of variations. We let Y_{ij}^c be the optical density for the i-th control subject on the j-th plate (or the j-th replication).

For the negative control data, we have $Y_{ij}^c = \beta_i (X_{ij}^c)$, and to distinguish the two types

of variation, we use the following construction.

$$log(X_{ij}^c) \sim N(\tau_j, \kappa^2)$$

$$\tau_j \sim N(\mu^{neg}, \lambda^2)$$

Here, κ^2 represents the plate-to-plate variation of a control's replicated samples, and λ^2 captures the variation between labeled control subjects. The advantage of this formulation is that the unconditional distribution of $log(X_{ij}^c)$ is itself normal, with mean μ_{neg} and variance $\kappa^2 + \lambda^2$, where we naturally set $\sigma_{neg}^2 = \kappa^2 + \lambda^2$.

Next, since we would like the plate effects to be strictly positive and centered around 1 (corresponding to "neutral" plate effect), we let β_i come from a lognormal distribution, and $log(\beta_i) \sim N(-\frac{\nu^2}{2}, \nu^2)$. (Since the mean of a log-normal distribution is given by $e^{\mu+\frac{\sigma^2}{2}}$, this parametrization ensures that the mean of β_i is 1).

Finally, we impose the following uninformative prior distributions on our parameters:

$$\mu_{neg}, \mu_{pos} \sim Unif(-\infty, \infty)$$

$$\frac{1}{\sigma_{pos}^2} \sim Gamma(0, 0)$$

$$\lambda^2, \kappa^2 \sim Unif(0, \infty)$$

$$\alpha \sim Unif(0, 1)$$

$$\frac{1}{\nu^2} \sim Gamma(5, 5)$$

.

Our ultimate goal is to simulate draws from the posterior distribution of our unknown

parameters.

$$P(\mu_{neg}, \mu_{pos}, \sigma^2_{neg}, \sigma^2_{pos}, \alpha, \beta, \nu^2|\mathbf{Y})$$

To simplify our notation, let us define $y_{ij} = log(Y_{ij})$, $y^c_{ij} = log(Y^c_{ij})$ $b_i = log(\beta_i)$, and let $\theta = (\mu_{neg}, \mu_{pos}, \sigma^2_{pos}, \kappa^2, \lambda^2)$, so that θ represents the set of parameters with improper flat prior distributions. Taking advantage of the conditional independence of $\mathbf{y}$ and $\mathbf{y^c}$, we write out the joint posterior as:

$$P(\mathbf{I}, \vec{b}, \vec{\tau}, \nu^2, \alpha, \theta|\mathbf{Y}) \propto$$

$$P(\mathbf{y}|\mathbf{I}, \vec{b}, \vec{\tau}, \nu^2, \alpha, \theta)P(\mathbf{y^c}|\mathbf{I}, \vec{b}, t\vec{a}u, \nu^2, \alpha, \theta)P(\mathbf{I}|\vec{b}, \vec{\tau}, \nu^2, \alpha, \theta)P(\vec{b}, \vec{\tau}|\nu^2, \alpha, \theta)P(\nu^2|\alpha, \theta)P(\alpha, \theta)$$

The components of this posterior distribution are:

$$\begin{aligned}
P(\mathbf{y}|\mathbf{I}, \mathbf{b}, \vec{\tau}, \nu^2, \alpha, \theta) &\propto \prod_{i,j} \left(\frac{1}{\sigma_{neg}} e^{\frac{-(y_{ij}-b_i-\mu_{neg})^2}{2\sigma^2_{neg}}} \right)^{1-I_{ij}} \left(\frac{1}{\sigma_{pos}} e^{\frac{-(y_{ij}-b_i-\mu_{pos})^2}{2\sigma^2_{pos}}} \right)^{I_{ij}} \\
P(\mathbf{y^c}|\mathbf{I}, \mathbf{b}, \vec{\tau}, \nu^2, \alpha, \theta) &\propto \prod_{i,j} \left(\frac{1}{\kappa} e^{\frac{-(y^c_{ij}-b_i-\tau_j)^2}{2\kappa^2}} \right) \\
P(\mathbf{I}|\mathbf{b}, \vec{\tau}, \nu^2, \alpha, \theta) &= \alpha^{\sum_{i,j} I_{ij}} (1-\alpha)^{\sum_{i,j}(1-I_{ij})} \\
P(\mathbf{b}|\vec{\tau}, \nu^2, \alpha, \theta) &\propto \prod_i \frac{1}{\nu} e^{\frac{-(b_i+\frac{\nu^2}{2})^2}{2\nu^2}} \\
P(\vec{\tau}|\nu^2, \alpha, \theta) &\propto \prod_j \frac{1}{\lambda} e^{\frac{-(\tau_j-\mu_{neg})^2}{2\lambda^2}} \\
P(\nu^2|\alpha, \theta) &\propto \frac{1}{(\nu)^6} e^{\frac{-5}{\nu^2}} \\
P(\alpha, \theta) &\propto 1
\end{aligned}$$

where $\mathbf{Y} = (\mathbf{y}, \mathbf{y^c})$. We obtain samples from the posterior distribution via a Markov

Chain Monte Carlo (MCMC) scheme, consisting of Gibbs Sampling and Metropolis-Hastings steps (Casella and George, 1992; Hastings, 1970).

For each iteration of our MCMC implementation, we obtain a vector of parameters sampled from the joint posterior distribution. Combined with the optical density readings, we can calculate the posterior probability that a particular sample comes from the infected group. We then average these posterior probabilities (on a well-by-well basis) over all the iterations within that chain. Our classification rule is based on cutoff values for these posterior probabilities. If a sample's average posterior probability is less than the predetermined cutoff, we classify the sample as negative. If the probability of being positive exceeds the cutoff, we classify that sample as positive. We then compute usual measure of sensitivity and specificity of our procedure by determining how many of our classified negatives/positives match up with the actual negative/positive state of the sample (which we know definitively in this case, since the data is simulated). A higher probability cutoff will naturally lead to fewer true positives (and hence a lower sensitivity), and also lead to more true negatives (and thus higher specificity).

3.4. Details of MCMC Implementation

Due to the conjugacy of our prior distributions with the joint-likelihood function, many of conditional posterior distributions come from standard distributions, and we can sample them directly.

Sampling the Well Indicators

The positive/negative labels for each well come from Bernoulli distributions.

$$I_{ij}|y_{ij}, b_i, \nu^2, \alpha, \theta \sim Bin(1, p_{ij})$$

$$p_{ij} = \frac{\alpha\phi(z_{ij}^{pos})}{\alpha\phi(z_{ij}^{pos}) + (1-\alpha)\phi(z_{ij}^{neg})}$$

where ϕ is the pdf for a standard normal, $z_{ij}^{pos} = \frac{y_{ij}-b_i-\mu_{pos}}{\sigma_{pos}}$, and $z_{ij}^{neg} = \frac{y_{ij}-b_i-\mu_{neg}}{\sigma_{neg}}$

Sampling the Normal-Normal Mixture Parameters

Letting nC be the number of controls on each plate, $n_0 = \sum(1 - I_{ij})$ and $n_1 = \sum I_{ij}$, the conditional posterior distribution for the disease prevalence α is a beta distribution:

$$\alpha|\mathbf{y}, \mathbf{I}, \vec{b}, \nu^2, \theta \sim Beta(1 + n_1, 1 + n_0)$$

Letting $d_{ij} = y_{ij} - b_i$, the conditional densities of the two means of the mixture model are the following normal distributions.

$$\mu_{neg}|\mathbf{y}, \mathbf{I}, \vec{b}, \vec{\tau}, \nu^2\, \alpha, \theta_{-1} \sim N\left(\frac{\frac{\sum d_{ij}\mathbb{1}I_{ij}=0}{\sigma_{neg}^2} + \frac{\sum \tau_j}{\lambda^2}}{\frac{n_0}{\sigma_{neg}^2} + \frac{nC}{\lambda^2}}, \frac{1}{\frac{n_0}{\sigma_{neg}^2} + \frac{nC}{\lambda^2}}\right)$$

$$\mu_{pos}|\mathbf{y}, \mathbf{I}, \vec{b}, \vec{\tau}, \nu^2, \alpha, \theta_{-2} \sim N(\frac{\sum d_{ij}\mathbb{1}I_{ij} = 1}{n_1}, \frac{\sigma_{pos}^2}{n_1})$$

Since the inverse-gamma prior distribution for σ^2_{pos} is conjugate to the normal likelihood, its posterior distribution is also inverse-gamma. Letting $SS_{pos} = \sum(z^{pos}_{ij} - \mu_{pos})^2$, the conditional posterior distribution of the variance of the positive group is given by:

$$\sigma^2_{pos}|\mathbf{y}, \mathbf{I}, \vec{b}, \vec{\tau}, \nu^2, \alpha, \theta_{-4} \sim InvGamma(\frac{n_1}{2}, \frac{SS_{pos}}{2})$$

Sampling the Plate Effect Parameters

Next, we sample the (log) plate effect parameters, $\mathbf{b}$. If we let $z^{\beta}_{ij} = y_{ij} - \mu_{ij}$, then $z^{\beta}_{ij} \sim N(b_i, \sigma^2_{ij})$ where μ_{ij} and σ^2_{ij} correspond to the positive/negative assignment of observation y_{ij}, y^c_{ij}. This implies that the corresponding conditional posterior distribution for b_i is:

$$b_i|y_{ij}, I_{ij}, \nu^2, \alpha, \theta \sim N(m_i, s^2_i) \text{ where}$$

$$m_i = \frac{\sum_j \frac{z^{\beta}_{ij}}{\sigma^2_{ij}} + \frac{1}{2}}{\sum_j \frac{1}{\sigma^2_{ij}} + \frac{1}{\tau^2}}$$

$$s^2_i = \frac{1}{\sum_j \frac{1}{\sigma^2_{ij}} + \frac{1}{\tau^2}}$$

Sampling the variance ν^2 of the log of the plate effects $\mathbf{b}$ is trickier. Since the distribution of $\mathbf{b}$ is constrained so that β has mean 1, the modeled distribution of b_i has the form $b_i \sim N(-\frac{\nu^2}{2}, \nu^2)$, which is no longer conjugate to the $InvGamma(5, 5)$ prior distribution imposed on ν^2. Thus, the posterior conditional distribution $\nu^2|y_{ij}, I_{ij}, b_i, \alpha, \theta$ is sampled indirectly using a Metropolis-Hastings step.

Sampling the Means of the Negative Controls, $\vec{\tau}$

Letting nP be the total number of plates (and hence number of replications for each negative control subject), and $d^c_{ij} = y^c_{ij} - b_i$, the posterior conditional distributions

for each τ_j are normal.

$$\tau_j|\mathbf{y}, \mathbf{I}, \vec{b}, \nu^2, \alpha, \theta \sim N\left(\frac{\frac{\mu_{neg}}{\lambda^2} + \frac{\sum_i d_{ij}^c}{\kappa^2}}{\frac{1}{\lambda^2} + \frac{nP}{\kappa^2}}, \frac{1}{\frac{1}{\lambda^2} + \frac{nP}{\kappa^2}}\right)$$

Sampling the Two Variance Components of σ^2_{neg}

Finally, we need to sample the two components of σ^2_{neg}, κ^2 and λ^2. The conditional posterior distribution for the negative control parameters (conditioned on μ_{neg} and $\mathbf{b}$), where nC is the number of control subjects, and nP is number of plates, is now:

$$P(\tau, \kappa^2, \lambda^2|\mathbf{y^c}, \mathbf{y^{neg}}) \propto P(\tau|\lambda^2)P(\mathbf{y^c}|\tau, \kappa^2)P(\mathbf{y^{neg}}|\lambda^2, \kappa^2)$$

$$\propto \frac{1}{(\lambda^2)^{\frac{nC}{2}}} e^{\left(\frac{-\sum_i(\tau_j-\mu_{neg})^2}{2\lambda^2}\right)} \frac{1}{(\kappa^2)^{\frac{nC*nP}{2}}} e^{\left(\frac{-\sum_{i,j}(y_{ij}^c-b_i-\tau_j)^2}{2\kappa^2}\right)} \frac{1}{(\lambda^2+\kappa^2)^{\frac{n0}{2}}} e^{\left(\frac{-\sum_{i,j}(y_{ij}^{neg}-b_i-\mu_{neg})^2}{2(\kappa^2+\lambda^2)}\right)}$$

Since the resulting conditional distributions for λ^2 and κ^2 are non-standard, we incorporate another Metropolis step to sample from λ^2 and then κ^2 (Alternatively, one could also sample these two parameters jointly (Hastings, 1970)).

Implementation

The entire MCMC procedure above can be implemented in R using the *ImmunoassayMixture()*, whose output is a matrix containing the iterations for all the key parameters in the model. This function also allows one to monitor the chains carefully to ensure that the sampler is converging to non-degenerate parameters. In particular, when the two components of the mixture model have too much overlap, the MCMC can lead to parameter estimates that are poorly identified (Lindsay and Roeder, 1993). *ImmunoassayMixture()* and the corresponding documentation file can be found in the

supplemental materials.

3.5. Simulation Study

In Bayesian models based on noninformative priors, we want the model to produce results that are consistent with classical frequentist interpretations (Hobert et al., 2011; Morris, 1983; Rubin, 1984). To assess the frequentist performance of our methodology, we will consider the following:

1. The overall coverage rate of our posterior distributions. Do our 95% Bayesian posterior intervals in fact cover their respective "true" parameter values 95% of the time (sample-to-sample variation)?

2. Do the posterior means of our parameter estimates accurately capture the true parameter values, or are they biased upward or downward?

3. Does adding a substantial number of negative controls greatly increase the accuracy of our procedures?

In testing these considerations, we first generated 100 distinct simulated datasets, setting "true values" for α, μ_{neg}, μ_{pos}, κ^2,λ^2, σ^2_{pos}, and ν^2. These values were chosen to approximately reflect the typical optical measurements in out data. Using these parameters, we then sampled values for $\mathbf{b}$ and τ. For simplicity, we assumed there were 20 plates, and 40 unlabeled wells on each plate, and that there were 10 negative controls replicated on each plate. For each dataset, we ran our MCMC sampler for 6,500 iterations (the first 2,000 of which were discarded, after checking that our sampler has converged to the correct joint distribution) and recorded the 95% posterior intervals for each of these parameters, as well as the posterior means. After running this procedure for all 50 datasets, we examined each parameter and observed both

the fraction of times that parameter's posterior interval contained its true value, and also the average squared deviation between the posterior mean and the true value.

Table 1 shows our results for simulated datasets with 20 plates, and 10 negative controls per plate. For the parameters of interest, we observed seemingly good coverage rates. Namely, the fraction of datasets that resulted in intervals that contained each of the "true" values for these parameters ranged between 90 and 98%, close to the desired 95% coverage.

Another check as to the accuracy of our procedure is to see how the posterior means differ from the true underlying parameter. Here, we define bias as the simple mean of $(\hat{\theta} - \theta)$ across the 100 simulations, where $\hat{\theta}$ are the posterior means of the respective parameters. Here, the biases of these parameters are all positive. However, the magnitudes of these biases are generally pretty small compared to the true values, and thus should not be of too much concern.

Though the average squared deviation of a posterior mean and a fixed parameter is a bit of an awkward metric, it does represent the overall accuracy of this procedure. These MSE values seem quite low. Most encouraging are the small errors in estimating κ^2, λ^2, and σ^2_{neg}, as this is the most subtle part of our model, and most complicated to implement. The only possible concern is the relatively large bias and MSE for ν^2. This probably just reflects the uncertainty of measuring an effect where we only have 20 plates from which to infer the between-plate variation.

To get a sense as to the effect of number of plates in the accuracy of our results, we performed another simulation with just 5 plates instead of 20 (each plate still consisted of 10 negative controls and 40 unknown subjects). The coverage rates for these parameters are only slightly lower than they were above, except for ν^2, which

Parameter	True Value	Coverage Rate	Mean Squared Error	Bias
μ_{neg}	-1.88	0.94	0.0399	0.0332
μ_{pos}	0.20	0.95	0.0480	0.0246
σ^2_{neg}	0.27	0.96	0.0003	0.0027
σ^2_{pos}	0.36	0.94	0.0105	0.0259
α	0.125	0.94	0.0003	0.0052
λ^2	0.17	0.93	0.0004	0.0012
κ^2	0.10	0.91	0.0001	0.0015
ν^2	0.49	0.97	0.0639	0.1285

Table 1: Simulation Results for 20 plates, 10 Negative Controls per plate

has dropped to 87% (Table 2). However, we begin to see drastic differences in the MSE's and biases, which are now an order of magnitude larger than they were when we used more plates. This suggests that our Bayesian procedure is quite sensitive to the number of plates we incorporate into our model.

Parameter	True Value	Coverage Rate	"Mean Squared Error"	"Bias"
μ_{neg}	-1.88	0.90	0.4369	0.3518
μ_{pos}	0.20	0.92	0.4307	0.2888
σ^2_{neg}	0.27	0.94	0.0014	0.0109
σ^2_{pos}	0.36	0.95	0.0563	0.0747
α	0.125	0.96	0.0014	0.0164
λ^2	0.17	0.97	0.0013	0.0027
κ^2	0.10	0.96	0.0005	0.0082
ν^2	0.49	0.87	1.2519	0.7318

Table 2: Simulation Results for 5 plates, 10 Negative Controls per plate

Lastly, we asses the performance of our model when we reduce the number of controls on each plate. Here, we use 20 plates, but now use just 3 controls (replicated on each plate) instead of 10. Compared with the 10-control simulation, our biases and MSE's seem more extreme. Our coverage rates though are fairly good, except for μ_{neg}, whose coverage rate drops to 89% (Table 3). This could simply be a consequence of the fact that when we have fewer controls, we have fewer negative data points, so our estimates for the parameters of the subgroup of negative patients would likely be less

precise.

Parameter	True Value	Coverage Rate	"Mean Squared Error"	"Bias"
μ_{neg}	-1.88	0.89	0.2121	0.2187
μ_{pos}	0.20	0.94	0.2320	0.1787
σ^2_{neg}	0.27	0.98	0.0006	-0.0006
σ^2_{pos}	0.36	0.96	0.0221	0.0362
α	0.125	0.96	0.0008	0.0079
λ^2	0.17	0.95	0.0017	-0.0193
κ^2	0.10	0.91	0.0013	0.0187
ν^2	0.49	0.91	0.5501	0.4301

Table 3: Simulation Results for 20 Plates, 3 Negative Controls per Plate

We also evaluated our classification of positive vs. negative samples via a simulation study. For the first analysis, we look carefully at the model with 20 plates, but only 3 negative controls. In Figure 3, we plot the sensitivity and specificity of our criterion as a function of the posterior probability cutoff. For our procedure, the sensitivity remains fairly high for cutoff probabilities as high as 0.8 (with sensitivity of 0.8), before dropping off for more stringent tests. The specificity for our procedure is extremely high for probability cutoffs as low as 0.1. Thus, nearly all true negative samples have tiny posterior probabilities of being in the positive group. These extremely high sensitivity and specificity characteristics for our procedure are a consequence of the fact that the lognormal mixture distribution for the optical densities leads to fairly high separation between positive and negative values.

Next, we look at the Receiver Operating Characteristic (ROC) for our method, and compare it to the classification rates using the 3 SD Rule (Figure 4) (Opsteegh et al., 2010; Greiner et al., 2000). Though this type of analysis was originally developed as a frequentist technique (Kurkjian et al., 2005), it extends nicely to the Bayesian framework (Choi et al., 2006; Wang et al., 2007), and has been used previously in Bayesian analysis of ELISA tests (Limmathurotsakul et al., 2011; Nielsen et al., 2002).

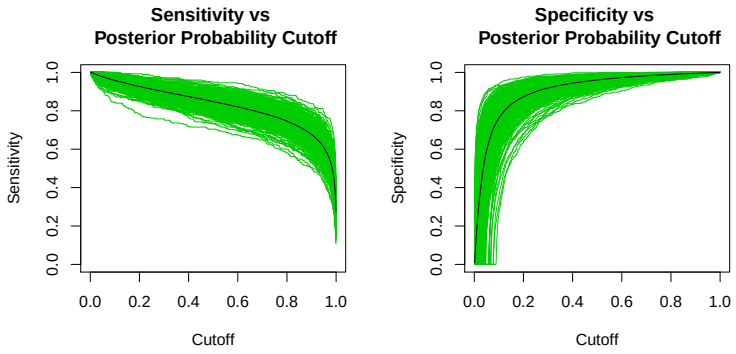

Figure 3: Sensitivity and Specificity for 20 Plates, 3 Negative Controls. Green curves represent individual simulations, the black line reflects the average rates for all 1000 simulations.

Our method clearly outperforms the 3 SD rule, as the blue X (denoting the average ROC values of the 3 SD rule across 1000 datasets) and the majority of the red circles (corresponding to individual simulations) lie to the lower right of the black line (our procedure). The 3 SD rule results in roughly a 0.94 and 0.89 sensitivity and specificity respectively. Using our method, if we required 0.94 sensitivity, we could improve specificity to 0.98. If instead we require a specificity of 0.89, we could improve our specificity to nearly 0.99. Thus, one could say that when using just 3 controls, our method dominates the industry standard in terms of classification rates.

We perform a similar analysis for a series of microtiter plates with 7 controls, instead of 3, to more mimic our real optical density data. In Figure 5, the sensitivity and specificity curves are more ideal including the additional 4 controls. The same probability cutoffs lead to slightly higher sensitivity and specificity rates than before.

Looking at the ROC curves (Figure 6), we see that the 3SD rule performs much better now as compared to how it performed in the setting with just 3 controls, as the red dots hover much closer to the black line. This makes sense, as the 3 standard deviation cutoff for each plate is now computed using 7 controls, as opposed to just 3.

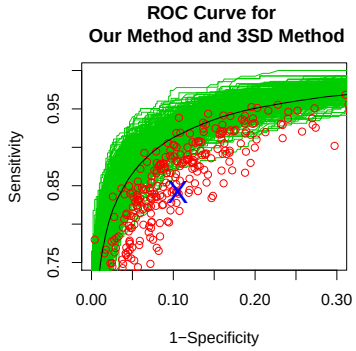

Figure 4: Plot of Sensitivity vs. 1-Specificity for the two methods (3 controls). Green curves represent simulated posterior ROC curves under our method. Black line reflects the average ROC curve across all 1000 simulations. Red dots correspond to the sensitivity/specificities of the individual simulations under the 3SD rule. The blue "X" is the average sens./spec. of the 3SD rule under all 1000 simulations.

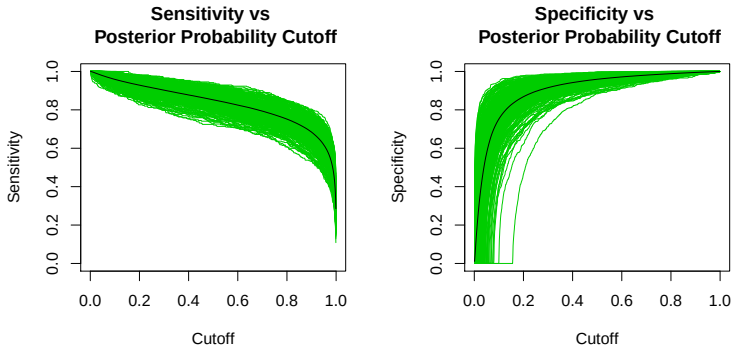

Figure 5: Sensitivity and Specificity for 20 Plates, 7 Negative Controls. Green curves represent individual simulations, the black line reflects the average rates for all 1000 simulations.

However, while better than before, the 3SD classification procedure is still dominated by our method.

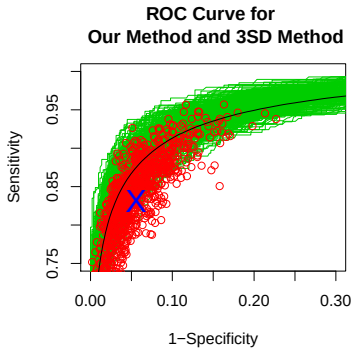

Figure 6: Plot of Sensitivity vs. 1-Specificity for the two methods (7 controls). Green curves represent simulated posterior ROC curves under our method. Black line reflects the average ROC curve across all 1000 simulations. Red dots correspond to the sensitivity/specificities of the individual simulations under the 3SD rule. The blue "X" is the average sens./spec. of the 3SD rule under all 1000 simulations.

3.6. Analysis of Chagas Disease Data

Applying our model to the 18 microtiter plates introduced in Section 2, we obtain the posterior estimates for our parameters given in Table 4.

Parameter	2.5% Quantile	50% Quantile	97.5% Quantile
μ_{neg}	-1.614	-1.364	-0.989
μ_{pos}	0.019	0.602	0.979
σ^2_{neg}	0.189	0.229	0.261
σ^2_{pos}	0.026	0.064	0.378
α	0.027	0.042	0.086
λ^2	0.066	0.110	0.157
κ^2	0.089	0.116	0.151
ν^2	0.125	0.256	0.609

Table 4: Posterior Intervals for Parameters

Next, we examine the magnitude of the estimated plate effects, as determined by the median $\log(\beta)$ (Figure 7). Given how we constructed the model for the plate effects,

it is not surprising that the 18 parameters are centered around 0. The magnitudes of these effects is surprisingly strong. These values suggest that if we take a sample from one plate, and transfer it over to another, the log of the measured optical density can easily increase by 1.0 units. This is consistent with the parameter values above, where ν^2 is noticeably larger than σ^2_{neg} and σ^2_{pos}, suggesting that a larger amount of the variation in the optical densities can be attributed to the plate effect than to the heterogeneity of the actual test subjects. Also somewhat surprising is the nature in which the magnitudes of the plate effects cluster together, in particular for those around -0.5.

For this particular application, our method ends up being far more conservative in terms of classifying patients as having the disease. In Figure 8, we look at 4 representative plates, where we overlay our mixture density onto the histogram of the log of the measurements (using the posterior mean of our parameters to generate the density curves). The vertical lines represent the plates' respective cutoffs under the 3SD rule. For every one of our plates, the 3SD classification threshold is to the left of the mode of the positive population density in our mixture. Thus, for any posterior probability cutoff we choose in our method, our procedure will be far more conservative than the one based on the 3SD rule, for each of the plates in our data.

When we use these two procedures to compare the classifications, there are four possible regimes;

- Observations classified as positive under both methods.
- Observations classified as negative under both methods.
- Observations classified as positive under the 3SD rule, but negative under our

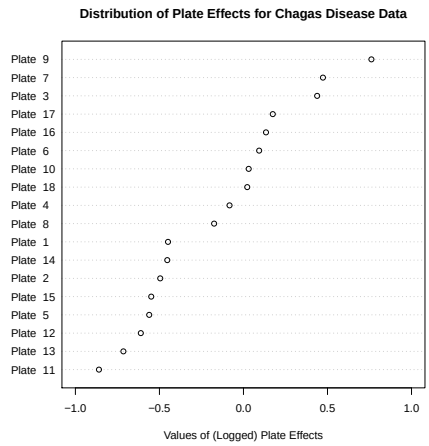

Figure 7: Dot plot of the magnitudes of the 18 estimated plate effects, sorted in increasing order. The lowest dashed line corresponds to the plate with the most negative effect, the highest dashed line to the most positive of the 18 plates, and so on.

Bayesian model.

- Observations classified as negative under the 3SD rule, but positive under our Bayesian model.

Due to the stark differences in the sensitivity of these two tests, there were no individual subjects in that fell into this last regime (negative under 3SD, but positive under our model). Figure 9 shows how often observations fell into one of these three regimes as a function of the posterior probability cutoff in our method.

Under our Bayesian model, the estimated posterior probabilities for particular observations belonging to the positive group are highly polarized. Small optical densities lead to posterior probabilities that are often less than 0.001, while those samples with high antibody concentrations often have posterior probabilities higher than 0.99. Since there are very few observations whose probabilities are between 0.1 and 0.9, the curves for the three regimes remain mostly flat in this region.

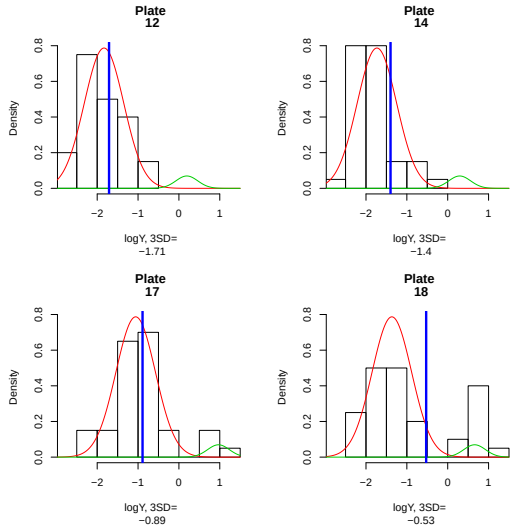

Figure 8: Histogram of the log of the optical densities, overlayed with a mixture density derived from our procedure. For each plate, the leftmost bell curve is the estimated Gaussian density curve for the negative subjects, the smaller rightmost bell curve is the density for the positives. The vertical line represents the (plate-specific) cutoff values derived from the 3SD rule ($\log(\bar{Y} + 3s_Y)$)

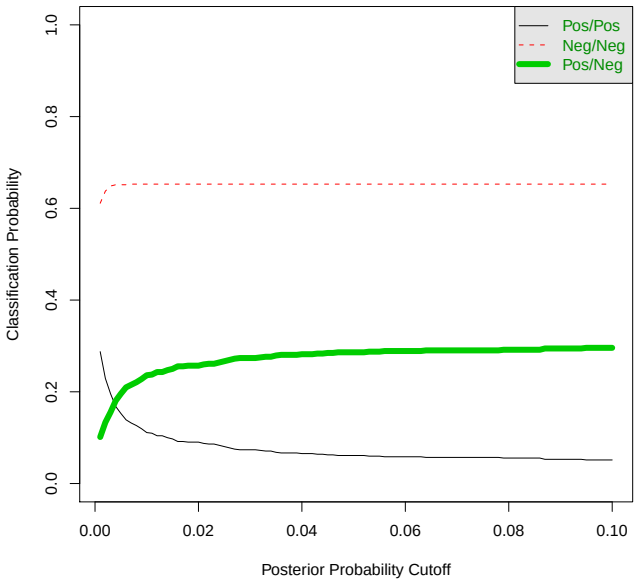

Figure 9: Fraction of patients who are diagnosed as positive under both methods, negative under both methods, and positive under the 3SD rule, and negative under our Bayesian model, as a function of posterior probability cutoff.

3.7. Discussion

Our goal has been to develop a robust model that allows one to perform more accurate diagnoses for diseases when testing samples on a series of different microtiter plates. Our classification method, based on computing posterior probabilities from a Bayesian mixture model, works quite well in simulation settings. In particular, it provides more plausible results than the conventional 3SD rule, and has improved sensitivity and specificity. Furthermore, the Bayesian approach has excellent frequentist properties. One can further improve the utility of this model by increasing the number of negative controls that are replicated on each plate.

Chapter 4

Robustness of our Bayesian Method

4.1. Introduction

While our method has promising results when the underlying data comes from our specified model, the most useful methodologies are those that produce reasonable predictions even when the data deviates slightly from the proposed model, or if there are mild violations of the model assumptions.

One of the model's key assumptions is that each unknown sample is randomly allocated to one of the microtiter plates. If unknown samples are put into the plates in sequential order (that is, filling up plate 1 before adding samples to plate 2, and so on), then the β parameters might not be valid measures of the underlying plate-effect, but rather be a measure of the overall antibody levels for that particular batch of samples, which may all share some confounding factor that systematically affects

their measured optical densities. For instance, the samples of a particular batch could all come from a community with a particularly severe outbreak of Chagas disease, or could all be prepared in the same laboratory. In the presence of such confounding factors, application of our model may lead to too many false negatives on plates with patients from highly infected villages, and too many false positives for samples placed on plates that come from fairly healthy regions. Randomizing samples to plates has the benefit of enabling us to make unbiased estimates of the plate effects, which in turn, lets us pool information across plates to make better classifications using our model.

4.2. Simulations and Model Performance under Model Violations

To assess our method's robustness, we will investigate both our model's long-run frequentist performance and its classification accuracy when one violates the assumption that individual antibody samples are randomly assigned to the various plates. Here, we consider the particular setting where half of the plates have a normal disease prevalence α_{low}, and half the plates come from communities with a somewhat higher rate of Chagas disease, α_{high}. In all these simulations, we will use 7 negative controls and 20 plates. We will use the same exact true parameters we used for the previous simulations, with the exception of α, which is now simply defined as the average of α_{low} and α_{high} (since half the plates are in each group).

Table 5 shows the frequentist behavior of a model in which the disease prevalence in the last ten plates is only slightly higher than the populations that are represented on the first ten plates, $\alpha_{low} = 0.125$, $\alpha_{high} = 0.225$. Again, we use 100 simulations.

Here, the coverage rates for the various parameters, including the new α, are still very close to 95%. The magnitudes of the MSE's and Bias estimates are slightly higher

Parameter	True Value	Coverage Rate	Mean Squared Error	Bias
μ_{neg}	-1.88	0.93	0.0456	0.0789
μ_{pos}	0.20	0.96	0.0842	0.0252
σ^2_{neg}	0.27	0.95	0.0010	0.0046
σ^2_{pos}	0.36	0.94	0.0553	0.0398
α	**0.175**	0.93	0.0013	0.0113
λ^2	0.17	0.95	0.0014	-0.0019
κ^2	0.10	0.93	0.0007	0.0065
ν^2	0.49	0.94	0.0703	0.1409

Table 5: Simulation Results for 20 plates, 7 Negative Controls, $\alpha_{high} = 0.225$

than they were for the simulation we ran earlier under the true model assuming homogeneity in α, but not alarmingly so.

Next we consider a more egregious violation of this assumption, where the heterogeneity of α is more extreme. Here, we keep $\alpha_{low} = 0.125$, but now let $\alpha_{high} = 0.325$, so that the rate of Chagas disease in the second subpopulation is more than double that of the first (Table 6). Here, the overall population prevalence is now $\alpha = 0.225$.

Parameter	True Value	Coverage Rate	Mean Squared Error	Bias
μ_{neg}	-1.88	0.91	0.0489	0.0730
μ_{pos}	0.20	0.94	0.0876	-0.0258
σ^2_{neg}	0.27	0.96	0.0012	0.0043
σ^2_{pos}	0.36	0.92	0.0433	0.0759
α	**0.225**	0.91	0.0024	0.0180
λ^2	0.17	0.94	0.0019	0.0003
κ^2	0.10	0.92	0.0 009	0.0040
ν^2	0.49	0.94	0.0707	0.1479

Table 6: Simulation Results for 20 plates, 7 Negative Controls, $\alpha_{high} = 0.325$

Now we see that the our coverage rates dip below 95 for many of our parameters. The most notable dropoffs are the coverages for σ^2_{pos} and α. Also, the magnitude of the MSE's and the biases have increased even more so than they did for the more mild heterogeneity in α. Thus, while the frequentist behavior of our model is robust to mild violations of the constant prevalence assumption, its performance suffers when

the heterogeneity is more severe.

Next, we see what effect the degree of the prevalence heterogeneity has on the performance of both our disease classification procedure, as well as on the 3SD Rule procedure. Figures 10 and 11 compare the sensitivities and specificities as a function of posterior probability cutoff in these two scenarios.

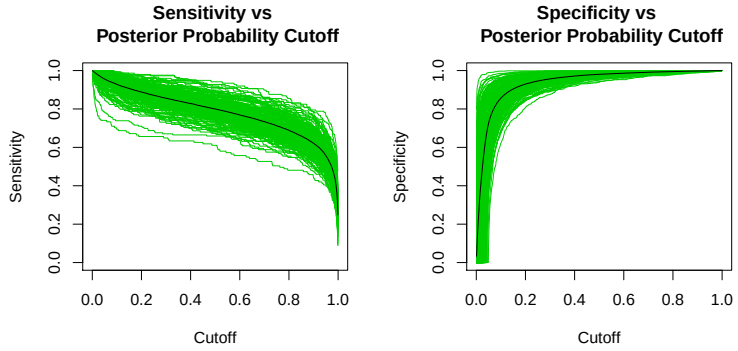

Figure 10: Sensitivity and Specificity for $\alpha_{high} = 0.225$. Green curves represent individual simulations, the black line reflects the average rates for all 1000 simulations.

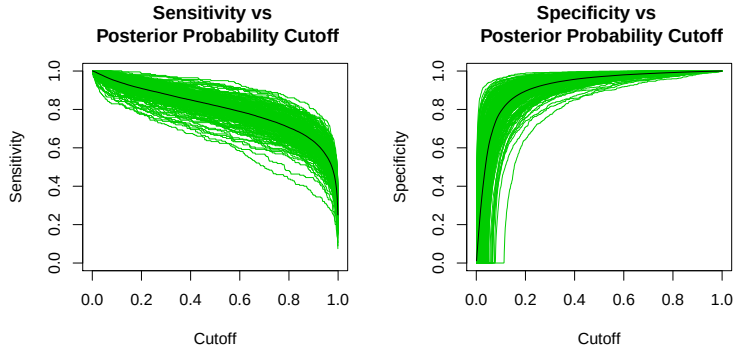

Figure 11: Sensitivity and Specificity for $\alpha_{high} = 0.325$. Green curves represent individual simulations, the black line reflects the average rates for all 1000 simulations.

For the setting with more extreme heterogeneity ($\alpha_{high} = 0.325$), the specificity at each probability cutoff seems to be noticeably lower than the corresponding specificities of both the homogeneous model and the one with $\alpha_{high} = 0.225$. The sensitivity profiles though appear quite similar for all three settings.

To compare our procedure to the 3SD rule, we once again analyze the ROC curves. As we can see in Figures 12 and 13, despite the violations of the model assumptions, our Bayesian classification procedure still out performs the 3SD Rule, as the blue X (denoting the average ROC values of the 3 SD rule across 1000 datasets) and the majority of the red circles (corresponding to individual simulations) lie to the lower right of the black line (Average Sensitivity and 1-Specificity for our procedure). However, the gap between the blue X and the black line is notably smaller than it was when α was in fact heterogeneous for all plates. This resonates with intuition. One of the main advantages of our Bayesian method over the 3SD rule was that we borrowed strength across all the plates, and under our model assumption, the distribution of optical densities (after controlling for the individual plate effects) were equivalent. When the samples on 50% of the plates come from a different subpopulation of patients, the advantage of sharing information is somewhat mitigated.

Note that the location of the X in our ROC curves doesn't noticeably change when we violate the homogeneity assumption in this manner. This also makes sense, since the 3SD rule, by definition, is an unpooled procedure, and only considers how a patient's sample compares to the negative controls on that given plate.

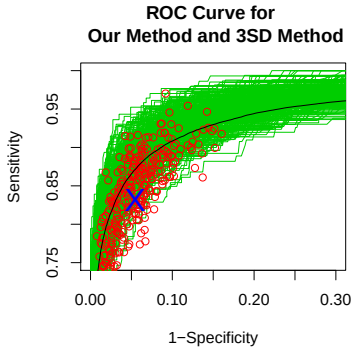

Figure 12: ROC plot for the two methods when $\alpha_{high} = 0.225$. Green curves represent simulated posterior ROC curves under our method. Black line reflects the average ROC curve across all 1000 simulations. Red dots correspond to the sensitivity/specificities of the individual simulations under the 3SD rule. The blue "X" is the average sens./spec. of the 3SD rule under all 1000 simulations.

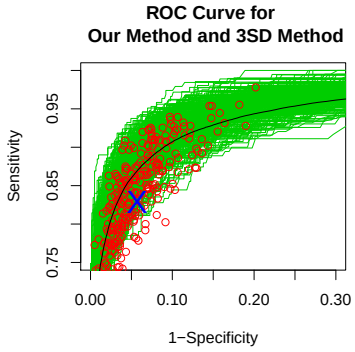

Figure 13: Same ROC plot for the two methods, this time for $\alpha_{high} = 0.325$.

Chapter 5

Validity of the Bayesian Mixture Model on Chagas Dataset

5.1. Model Checking

5.1.1. Basics of Model Checking

When fitting a complicated model to a dataset, how do we know whether or not the proposed model is "correct?" One reasonable answer is that a model is reasonably accurate if when one replicates data under the model's assumptions via simulation, these simulated datasets resemble the observed data in several key attributes. The amount that replicated datasets systematically deviate from what is observed can be thought of as an indication to the degree in which a model fails to capture underlying patterns in the true data.

There are two key issues that arise when using simulation to check a model.

1. How does one generate reference datasets that can be considered alternative versions of the real data? What do we even mean by a dataset replication?

2. What criteria does one use to compare the simulated and observed data? What numerical and graphical summaries can be used to measure a model's goodness of fit, or lack thereof?

5.1.2. Parametric Bootstrap Methods

In frequentist statistics, there are a plethora of methods used to establish validity of various models, including non-parametric bootstrap, residual resampling, and cross-validation technique. A particularly elegant approach involves creating null datasets by conditioning on minimal sufficient statistics under the null hypothesis, which removes the nuisance parameters, and is an example of an *exact test*. Langsrud (2005) demonstrates how to apply this principle of sufficiency to perform exact tests of a Gaussian linear model (Langsrud, 2005). Unfortunately, in more complicated models, it is often not possible to find a minimal sufficient statistic.

The classical method for creating null datasets that we will use here is the parametric bootstrap. The idea is to find a suitable point estimate $\hat{\theta}$ for the parameter vector θ and repeatedly sample artificial datasets $\mathbf{y}^*$ from $P(y^*|\hat{\theta})$. Unlike the Bayesian procedures, which look at variation of the parameters given a fixed data set (more on this in a moment), the parametric bootstrap looks at the variation among different datasets that have the same underlying model and nuisance parameters (Fushiki et al., 2004; Buja et al., 2009).

We numerically measure the discrepancy between a model and our observed data by defining an appropriate test statistic. A classical test statistic, $T(y)$ is a scalar

summary of the data that can be used as a benchmark when comparing the observed data to the series of reference datasets (since each y^* will have its own value of T).

5.1.3. Posterior Predictive Checks

Under the Bayesian scheme of posterior predictive checks, we generate alternate versions of the data by simulating from the posterior predictive distribution. Let y be the observed data, θ be the vector of underlying parameters (including the hyperparameters that govern the prior distributions of the hierarchical model), and y^{rep} be the hypothetical new data. Then the posterior predictive distribution of y^{rep} is defined as

$$P(y^{rep}|y) = \int P(y^{rep}|\theta)P(\theta|y)d\theta$$

To generate replicate datasets, we first sample several values of θ from its posterior distribution $P(\theta|y)$. Then for each value of θ, we generate a new dataset from the model, $P(y^{rep}|\theta)$. Notice that the reference distributions y^{rep} have two types of uncertainty built in; sampling uncertainty about the data given a particular value of θ, and uncertainty in the value of θ itself.

Just as in the parametric bootstrap approach, one then assess the goodness-of-fit of our model by an appropriate *Bayesian test statistic*. The Bayesian test statistic is very much like the frequentist analog described above, except that now, $T(y^{rep}, \theta)$ depends also on the drawn parameter θ in addition to the data.

5.1.4. P-Values

In classical statistics, the P-value is defined as the probability of obtaining a test statistic $T(y)$ as or more extreme than one observed in the real data, if the underlying

null hypothesis is true. Mathematically, this can be expressed as

$$p_C = Pr(T(y^{rep}) > T(y)|\theta)$$

Note that in the typical frequentist setting, the value of θ is either assigned a null-hypothesized value, or is set to a point estimate, such as its maximum likelihood value.

In a Bayesian model, we compare the observed data to the posterior predictive distribution (as opposed to the distribution of data under a single value of θ). Since replicated data sets are generated from different values of θ, the Bayesian test statistics can be functions of both the data and the unknown parameters. In the same spirit as the classical P-Value, we define the Bayesian predictive P-Value as

$$p_B = Pr(T(y^{rep}, \theta) > T(y, \theta)|y)$$

where the probability is taken over the joint distribution of the parameter and the data, $P(\theta, y^{rep}|y)$. One natural advantage of the Bayesian P-Values over their classical analogues is that the former does not require any special methods to handle nuisance parameters, as the test statistic is averaged over all possible values of θ (Gelman et al., 1996a).

5.2. Applying the Bootstrap Methods to Microtiter Data

5.2.1. Mechanics of the Bootstrap Method

In doing our check, our idea is based around creating "alternative" datasets that one would expect to see if our mixture model is correct. To apply the parametric

bootstrap, we use a reasonable point estimate for θ from which all subsequent datasets will be drawn. Here we use $\bar{\theta^{real}}$, which corresponds to the component-wise posterior mean of $P(\theta|Y^{real})$. We then generate alternative datasets $\mathbf{y}^*$ from $P(\mathbf{y}|\bar{\theta^{real}})$. Once we generate these alternate datasets, we use a goodness-of-fit test statistics $T(y, \theta)$ to compare our real data to the hypothetical data under our model. Our procedure is applied as follows:

1. Apply the MCMC procedure to our Chagas Disease microtiter plates, and use the output to compute $\bar{\theta}$ and $T^{real} = T(\mathbf{y^{real}}, \bar{\theta})$.

2. Generate N alternative datasets $\mathbf{y}^* = (y^*, y^{c*})$, each one created as follows:

   (a) Draw indicator variables $I^*_{ij} \sim Bernoulli(\bar{\alpha})$

   (b) For the plate/well pairs (i, j) where $I^*_{ij} = 0$, draw $y^*_{ij} \sim N(\bar{\mu_{neg}}, \bar{\sigma^2_{neg}})$.

   (c) For the plate/well pairs (i, j) where $I^*_{ij} = 1$, draw $y^*_{ij} \sim N(\bar{\mu_{pos}}, \bar{\sigma^2_{pos}})$.

   (d) For the plate/control pairs (i, j), draw $y^{c*}_{ij} \sim N(\bar{\tau_j}, \bar{\kappa^2})$.

3. Run the MCMC procedure on each $\mathbf{y}^* = (y^*, y^{c*})$, and compute $\bar{\theta^*}$.

4. For each alternative dataset, compute $T^* = T(\mathbf{y}^*, \bar{\theta^*})$. These values will serve as a null distribution for this statistic under our model.

5. Compute the bootstrapped P-Value. Since large values of the T imply large deviations from the model, the P-Value is defined as

$$P(T^* > T^{real}) = \frac{1}{N} \sum I_{T^* > T^{real}}$$

5.2.2. Bootstrapping Using the Bayesian Kolmogorov - Smirnov Statistic (BKS)

To define the first of out test statistics T, the Bayesian Kolmogorov-Smirnov Statistic (BKS), let $\mathbf{y}$ be the log of the optical measurements for the unknown subjects, and let $\bar{\theta}$ be the component-wise posterior mean vector of $(\mu_{neg}, \sigma^2_{neg}, \mu_{pos}, \sigma^2_{pos}, \alpha, \mathbf{b})$, corresponding to the means and variances of the negative and positive subgroups respectively, the mixing parameter denoting the fraction of the population that is positive, and the vector of (log) plate effects. Then for n unknown wells, we define the empirical distribution function $\hat{F}(y)$ as :

$$\hat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} I_{y_i \leq y}$$

The theoretical cumulative distribution function for the normal-normal mixture model, using the posterior means as plug-in estimates, is:

$$F(y, \bar{\theta}) = (1-\alpha)\Phi(\frac{y - \bar{\mu_{neg}} - \bar{b_i}}{\bar{\sigma_{neg}}}) + \alpha\Phi(\frac{y - \bar{\mu_{pos}} - \bar{b_i}}{\bar{\sigma_{pos}}})$$

In the spirit of the classical Kolmogorov-Smirnov Statistic, we define the BKS as:

$$BKS(\mathbf{y}, \bar{\theta}) = \sup_y |\hat{F}(y) - F(y, \bar{\theta})|$$

5.2.3. BKS Bootstrap Analysis on Our Chagas Disease Dataset

After fitting the Bayesian mixture model to our microtiter plate data, we used the posterior distribution of the parameters to create 200 alternative datasets that could theoretically have arisen under our model, using the bootstrap approach described

above. The bootstrapped distribution of the BKS statistics $B^* = BKS(\mathbf{y}^*, \bar{\theta^*})$ is given in Figure 14. The BKS statistic for the actual dataset, $B = BKS(\mathbf{y}, \bar{\theta})$, is 0.0321. Of our 200 bootstrapped values of B^*, 20 had values larger than 0.0321. Thus, the posterior P-Value for our dataset is 0.10. While this is a somewhat low number, a P-value of 10% is still within the reasonable realm of values one should get if the true data indeed arises from our proposed mixture model.

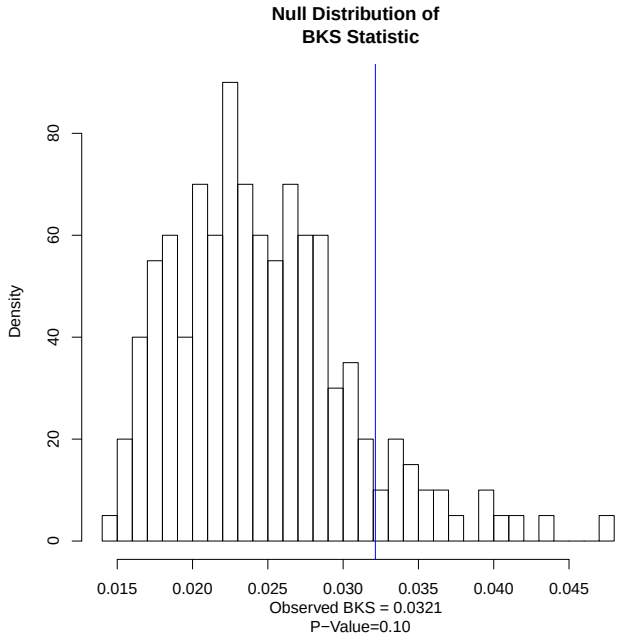

Figure 14: Null Distribution of the Bayesian Kolmogorov-Smirnov statistic. Blue line corresponds to B=0.0321, the BKS of the observed data. The P-Value is the area of the curve to the right of this, which is 0.10.

5.2.4. The Bayesian Anderson - Darling Statistic (BAD)

Another common frequentist method for testing the validity of the model is the Anderson-Darling Test (Anderson and Darling, 1952). It originally arose as a statistically more powerful procedure for testing departures from normality, by giving more weight to the tail ends of the distribution than towards the middle, hence better capturing the model's fit on the extremities of the data. In contrast to the classical

Kolmogorov-Smirnov statistic, whose distribution is related to the maximum of a Brownian bridge process regardless of the model being tested, the distribution of the Anderson-Darling statistic depends on the underlying distribution of interest.

To incorporate this test into our Bayesian framework, we once again define the theoretical CDF functions with our posterior means used as plug-in estimates,$F(y, \bar{\theta})$, as

$$F(y, \bar{\theta}) = (1-\alpha)\Phi(\frac{y - \bar{\mu_{neg}} - \bar{b_i}}{\bar{\sigma_{neg}}}) + \alpha\Phi(\frac{y - \bar{\mu_{pos}} - \bar{b_i}}{\bar{\sigma_{pos}}})$$

In the spirit of the frequentist test, we define the Bayesian Anderson-Darling statistic (BAD) as:

$$BAD(\mathbf{y}, \bar{\theta}) = -N - S$$

$$N = \text{ Number of observations}$$

$$S = \sum_{i=1}^{N} \frac{(2i-1)}{N} \left[\log(F(y_{(i)}, \bar{\theta})) + \log(1 - F(y_{(N+1-i)}, \bar{\theta}))\right]$$

where $y_{(i)}$ corresponds to the i^{th} order statistic of our observed log-densities y. Once again, large values of BAD denote large deviations from the underlying normal-normal mixture model.

To compute a bootstrapped P-Value for this new statistic, we follow the same exact procedure we used in the last section, except that we replace $B = BKS(\mathbf{y}, \bar{\theta})$ with $B = BAD(\mathbf{y}, \bar{\theta})$, and $B^* = BKS(\mathbf{y}^*, \bar{\theta^*})$ with $B^* = BAD(\mathbf{y}^*, \bar{\theta^*})$.

5.2.5. BAD Bootstrap Analysis on Our Chagas Disease Dataset

After fitting the Bayesian mixture model to our microtiter plate data, we again used the posterior distribution of the parameters to create 200 alternative datasets that could theoretically have arisen under our model, using the bootstrap approach described above. The posterior predictive distribution of the BAD statistics $B^* = BAD(\mathbf{y}^*, \bar{\theta^*})$ is given in Figure 15. The BAD statistic for the actual dataset, $B = BAD(\mathbf{y}, \bar{\theta})$, is 0.903. Of our 200 bootstrapped values of B^*, 156 had values larger than 0.903. Thus, the posterior P-Value for our dataset is 0.77, which is even more encouraging than the P-Value for the BKS. The high P-Value suggests that, for at least under this particular validation criterion, our dataset seems very plausible with our proposed model.

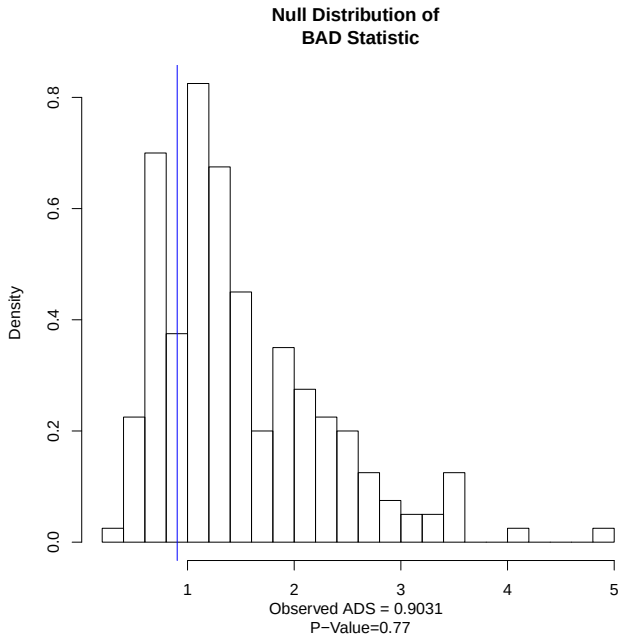

Figure 15: Null Distribution of the Bayesian Anderson-Darling statistic. Blue line corresponds to B=0.903, the BAD of the observed data. The P-Value is the area of the curve to the right of this, which is 0.77.

5.3. Bayesian Posterior Predictive Checks on Microtiter Data

5.3.1. Mechanics of Posterior Predictive Checks for Chagas Data

The recipe for the full Bayesian posterior predictive check is quite similar to the one used in our earlier bootstrap methods. The differences are subtle, but crucial (Upadhyay et al., 2000; Gelman, 2004).

1. Apply the MCMC procedure to our Chagas Disease microtiter plates to obtain draws from the posterior distribution of θ.

2. Sample N vectors of θ^{rep} from the posterior distribution $P(\theta|\mathbf{y^{real}})$.

3. For each θ^{rep}, generate an alternative dataset $\mathbf{y^{rep}}$ as follows:

   (a) Draw indicator variables $I_{ij}^{rep} \sim Bernoulli(\alpha^{rep})$

   (b) For the plate/well pairs (i, j) where $I_{ij}^{rep} = 0$, draw $y_{ij}^{rep} \sim N(\mu_{neg}^{rep}, \sigma_{neg}^{2rep})$.

   (c) For the plate/well pairs (i, j) where $I_{ij}^{rep} = 1$, draw $y_{ij}^{rep} \sim N(\mu_{pos}^{rep}, \sigma_{pos}^{2rep})$.

   (d) For this analysis, we do not need to generate replicate data for the negative controls.

4. For each value of θ^{rep}, use our test statistic to compute $T^{real} = T(\mathbf{y^{real}}, \theta^{rep})$

5. For each value of θ^{rep} and corresponding alternate dataset $\mathbf{y^{rep}}$, compute our test statistic $T^{rep} = T(\mathbf{y^{rep}}, \theta^{rep})$

6. Compute the posterior P-Value. Since the test statistics we are using are measures of discrepancies, an abundance of values of T^{real} that exceed T^{rep} would

indicate a poorly fitting model. Thus the posterior P-Value is defined as

$$P(T^{rep} > T^{real}) = P(T^{rep} - T^{real} > 0)$$

$$= \frac{1}{N} \sum I_{T^{rep} > T^{real}}$$

5.3.2. Bayesian Inference with BKS

Using the output from our Bayesian mixture model on our microtiter plate data, we sampled 200 draws of the vector θ from the posterior distribution $P(\theta|\mathbf{y^{real}})$, and then used our model to generate a replicate dataset for each draw of θ. For draws $k = 1, 2, \ldots, 200$, we computed $BKS(\mathbf{y^{real}}, \theta_k)$ and $BKS(\mathbf{y_k^{rep}}, \theta_k)$, where once again,

$$BKS(\mathbf{y}, \theta) = \sup_y |\hat{F}(y) - F(y, \theta)|$$

The posterior P-Value is the fraction of times $BKS(\mathbf{y^{rep}}, \theta)$ exceeds $BKS(\mathbf{y}, \theta)$. In figure 16, the P-value corresponds to the area of the histogram of $T^{rep} - T^{real}$ that lies to the right of 0, and equivalently, the fraction of points that lie above the 45-degree line in the scatterplot of T^{rep} vs T^{real}. For this dataset, the posterior P-Value came to 0.355, which is a decent result, and is another indication that our Chagas data does not deviate too severely from our proposed mixture model.

5.3.3. Bayesian Inference with BAD

We apply an analogous procedure in using the Anderson-Darling statistic as the discrepancy measure for our posterior predictive check. We once again sampled 200 draws of the vector θ from the posterior distribution $P(\theta|\mathbf{y^{real}})$, and then generate respective alternative datasets for each draw of θ. For draws $k = 1, 2, \ldots, 200$, we

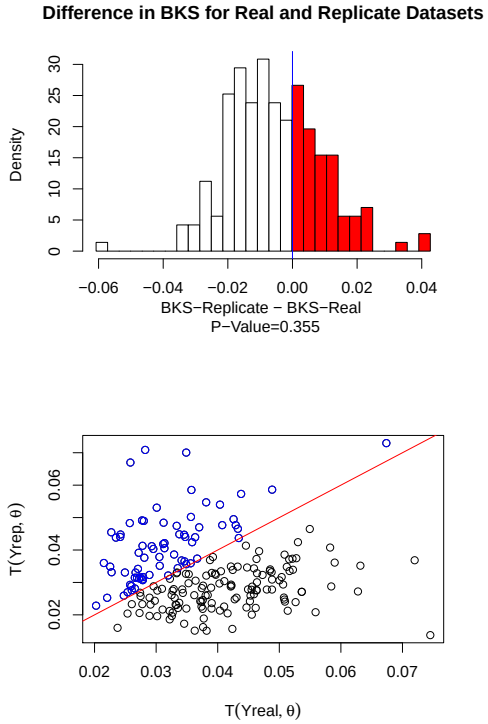

Figure 16: Top figure shows a histogram of $T^{rep} - T^{real}$. The area of the graph to the right of 0 (shaded in red) corresponds to a P-Value of 0.355. Bottom figure shows the scatterplot of T^{rep} vs T^{real} for the 200 draws of θ. The fraction of dots above the red 45-degree line corresponds to this same P-Value.

computed $BAD(\mathbf{y^{real}}, \theta_k)$ and $BAD(\mathbf{y_k^{rep}}, \theta_k)$, using the same formula as before, only replacing the posterior mean $\bar{\theta}$ with the individual posterior draws of θ. The posterior P-Value is again the fraction of times $BAD(\mathbf{y^{rep}}, \theta)$ exceeds $BAD(\mathbf{y}, \theta)$, and this can be seen graphically in both the histogram of $T^{rep} - T^{real}$, and the scatterplot of these two sets of statistics (Figure 17). For the BAD posterior predictive check, the P-value comes out to 0.275, another encouraging sign that our model is performing well for the Chagas disease data.

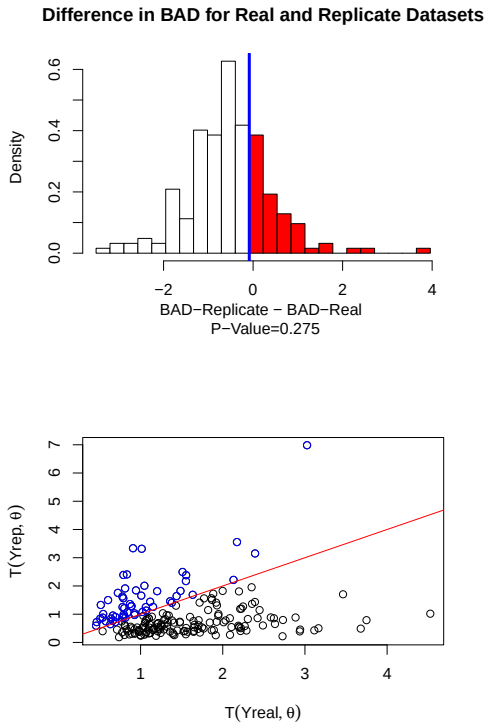

Figure 17: Top figure shows a histogram of $T^{rep} - T^{real}$. The area of the graph to the right of 0 (shaded in red) corresponds to a P-Value of 0.275. Bottom figure shows the scatterplot of T^{rep} vs T^{real} for the 200 draws of θ. The fraction of dots above the red 45-degree line corresponds to this same P-Value.

Chapter 6

Conclusions and Further Directions

In this dissertation, we have developed a Bayesian hierarchical model that allows one to perform more accurate diagnoses for diseases when testing samples on a series of different microtiter plates. Our classification method, based on computing posterior probabilities from a Bayesian mixture model, works quite well in simulation settings. In particular, it provides more plausible results than the conventional 3SD rule, and has improved sensitivity and specificity. Furthermore, the Bayesian approach has excellent frequentist properties, and for this particular data set, provides an estimate of the overall prevalence rate of Chagas disease that is consistent with other studies. Our method is fairly robust to certain violations of our model's assumptions. Namely, if there are systematic differences in the underlying patients that are assigned to each plate, this Bayesian approach still outperforms the current crude classification procedures. Finally, we demonstrate how our procedure can be applied to a real life example involving the diagnosis of Chagas disease.

There are several avenues one could pursue to expand on our research. For one, we've

only tested our model's robustness with regards to prevalence heterogeneity. One could consider testing our method under other assumption violations. For instance, we have assumed that the optical densities of a samples are conditionally independent, once one controls for the plate effect. However, depending on the opaqueness or heterogeneity in the glass of the microtiter plates, it is plausible that there are spacial correlations between optical densities in wells that are in the same row or column.

One might also want to consider our model's performance when the optical densities are not lognormal, but rather come from another skewed distribution. The Box-Cox transformations (of which our lognormal setting is a special case) provide a whole family of distributions that could plausibly describe a mixture of positive and negative patients. If our model is not robust to this discrepancy, one solution would be to incorporate the Box-Cox transformation parameter λ directly into our Bayesian model, instead of fixing its value at $\lambda = 1$ as we do now (Sweeting, 1984; Lee et al., 2005).

Another extension would be to subject our Chagas dataset to more rigorous model checking. Here, the BKS and BAD statistics introduced in the thesis only check one particular facet of the data, namely the appropriateness of the lognormal mixture assumption. This does not, however test some of the more complex features of our model, such as the interrelationship between the negative controls and the unknown samples on a particular plate. Finally, one may want to make use of the optical densities for the positive controls that are used on each plate, but not included in our current analysis.

Appendix A

R Package for Running MCMC Mixture Model

The main function of our package, ImmunoassayMixture, runs the MCMC procedure above, and produces draws from the joint posterior distribution of all the relevant parameters of our model.

A.1. Usage

ImmunoassayMixture(Y,Plate,YcontrolMat,withControls=F,...)

A.2. Arguments

- **Y**: A numerical vector of the raw (unlogged) optical densities of all the unknown samples.
- **Plate**: An integer-valued vector of same length of **Y** denoting the plate index on which each observation is located.

- **YcontrolMat**: A numerical matrix denoting the raw optical densities of the negative controls. For instance, **YcontrolMat[2,3]** corresponds to the measurement on the second plate, of the third negative control subject. Only used if **withControls=T**.

- **withControls**: A boolean input denoting whether or not your model makes use of negative control data. By default, this equals False (F).

- **NUMITERS**: An integer, corresponding to the number of iterations one would like to carry out in the simulation. Default is set to 6,500.

- **BURN**: Integer denoting the number of preliminary iterations to be discarded from the output matrix. Default is 0.

- **THIN**: Integer denoting the "thinning factor" of the output matrix. For example, if **THIN=5**, we keep only every fifth iteration post burn-in. Default is 1, meaning we keep all the iterations.

- **CountFrequency**: An integer denoting how often to print the "status" of the MCMC chain. If **CountFrequency=20**, then at every 20th iteration, the function will print out the iteration number, and the total cumulative time (in seconds) that has elapsed. By default, this is set to 100.

- **Init.mu**: Two-component numerical vector corresponding to the initial values of μ_{neg}, μ_{pos}, the means of the logged optical densities of the negative and positive subgroups respectively. By default, this is set to (0,0).

- **Init.sigSq**: Two-component numerical vector corresponding to the initial values of σ^2_{neg}, σ^2_{pos}, the means of the logged optical densities of the negative and

positive subgroups respectively. By default, this is set to (1,0.5). Note that when **withContols=T**, $\sigma^2_{neg} = \lambda^2 + \kappa^2$, so the initial values for these three parameters must have this relationship.

- **Init.alpha**: Initial value of α, the fraction of the underlying population that is positive. Default is 0.06.

- **Init.Beta**: Vector of length *NumPlates*, corresponding to initial values for the multiplicative plate effects β(unlogged). By default, the initial plate effects are set to 1 (and hence $\log(\beta_i) = 0$ for all plates).

- **Init.nusq**: Initial value of ν^2, the variance of the log of the plate effects. Default value is 0.25

- **Init.lambdasq**: Initial value for λ^2, the variance of the means of the log-measurements of the negative control subjects. Should follow the constraint $\sigma^2_{neg} = \lambda^2 + \kappa^2$. Initial value is 1.2.

- **Init.kappasq**: Initial value for κ^2, the within-person variance of the repeated logged measurements of the control subjects, after controlling for the plate effects. Should follow the constraint $\sigma^2_{neg} = \lambda^2 + \kappa^2$. Initial value is 0.8.

- **Init.nusq**: Initial value for ν^2, the variance of the (logged) plate effects. Initial value is 0.25.

- **Prior.Nusq**: A two-component vector corresponding to the shape and rate hyperparameters of the inverse-gamma prior distribution of ν^2. By default, this is set to (0,0), corresponding to a flat (uninformative) prior distribution.

A.3. Output

- **Output.Params**: The full matrix of outputs corresponding the posterior draws of all the key parameters of the model. The number of rows depends on **NUMITERS**,**BURN**, and **THIN**. The number of columns depends on the number of plates, and if applicable, the number of negative controls. A few caveats:

  - The output variable **Output.Params** is global. Thus, it is recommended that at the conclusion of each run of the MCMC, one saves **Output.Params** under a new name (or in a separate file), since any subsequent usage of the function will ultimately overwrite **Output.Params**.

  - This variable is only defined/redefined at the very end of the MCMC procedure. If one interrupts the function, **Output.Params** will simply be the last output matrix (or undefined, if the function has never run to completion).

- **Output.Params.Temp** This is a globally defined matrix similar to **Output.Params**, with the main differences being that it is redefined at each iteration, and that no iterations are discarded via burning or thinning. Thus, if one interrupts the MCMC part way through (or if it crashes), one can look at **Output.Params.Temp** to observe the current progress of the chains at the time of the interruption. **Output.Params.Temp** can therefore be used as a debugging tool.

BIBLIOGRAPHY

T. W. Anderson and D. A. Darling. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statistics*, 23:193–212, 1952. ISSN 0003-4851.

C. Bern, M. Verastegui, R. H. Gilman, C. Lafuente, G. Galdos-Cardenas, M. Calderon, J. Pacori, M. Del Carmen Abastoflor, H. Aparicio, M. F. Brady, L. Ferrufino, N. Angulo, S. Marcus, C. Sterling, and J. H. Maguire. Congenital Trypanosoma cruzi transmission in Santa Cruz, Bolivia. *Clin. Infect. Dis.*, 49(11): 1667–1674, Dec 2009.

A. Buja, D. Cook, H. Hofmann, M. Lawrence, E. Lee, D. Swayne, and H. Wickham. Statistical inference for exploratory data analysis and model diagnostics. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 367(1906): 4361–4383, 2009. ISSN 1364-503X. doi: 10.1098/rsta.2009.0120. URL `http://dx.doi.org/10.1098/rsta.2009.0120`. With electronic supplementary materials available online.

G. Casella and E. George. Explaining the Gibbs sampler,. *Amer. Statist.*, 46:167–174, 1992.

S. Chib and E. Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335, Nov. 1995. URL `http://www.jstor.org/stable/2684568`.

S. Chib and E. Greenberg. Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory*, 12(3):409–431, 1996. ISSN 0266-4666. doi: 10.1017/S0266466600006794. URL `http://dx.doi.org/10.1017/S0266466600006794`.

Y. Choi, W. Johnson, M. Collins, and I. Gardner. Bayesian inferences for receiver operating characteristic curves in the absence of a gold standard. *Journal of Agricultural, Biological, and Environmental Statistics*, 11:210–229, 2006. ISSN 1085-7117. URL `http://dx.doi.org/10.1198/108571106X110883`. 10.1198/108571106X110883.

S. Delgado, R. Castillo Neyra, V. Quispe Machaca, J. Ancca Jurez, L. Chou Chu, M. Verastegui, G. Moscoso Apaza, C. Bocngel, A. Tustin, C. Sterling, A. Comrie, C. Nquira, J. Cornejo Del Carpio, R. Gilman, C. Bern, and M. Levy. A history of chagas disease transmission, control, and re-emergence in peri-rural la joya, peru. *PloS Neglected Tropical Diseases, 5(2): e970.*

T. Fushiki, F. Komaki, and K. Aihara. On parametric bootstrapping and

Bayesian prediction. *Scand. J. Statist.*, 31(3):403–416, 2004. ISSN 0303-6898. doi: 10.1111/j.1467-9469.2004.02_127.x. URL http://dx.doi.org/10.1111/j.1467-9469.2004.02_127.x.

A. Gelman. Iterative and non-iterative simulation algorithms. *The American Statistician*, pages 433–438, May 1992.

A. Gelman. Exploratory data analysis for complex models. *J. Comput. Graph. Statist.*, 13(4):755–787, 2004. ISSN 1061-8600. doi: 10.1198/106186004X11435. URL http://dx.doi.org/10.1198/106186004X11435. With a discussion by Andreas Buja and a rejoinder by the author.

A. Gelman, X. Meng, and H. Stern. Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica*, 6(4):733–807, 1996a. ISSN 1017-0405. With comments and a rejoinder by the authors.

A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient Metropolis jumping rules. In *Bayesian statistics, 5 (Alicante, 1994)*, Oxford Sci. Publ., pages 599–607. Oxford Univ. Press, New York, 1996b.

A. Gelman, G. L. Chew, and M. Shnaidman. Bayesian analysis of serial dilution assays. *Biometrics*, 60:407–417, Jun 2004.

S. Geman. Stochastic relaxation methods for image restoration and expert systems. In *Maximum-entropy and Bayesian methods in science and engineering, Vol. 2 (Laramie, WY, 1985 and Seattle, WA, 1986/1987)*, Fund. Theories Phys., pages 265–311. Kluwer Acad. Publ., Dordrecht, 1988.

J. Geweke. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57(6):1317–1339, 1989. ISSN 0012-9682. doi: 10.2307/1913710. URL http://dx.doi.org/10.2307/1913710.

M. Greiner, D. Pfeiffer, and R. D. Smith. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev. Vet. Med.*, 45: 23–41, May 2000.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, Apr 1970.

K. M. Higgins, M. Davidian, G. Chew, and H. Burge. The effect of serial dilution error on calibration inference in immunoassay. *Biometrics*, 54:19–32, Mar 1998.

J. P. Hobert, V. Roy, and C. Robert. Improving the Convergence Properties of the

Data Augmentation Method with an Application to Bayesian Mixture Modeling. *Statist. Sci.*, 26(3):332–351, 2011.

A. Irion, H. P. Beck, and T. Smith. Assessment of positivity in immuno-assays with variability in background measurements: a new approach applied to the antibody response to Plasmodium falciparum MSP2. *J. Immunol. Methods*, 259:111–118, Jan 2002.

L. V. Kirchhoff, L. M. Weiss, M. Wittner, and H. B. Tanowitz. Parasitic diseases of the heart. *Front. Biosci.*, 9:706–723, Jan 2004.

K. M. Kurkjian, L. E. Vaz, R. Haque, C. Cetre-Sossah, S. Akhter, S. Roy, F. Steurer, J. Amann, M. Ali, R. Chowdhury, Y. Wagatsuma, J. Williamson, S. Crawford, R. F. Breiman, J. H. Maguire, C. Bern, and W. E. Secor. Application of an improved method for the recombinant k 39 enzyme-linked immunosorbent assay to detect visceral leishmaniasis disease and infection in Bangladesh. *Clin. Diagn. Lab. Immunol.*, 12:1410–1415, Dec 2005.

Ø. Langsrud. Rotation tests. *Stat. Comput.*, 15(1):53–60, 2005. ISSN 0960-3174. doi: 10.1007/s11222-005-4789-5. URL `http://dx.doi.org/10.1007/s11222-005-4789-5`.

J. Lee, T. Lin, K. Lee, and Y. Hsu. Bayesian analysis of Box-Cox transformed linear mixed models with ARMA(p, q) dependence. *J. Statist. Plann. Inference*, 133(2):435–451, 2005. ISSN 0378-3758. doi: 10.1016/j.jspi.2004.03.015. URL `http://dx.doi.org/10.1016/j.jspi.2004.03.015`.

M. Levy, D. Small, D. Vilhena, N. Bowman, V. Kawai, J. Cornejo del Carpio, E. Cordova-Benzaquen, R. Gilman, C. Bern, and J. Plotkin. Retracing micro-epidemics of Chagas disease using epicenter regression. *PLoS Comput. Biol.*, 7 (9):e1002146, 9, 2011. ISSN 1553-734X. doi: 10.1371/journal.pcbi.1002146. URL `http://dx.doi.org/10.1371/journal.pcbi.1002146`.

D. Limmathurotsakul, N. Chantratita, N. Teerawattanasook, K. Piriyagitpaiboon, A. Thanwisai, V. Wuthiekanun, N. P. Day, B. Cooper, and S. J. Peacock. Enzyme-linked immunosorbent assay for the diagnosis of melioidosis: better than we thought. *Clin. Infect. Dis.*, 52:1024–1028, Apr 2011.

B. G. Lindsay and K. Roeder. Uniqueness of estimation and identifiability in mixture models,. *Canad. J. Statist.*, 21:139–147, 1993.

N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087, 1953.

C. Morris. Parametric empirical Bayes inference: theory and applications. *J. Amer. Statist. Assoc.*, 78(381):47–65, 1983. ISSN 0162-1459. With discussion.

L. H. Moulton, F. C. Curriero, and P. F. Barroso. Mixture models for quantitative HIV RNA data. *Stat Methods Med Res*, 11:317–325, Aug 2002.

S. S. Nielsen, C. Gronbaek, J. F. Agger, and H. Houe. Maximum-likelihood estimation of sensitivity and specificity of ELISAs and faecal culture for diagnosis of paratuberculosis. *Prev. Vet. Med.*, 53:191–204, Mar 2002.

M. Opsteegh, P. Teunis, M. Mensink, L. Zuchner, A. Titilincu, M. Langelaar, and J. van der Giessen. Evaluation of ELISA test characteristics and estimation of Toxoplasma gondii seroprevalence in Dutch sheep using mixture models. *Prev. Vet. Med.*, 96:232–240, Sep 2010.

D. B. Rubin. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.*, 12(4):1151–1172, 1984. ISSN 0090-5364. doi: 10.1214/aos/1176346785. URL `http://dx.doi.org/10.1214/aos/1176346785`.

T. Sweeting. On the choice of prior distribution for the Box-Cox transformed linear model. *Biometrika*, 71(1):127–134, 1984. ISSN 0006-3444. doi: 10.1093/biomet/71.1.127. URL `http://dx.doi.org/10.1093/biomet/71.1.127`.

R. L. Tarleton, R. Reithinger, J. A. Urbina, U. Kitron, and R. E. Gurtler. The challenges of Chagas Disease– grim outlook or glimmer of hope. *PLoS Med.*, 4(12): e332, Dec 2007.

L. Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22 (4):1701–1762, 1994. ISSN 0090-5364. doi: 10.1214/aos/1176325750. URL `http://dx.doi.org/10.1214/aos/1176325750`. With discussion and a rejoinder by the author.

A. Tustin, D. Small, S. Delgado, R. Castillo Neyra, M. Verastegui, J. Ancca Juarez, M. Quispe Machaca, R. Gilman, C. Bern, and M. Levy. Use of individual level covariates to improve latent class analysis of Trypanosoma cruzi diagnostic tests. *Epidemiologic Methods.*, To appear.

S. K. Upadhyay, N. Vasishta, and A. F. M. Smith. Bayes inference in life testing and reliability via Markov chain Monte Carlo simulation. *Sankhyā Ser. A*, 62(2): 203–222, 2000. ISSN 0581-572X.

C. Wang, B. Turnbull, Y. Grhn, and S. Nielsen. Nonparametric estimation of roc curves based on bayesian models when the true disease state is unknown. *Journal of Agricultural, Biological, and Environmental Statistics*, 12:128–146,

2007. ISSN 1085-7117. URL http://dx.doi.org/10.1198/108571107X178095. 10.1198/108571107X178095.