UNIVERSITY OF CALIFORNIA
RIVERSIDE

Sequential Probability Ratio Tests for Generalized Linear Mixed Models

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Judy Xiang Li

August 2010

Dissertation Committee:

Dr. Daniel R. Jeske, Chairperson
Dr. Barry C. Arnold
Dr. Robert F. Luck

The Dissertation of Judy Xiang Li is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

Though only my name appears on the cover of this dissertation, a great many people have contributed to its production. My sincere gratitude goes to all those people who have made this dissertation possible:

Dr. Daniel R. Jeske, doctoral advisor and chair of my committee, for his patient guidance, brilliant advises, endless encouragement, and infectious enthusiasm throughout my Ph. D. studies. I also want to thank him for providing me the great opportunity to work at the statistical consulting collaboratory where I obtain a lot first hand knowledge of how to do statistical consulting, and how to solve applied problems. He is the best advisor ever!

Dr. Barry C. Arnold, dissertation committee member, for participating in my doctoral committee and the distinguished teaching of classes that build up my theoretical statistics and probability knowledge.

Dr. Robert F. Luck, dissertation committee member, for introducing me to the field of pest management and being readily available when needed.

Dr. Jeffrey A. Klein M.D., summer internship mentor and collaborator, for introducing me to the field of clinical trial studies, and providing medical insights on the design of clinical motivating example in this dissertation.

Jesús R. Lara, Dr. Mark Hoddle, Jane Delahoyde, Rick Shade, Ruth Amrich and Allison Bistline for helping to acquire necessary data for the analysis included in this dissertation.

DEDICATIONS

I sincerely dedicate this dissertation to:

my family, for their generous love, endless support and encouragement ever since I was born.

all my friends those who care about me and wish me the best.

Thank you all for being constant in my life and sharing the laughter and tears of my journey.

ABSTRACT OF THE DISSERTATION

Sequential Probability Ratio Tests
for Generalized Linear Mixed Models

by

Judy Xiang Li

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, August 2010
Dr. Daniel R. Jeske, Chairperson

The sequential probability ratio test (SPRT) is a hypothesis testing procedure, which evaluates data as it is collected. The original SPRT was developed by Wald for one-parameter families of distributions and later extended by Bartlett to account for nuisance parameters. We adapt Bartlett's SPRT to Generalized Linear Mixed Models (GLMM), in which the observations are non-identically and non-independently distributed and illustrate the approach taken with two applications. In the first application, we incorporate a Poisson GLMM into sequential procedure to design a multicenter randomized clinical trial that compares two preventive treatments for surgical site infections. In the second application, we incorporate a Negative Binomial spatial GLMM into sequential procedure to design a pest assessment protocol. We also consider a generative spatial model in the context of sequential procedures as an alternative to spatial GLMMs.

TABLE OF CONTENTS

Introduction…….

Chapter 1: Introduction of Sequential Probability Ratio Tests..…….…..……....

1.1. Wald's Sequential Probability Ratio Test (Wald's SPRT)…….....

1.1.1. Wald's SPRT with IID Data…….…...

1.1.2. Wald's SPRT with Non-IID Data……....

1.2. Bartlett's Sequential Probability Ratio Test (Bartlett's SPRT)…….....

1.2.1. Bartlett's SPRT with IID Data….……....

1.2.2. Bartlett's SPRT for GLMs……...

Chapter 2: Design of Sequential Hypothesis Testing Methodology for Poisson GLMM……....

2.1. Motivating Clinical Example……...

2.2. Fixed Sample Size Test……...

2.3. Wald's SPRT…….……..……...

2.4. Proposed SPRT for Poisson GLMM……...……...

2.4.1. GLMM-Bartlett SPRT……...……..

2.4.2. The Consistency of eBLP(τ_i)……..……....

2.4.3. Performance of GLMM-Bartlett SPRT…..……..……...

2.4.4. Computational Approximation……..……...……....

Chapter 3: Design of Sequential Hypothesis Testing Methodology for Spatial GLMM……....

3.1. Motivating Entomology Example……...……...

3.1.1. *O. Perseae* Data……..……..

3.1.2. Analysis of *O. Perseae* Data with Spatial GLMM…….....

3.2. Proposed Sampling Methodology……....

3.3. Illustration Example on Leafhopper Count Data…….

3.3.1. First Occasion Sampling……...

3.3.2. Subsequent Occasion Sequential Sampling…..……....

3.3.3. Fixed Sample Size Test…….

Chapter 4: Design of Sequential Hypothesis Testing Methodology for Generative Spatial Model……..

4.1. Poisson Generative Spatial Model……..……...….

4.1.1. Model Formulation…….….

4.1.2. Wald's SPRT for Poisson Generative Spatial Model…….....

4.1.3. Spatial Sampling Plan……..…....

4.1.4. OC and ASN Curves for Poisson Generative Spatial Model……...……...…

4.2. Negative Binomial Generative Spatial Model..…….….....

4.2.1. Model Formulation…….…….....

4.2.2. Transformed Bartlett's SPRT for NB Generative Spatial Model……....

4.2.3. OC and ASN Curves for NB Generative Spatial Model ……...……..

Chapter 5: Future Work……....

5.1. Extension of Sequential GLMM to Other Distributions ……....

5.2. Application of GLMM-Bartlett SPRT to Spatial GLMM…….....

5.2.1. Pseudo Likelihood Approach……...

5.2.2. Monte Carlo EM Algorithm…….……...….

5.3. Extension of Sequential GLMMs to Group Sequential Contexts……...….

Bibliography…….....

LIST OF FIGURES

Figure 1. ASN Curves for $H_0 : \mu = 1.9$ vs $H_0 : \mu = 2.3$ with Different Correlations of AR(1) Model

Figure 2. OC Curves for $H_0 : \mu = 1.9$ vs $H_0 : \mu = 2.3$ with Different Correlations of AR(1) Model

Figure 3. OC Curve for Wald SPRT Test of $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ with $(\zeta_0, \sigma) = (2.5, 0.2)$ and $(\zeta_0, \sigma) = (2.5, 0.4)$ and nominal type-1 and type-2 error rates of 0.001

Figure 4. ASN Curve for Wald SPRT and Fixed Sample Size Test of $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ with $(\zeta_0, \sigma) = (2.5, 0.2)$ and nominal type-1 and type-2 error rates of 0.001

Figure 5. ASN Curve for GLMM-Bartlett SPRT, Wald's SPRT and Fixed Sample Size Test of $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ with $(\zeta_0, \sigma) = (2.5, 0.2)$ and nominal type-1 and type-2 error rates of 0.01

Figure 6. Comparison of Exact and Approximate Likelihood

Figure 7. 90% Tolerance Bands (Estimated) for Conditional Negative Binomial Models

Figure 8. Sample Boundaries and Hypothetical Sequential Sample Path for Leafhopper Pest Counts

Figure 9. ASN Curve for Proposed Subsequent Occasion Sequential Procedure and Fixed Sample Size Test of $H_0 : \mu = 1.9$ vs $H_1 : \mu = 2.3$ with $\kappa = 6$ and nominal type-1 and type-2 errors rates of 0.1 on Illustrated Leafhopper Count Data

Figure 10. Hypothetical Block of 64 Trees

Figure 11. Heat Map of Hypothetical Pest Count Data based on Poisson Generative Spatial Model

Figure 12. Comparison of Exact and Approximated Likelihood for Poisson Generative Spatial Model

Figure 13. ASN Curves for Wald's SPRT for Poisson Generative Spatial Model ($H_0 : \lambda = 7$ versus $H_1 : \lambda = 10$)

Figure 14. Heat Map of Hypothetical Pest Count Data based on Negative Binomial Generative Spatial Model

Figure 15. ANS Curves for Transformed Bartlett's SPRT with Negative Binomial Generative Spatial Model and $\kappa = 3$

LIST OF TABLES

Table 1. Required sample sizes *n* for each group, to provide type-1 and type-2 error rates of 0.01 when using the NP test of $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$

Table 2. OC and ASN properties of Wald SPRT for $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ when nominal type-1 and type-2 error rates are 0.01

Table 3. OC and ASN properties of Proposed GLMM-Bartlett's SPRT for $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ when nominal type-1 and type-2 error rates are 0.01

Table 4. OC and ASN properties comparison of GLMM-Bartlett's SPRT with Quadrature rule and approximate likelihood for $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ when nominal type-1 and type-2 error rates are 0.01 and (ζ_0, σ) = (2.5, 0.2)

Table 5. Analysis of Variance Covariance Parameter Estimates for *O. Perseae* Count Data

Table 6. Analysis of Fixed Effects for *O. Perseae* Count Data

Table 7. Hypothetical First Occasion Initial Sample Leafhopper Counts Data

Table 8. Analysis of Fixed Effects for Hypothetical Leafhopper Count Data

Table 9. Variance-Covariance Parameter Values for Leafhopper Initial Sample Data

Table 10. Calculation of Starting Sample Size for Subsequent Occasion Sequential Hypothesis Test for Leafhopper Pest Counts

Table 11. OC and ASN properties for $H_0 : \mu_0 = 1.9$ vs $H_1 : \mu_1 = 2.3$ with $\kappa = 6.04$ and nominal type-1 and type-2 error rates are 0.1

Table 12. OC and ASN properties for Poisson Generative Spatial Model Testing $H_0 : \mu = 7$ versus $H_1 : \mu = 10$ when $\alpha = \beta = 0.1$.

Table 13. OC and ASN properties for NB Generative Spatial Model Testing $H_0 : \mu = 7$ versus $H_1 : \mu = 10$ when $\kappa = 3$ and $\alpha = \beta = 0.1$.

INTRODUCTION

It is well known that sequential hypothesis test plans can realize appreciable cost savings compared to fixed sample size test plans. The first sequential hypothesis test plan was developed by Wald for one-parameter families of distributions and later extended by Bartlett to handle the case of nuisance parameters. However, Bartlett's development assumed independent and identically distributed observations. In this dissertation, we adapt Bartlett's SPRT to the context of Generalized Linear Mixed Models (GLMM), in which the observations are non-identically and non-independently distributed and illustrate the approach with two applications.

The rest of this dissertation is organized as follows:

In Chapter 1, we briefly introduce the classic sequential methods of Wald's SPRT and Bartlett's SPRT. We also show that Bartlett's SPRT can be applied to generalized linear model (GLM) contexts.

In chapter 2, we are interested in the design of a multicenter randomized clinical trial that compares two preventive treatments for surgical site infections. In particular, we design a sequential probability ratio test methodology in the context of a Poisson GLMM. We validate the proposed SPRT with a simulation study that includes a comparison to a fixed sample size test and the Wald's SPRT.

In chapter 3, we adapt a Negative Binomial GLMM-based sequential procedure for use in pest management where it is common for the data to be counts that exhibit spatial correlations. We illustrate the existence of spatial correlations in pest counts by

analyzing the spatial structure in a data set for a foliar-feeding mite pest. Spatial correlations should be accounted for when designing a sequential hypothesis test plan to decide if the density of the pest is above an economic threshold. We propose a method for extending the scope of Bartlett's method to contexts where spatial GLMMs are appropriate for the data and we illustrate its implementation.

In chapter 4, we propose a generative spatial model-based sequential procedure as an alternative to the spatial GLMM-based sequential procedure discussed in chapter 3. The generative spatial model is compared and contrasted with the spatial GLMM.

Chapter 5 concludes the dissertation with a discussion of possible future work.

CHAPTER 1: Introduction of Sequential Probability Ratio Tests

1.1. *Wald's Sequential Probability Ratio Test (Wald's SPRT)*

As one of the oldest and most implemented sequential tests available in literature, Wald's SPRT was initiated during World War II (Wald, 1947). The goal of the SPRT is to sequentially discriminate between two simple hypotheses about the parameter of interest:

$$\mathrm{H}_0 : \theta = \theta_0 \qquad \mathrm{H}_1 : \theta = \theta_1 \qquad (\theta_0 \neq \theta_1),$$

subject to the nominal type-1 and type-2 error rates denoted by α and β respectively:

$$\Pr(\text{Reject } \mathrm{H}_0 \,|\, \mathrm{H}_0) = \alpha$$
$$\Pr(\text{Reject } \mathrm{H}_1 \,|\, \mathrm{H}_1) = \beta$$

Let $\{Y_i\}_{i=1}^{n}$ denote a sequence of random variables which have probability density function $f_n(\underset{\sim}{y}, \theta)$, where θ is the scalar parameter of interest. Let $\underset{\sim}{y}_n$ denote the observed $\{Y_i\}_{i=1}^{n}$ and define the logarithm of the probability ratio at stage $n \geq 1$:

$$LR_n = \ln \frac{f_n(\underset{\sim}{y}_n; \theta_1)}{f_n(\underset{\sim}{y}_n; \theta_0)} \qquad \text{for } n \geq 1$$

Wald's SPRT is defined by:

i. Continue collecting samples while $B < LR_n < A$

ii. Stop and accept H_0 if $LR_n \leq B$

iii. Stop and accept H_1 if $LR_n \geq A$

where the stopping bounds (A, B) are $A = \ln\frac{1-\beta}{\alpha}$ and $B = \ln\frac{\beta}{1-\alpha}$.

The Operating Characteristic (OC) function and the Average Sample Number (ASN) are usually used as the performance criteria for evaluating sequential procedures. The $\mathrm{OC}(\theta)$ is defined as the probability of accepting H_0 when θ is the true value of the parameter. The ASN function is defined as the mean number of samples required to make a decision about the simple hypotheses and it is also a function of θ. Smaller ASN values imply that the decisions are made faster and vice versa. It is desirable to have a test that has a small ASN function, while still achieving close to nominal $\mathrm{OC}(\theta_0)$ and $\mathrm{OC}(\theta_1)$ values. Wald provided formula to approximate the OC and ASN functions, and those formulas were very useful when computers were not readily available. It is straight forward to calculate the OC and ASN functions by simulations nowadays.

Sequential procedures are economical in the way that in most of the cases, a decision may be reached faster compared to a fixed sample size procedure that has the same type-1 and type-2 errors. Define N to be the number of samples required to make a decision, Wald and Wolfowitz proved that among all tests, whether fixed sample or sequential, for which $\mathrm{E}(N|\theta_i) < \infty, \quad i = 0, 1$, the SPRT with nominal error probabilities α and β minimizes both $E(N|\theta_0)$ and $E(N|\theta_1)$ (Wald and Wolfowitz 1948). Wald's SPRT is also perfectly general in that it applies to both independent and identically distributed (IID) data and non-IID data, both continuous and discrete random variables, and even random variables of different dimensions (Ghosh 1970). At the same time, Wald's SPRT has its own disadvantage and shortcoming, such as it does not specify a

maximum sample size, which means that cases could arise where an unsatisfactorily large number of samples are required to reach a conclusion. Wald's SPRT can only apply to hypotheses with one unknown parameter of interest, and it does not deal with the problem of nuisance parameters. Bartlett's SPRT is an extension to Wald's SPRT which accounts for nuisance parameter. We are going to talk about Bartlett's SPRT more in details in section 1.2.

1.1.1. *Wald's SPRT with IID Data*

Let $\{Y_i\}_{i=1}^{n}$ denote independent identically distributed random variables with probability function $f(y_i, \theta, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y_i-\theta)^2/2\sigma^2}$. We want to test the simple hypothesis about the mean which is $\mathrm{H}_0 : \theta = \theta_0$ versus $\mathrm{H}_1 : \theta = \theta_1$. Assuming that parameter σ^2 is known, Wald's SPRT has the following form:

$$\begin{aligned} &\text{i. Continue collecting samples while } \ln\frac{\beta}{1-\alpha} \le LR_n \le \ln\frac{1-\beta}{\alpha} \\ &\text{ii. Accept } \mathrm{H}_0, \quad \text{if} \quad LR_n \le \ln\frac{\beta}{1-\alpha} \\ &\text{iii. Accept } \mathrm{H}_1, \quad \text{if} \quad LR_n \ge \ln\frac{1-\beta}{\alpha} \end{aligned} \tag{1}$$

where

$$LR_n = \frac{\sum_{i=1}^{n}(y_i-\theta_0)^2 - \sum_{i=1}^{n}(y_i-\theta_1)^2}{2\sigma^2} = \frac{n(\theta_0^2-\theta_1^2) + 2(\theta_1-\theta_0)\sum_{i=1}^{n} y_i}{2\sigma^2}.$$

It can be seen that LR_n depends on $\underset{\sim}{y}$ through the cumulative sum of the n observations S_n, which is $S_n = \sum_{i=1}^{n} y_i$. When $\theta_1 > \theta_0$, the upper and lower stop boundaries for the Wald's SPRT test then can be simplified into:

$$\text{Accept } H_0, \quad \text{if} \quad S_n \le \frac{\sigma^2 B}{(\theta_1 - \theta_0)} + \frac{(\theta_1 + \theta_0)}{2} n$$

$$\text{Accept } H_1, \quad \text{if} \quad S_n \ge \frac{\sigma^2 A}{(\theta_1 - \theta_0)} + \frac{(\theta_1 + \theta_0)}{2} n$$

As n increases, S_n is compared to the two parallel stop boundaries. Sampling continues until either the upper line or the lower line is crossed. H_1 is accepted if the upper line is crossed and H_0 is accepted if the lower line is crossed. Similar derivations could be obtained when $\theta_1 < \theta_0$.

1.1.2. *Wald's SPRT with Non-IID Data*

The application of Wald SPRT doesn't require IID observation. When the random variables are non-IID, the sequential procedure is only slightly more complicated than under the IID assumption. In this section, we discuss the application of Wald's SPRT to non-IID data with a multivariate longitudinal data analysis example.

Let $\{Y_i\}_{i=1}^{n}$ denote a sequence correlated random variables with multivariate normal probability function:

$$f(\underset{\sim}{y}, \theta, \underset{\sim}{\sigma}^2) = \frac{1}{(2\pi)^{n/2} \left|\Sigma(\underset{\sim}{\sigma}^2)\right|^{1/2}} \exp\left(-\frac{1}{2} (\underset{\sim}{y} - \theta \underset{\sim}{1})' \Sigma(\underset{\sim}{\sigma}^2)^{-1} (\underset{\sim}{y} - \theta \underset{\sim}{1}) \right)$$

where $\Sigma(\underset{\sim}{\sigma}^2)$ is completely general (unstructured) $n \times n$ variance covariance matrix of the data which is expressed as $\Sigma(\underset{\sim}{\sigma}^2) = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2}^2 & \cdots & \sigma_{1,n}^2 \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{pmatrix}$ with $\underset{\sim}{\sigma}^2 = (\sigma_1^2, \sigma_{1,2}^2 \cdots, \sigma_n^2)$

$\underset{\sim}{\sigma}^2$ is known and we therefore denote $\Sigma(\underset{\sim}{\sigma}^2)$ with Σ. Wald's SPRT of testing simple hypotheses of $\mathrm{H}_0 : \theta = \theta_0$ versus $\mathrm{H}_1 : \theta = \theta_1$ ($\theta_0 \neq \theta_1$), with the proposed type-1 α and type-2 error β will be based on the log likelihood ratio:

$$LR_n = \ln \frac{f_n(\underset{\sim}{y}_n; \theta_1)}{f_n(\underset{\sim}{y}_n; \theta_0)} \qquad \text{for } n \geq 1$$
$$= \frac{1}{2}\left((\underset{\sim}{y} - \theta_0 \underset{\sim}{1})' \Sigma^{-1} (\underset{\sim}{y} - \theta_0 \underset{\sim}{1}) - (\underset{\sim}{y} - \theta_1 \underset{\sim}{1})' \Sigma^{-1} (\underset{\sim}{y} - \theta_1 \underset{\sim}{1}) \right)$$

The rejection/acceptance rule of the procedure is defined in (1):

i. Continue collecting samples while $\ln \frac{\beta}{1-\alpha} \leq LR_n \leq \ln \frac{1-\beta}{\alpha}$

ii. Accept H_0, if $LR_n \leq \ln \frac{\beta}{1-\alpha}$

iii. Accept H_1, if $LR_n \geq \ln \frac{1-\beta}{\alpha}$

To further understand Wald's SPRT procedure with correlated data, we write the test statistics LR_n as:

$$
\begin{aligned}
LR_n &= \ln \frac{f(y_1, \cdots y_n; \theta_1)}{f(y_1, \cdots y_n; \theta_0)} \qquad \text{for } n \geq 1 \\
&= \ln \frac{f(y_1, \cdots y_{n-1}; \theta_1) f(y_n | y_1, \cdots y_{n-1}; \theta_1)}{f(y_1, \cdots y_{n-1}; \theta_0) f(y_n | y_1, \cdots y_{n-1}; \theta_0)} \\
&= \ln \frac{f(y_1, \cdots y_{n-1}; \theta_1)}{f(y_1, \cdots y_{n-1}; \theta_0)} + \ln \frac{f(y_n | y_1, \cdots y_{n-1}; \theta_1)}{f(y_n | y_1, \cdots y_{n-1}; \theta_0)} \\
&= LR_{n-1} + \ln \frac{f(y_n | y_1, \cdots y_{n-1}; \theta_1)}{f(y_n | y_1, \cdots y_{n-1}; \theta_0)} \qquad (2)
\end{aligned}
$$

Denote $\ln \frac{f(y_n | y_1, \cdots y_{n-1}; \theta_1)}{f(y_n | y_1, \cdots y_{n-1}; \theta_0)}$ in (2) by λ_n, we see that λ_n is the increment of LR_n from sample size $n-1$ to sample size n during the sequential sampling procedure. Now let us look at how λ_n changes as the correlation among the data changes.

The conditional probability function of Y_n given $(Y_1 = y_1, \cdots, Y_n = y_{n-1})$ is:

$$
f(y_n | y_1, \cdots y_{n-1}) = \frac{1}{\sqrt{2\pi(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})}} \exp\left(- \frac{\left(y_n - \theta - \Sigma_{12}\Sigma_{22}^{-1}(y_1 - \theta, \cdots y_{n-1} - \theta)'\right)^2}{2(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})} \right)
$$

with $\Sigma_{11} = \sigma_n^2$, $\Sigma_{12} = \Sigma_{21}' = (\sigma_{n,1}^2, \cdots \sigma_{n,n-1}^2)$ and $\Sigma_{22} = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2}^2 & \cdots & \sigma_{1,n-1}^2 \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_{n-1}^2 \end{pmatrix}$.

It follows that,

$$
\lambda_n = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) - \frac{\left(y_n - \theta_1 - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{y}_{n-1} - \underset{\sim}{\theta}_1)\right)^2}{2(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})} +
$$

$$
\frac{1}{2}\ln 2\pi + \frac{1}{2}\ln(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) + \frac{\left(y_n - \theta_0 - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{y}_{n-1} - \underset{\sim}{\theta}_0)\right)^2}{2(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})}
$$

$$= \frac{\left(y_n - \theta_0 - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{y}_{n-1} - \underset{\sim}{\theta}_0)\right)^2 - \left(y_n - \theta_1 - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{y}_{n-1} - \underset{\sim}{\theta}_1)\right)^2}{2(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})}$$

$$= \frac{\left(\theta_1 - \theta_0 + \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{1}(\theta_0 - \theta_1)\right) \times \left(2y_n - \theta_0 - \theta_1 - \Sigma_{12}\Sigma_{22}^{-1}(2\underset{\sim}{y}_{n-1} - \underset{\sim}{\theta}_0 - \underset{\sim}{\theta}_1)\right)}{2(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})}$$

$$= \frac{(\theta_1 - \theta_0)(1 - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{1})}{2(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})} \times \left(2y_n - \theta_0 - \theta_1 - \Sigma_{12}\Sigma_{22}^{-1}(2\underset{\sim}{y}_{n-1} - \underset{\sim}{\theta}_0 - \underset{\sim}{\theta}_1)\right) \qquad (3)$$

As a special case, we study Σ as the first order autoregressive error structure, AR(1), which is constructed as:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho^n \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$$

With σ^2 be the variance parameter and ρ be the correlation parameter. Denote

$\Sigma_{11} = \sigma^2$, $\Sigma_{12} = \Sigma_{21}' = \sigma^2(\rho^{n-1}, \rho^{n-2}, \cdots \rho)$, $\Sigma_{22} = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho^{n-2} \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$, and

$\Sigma_{22}^{-1} = \frac{1}{(1-\rho^2)\sigma^2} \begin{pmatrix} 1 & -\rho & & 0 \\ -\rho & 1+\rho^2 & & \vdots \\ & & \ddots & -\rho \\ 0 & \cdots & -\rho & 1 \end{pmatrix}$. It is easy to see that $\Sigma_{12}\Sigma_{22}^{-1} = (0, 0, \cdots, \rho)$ and

$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \sigma^2(1-\rho^2)$ for any $n \geq 3$. Now (3) could be simplified into:

$$\lambda_n = \frac{(\theta_1 - \theta_0)}{2\sigma^2(1+\rho)}\left(2(y_n - \rho y_{n-1}) + (1-\rho)(\theta_0 + \theta_1)\right) \qquad (4)$$

We can calculate the expectation and the variance of (4) as

$$E(\lambda_n) = \frac{(\theta_1 - \theta_0)(2\theta - \theta_0 - \theta_1)(1-\rho)}{2\sigma^2(1+\rho)}$$

and,

$$Var(\lambda_n) = \frac{(\theta_1 - \theta_0)^2(1-\rho)}{\sigma^2(1+\rho)}.$$

Since $\lim_{\rho \to 1} E(\lambda_n) \to 0$, and $\lim_{\rho \to 1} Var(\lambda_n) \to 0$, λ_n converges in probability to 0 as ρ goes to 1. This tells us that as ρ gets large toward 1, the increment λ_n converges to 0 and the sample path of the sequential procedure will exhibit less changes. This is going to result in a large ASN values for the sequential test. At the same time, as ρ gets small towards -1, both expectation and variance of λ_n converges to infinite, and the sample path of sequential procedure will exhibit more changes, which decreases the ASN values. We therefore conclude that for Wald's SPRT, the correlation among the data has a great impact on the sample size required for the sequential test. High positive correlation among the observations slows down the changes of the sample path and increases the ASN values; on the other hand, low negative correlation reduces the ASN values.

Figure 1 shows the ASN curves of Wald's SPRT of testing $H_0 : \mu = 1.9$ vs $H_1 : \mu = 2.3$ with σ =0.4 for four different ρ values (0.9, 0.5, -0.5 and -0.9). The relationship between ρ and the ASN values is vividly characterized by Figure 1. We see the highest ASN curve is when $\rho = 0.9$, followed by $\rho = 0.5$, -0.5 and -0.9. **Figure 2**

shows the OC curves of Wald's SPRT for the four ρ values. It is notable that $\rho = -0.9$ yields the most conservative OC values, followed by $\rho = -0.5,\ 0.5$, and 0.9.

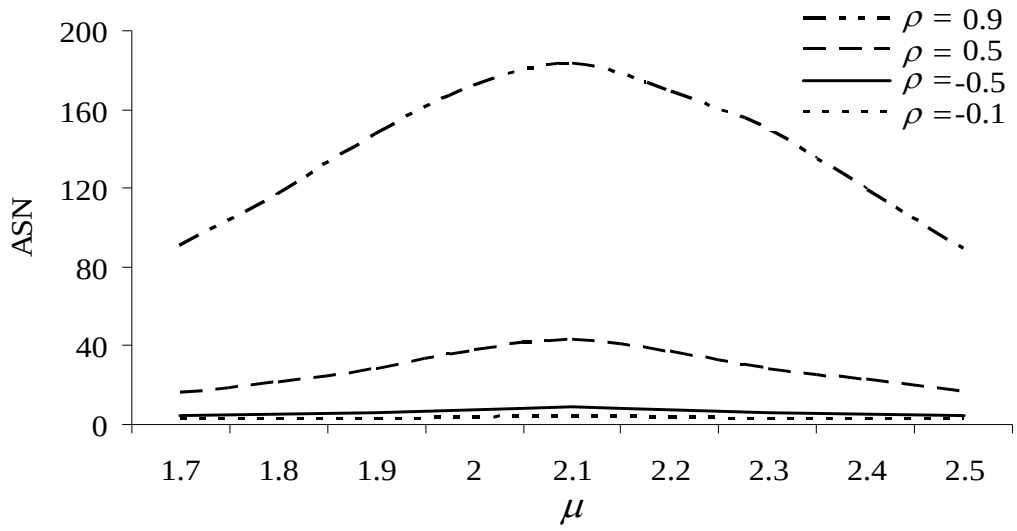

Figure 1. ASN Curves for $H_0 : \mu = 1.9$ vs $H_0 : \mu = 2.3$ with Different Correlations of AR(1) Model

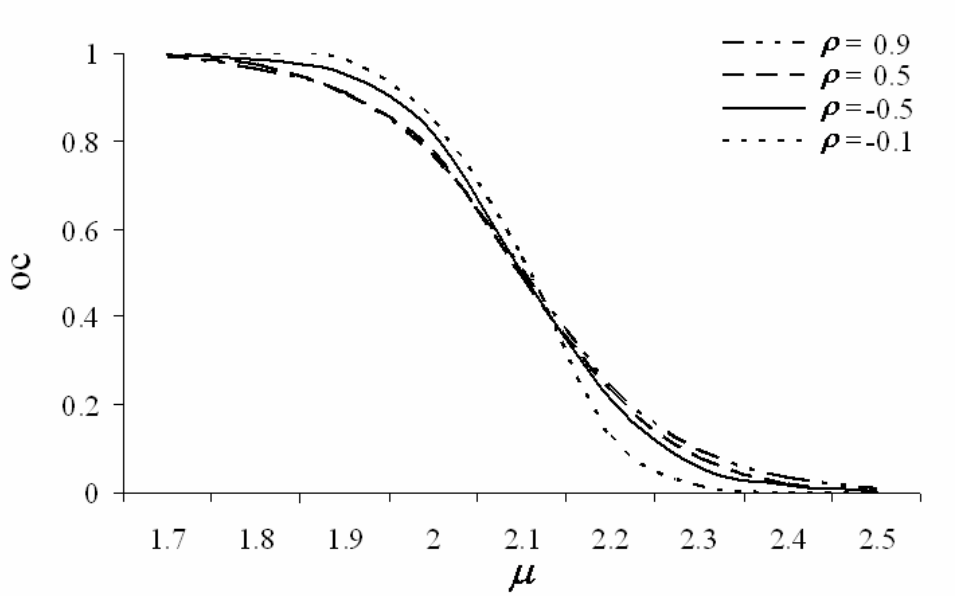

Figure 2. OC Curves for $H_0 : \mu = 1.9$ vs $H_0 : \mu = 2.3$ with Different Correlations of AR(1) Model

1.2. *Bartlett's Sequential Probability Ratio Test (Bartlett's SPRT)*

Wald's SPRT is very easy to implement and at the same time it does not require IID data. The limitation to use Wald's SPRT procedure is that it requires no unknown nuisance parameters. Bartlett's SPRT is an extension of Wald's SPRT, in the sense that they share exactly the same stop boundaries. However, Bartlett's SPRT has a provision for accounting for unknown nuisance parameters.

Consider the same simple hypotheses about the parameter of interest discussed in section 1.1:

$$\mathrm{H}_0 : \theta = \theta_0 \qquad \mathrm{H}_1 : \theta = \theta_1 \qquad (\theta_0 \neq \theta_1),$$

Bartlett (1946) proved in the IID case that a simple adjustment in the likelihood ratio is sufficient to preserve the asymptotic type-1 and type-2 error properties of the test procedure. Namely, if $\hat{\sigma}_n^2(\theta)$ denotes the conditional MLE of the unknown nuisance parameter σ^2, given a value of θ, an SPRT procedure can be based on the modified likelihood ratio

$$LR_n^* = \ln\left(\prod_{i=1}^{n} \frac{f(y_i ; \theta_1, \hat{\sigma}_n^2(\theta_1))}{f(y_i ; \theta_0, \hat{\sigma}_n^2(\theta_0))} \right)$$

and the same stop boundaries as for Wald's SPRT can be used:

i. Continue collecting samples while $B < LR_n^* < A$

ii. Stop and classify $\theta = \theta_0$ if $LR_n^* \leq B$

iii. Stop and classify $\theta = \theta_1$ if $LR_n^* \geq A$

where the stopping bounds (A, B) are two real numbers with $A = \ln\frac{1-\beta}{\alpha}$ and

$B = \ln\frac{\beta}{1-\alpha}$.

Bartlett's SPRT is different from Wald's SPRT in another way that Bartlett's SPRT does not have exact or approximate formula to compute the OC or ASN curves. But they are easily computed by simulation studies. In this section, we first discuss Bartlett's SPRT with IID data; we then show that Bartlett's SPRT can be applied to generalized linear model (GLM) contexts, where the data is non-IID.

1.2.1. *Bartlett's SPRT with IID Data*

The derivations of Bartlett's SPRT are based on multiple Taylor series expansions. We fill in the details of the derivations based on Govindarajulu (2004) as follows:

Without loss of generality, we assume only one nuisance parameter γ for the derivations. The hypothesis of interest is $\mathrm{H}_0 : \theta = \theta_0$ versus $\mathrm{H}_1 : \theta = \theta_1$, with $\theta_0 \neq \theta_1$. Namely, let $\hat{\gamma}_0$ and $\hat{\gamma}_1$ denote the conditional MLE of the unknown nuisance parameter γ, given θ_0 and θ_1 respectively. We firstly assume θ_1 is the true value of θ. Define $l(\theta, \gamma) = \sum_{i=1}^{n} \ln f(y_i; \theta, \gamma)$, and expand $l(\theta, \gamma)$ with a second order Taylor series expansion about $(\theta_1, \hat{\gamma}_1)$:

$$l(\theta,\gamma) \approx l(\theta_1,\hat{\gamma}_1) + \left.\frac{\partial l(\theta,\gamma)}{\partial\theta}\right|_{(\theta_1,\hat{\gamma}_1)} (\theta-\theta_1) + \left.\frac{\partial l(\theta,\gamma)}{\partial\gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\gamma-\hat{\gamma}_1) + \frac{1}{2}\left.\frac{\partial^2 l(\theta,\gamma)}{\partial^2\theta}\right|_{(\theta_1,\hat{\gamma}_1)} (\theta-\theta_1)^2$$
$$+ \frac{1}{2}\left.\frac{\partial^2 l(\theta,\gamma)}{\partial^2\gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\gamma-\hat{\gamma}_1)^2 + \frac{1}{2}\left.\frac{\partial^2 l(\theta,\gamma)}{\partial\theta\partial\gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\theta-\theta_1)(\gamma-\hat{\gamma}_1) \qquad (5)$$

Plug in $(\theta_0, \hat{\gamma}_0)$ for (θ, γ) in (5), we get the following result:

$$l(\theta_1, \hat{\gamma}_1) - l(\theta_0, \hat{\gamma}_0) \approx \left.\frac{\partial l(\theta,\gamma)}{\partial \theta}\right|_{(\theta_1,\hat{\gamma}_1)} (\theta_1 - \theta_0) + \left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\hat{\gamma}_1 - \hat{\gamma}_0) + \frac{1}{2}\left.\frac{\partial^2 l(\theta,\gamma)}{\partial^2 \theta}\right|_{(\theta_1,\hat{\gamma}_1)} (\theta_1 - \theta_0)^2$$
$$+ \frac{1}{2}\left.\frac{\partial^2 l(\theta,\gamma)}{\partial^2 \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\hat{\gamma}_1 - \hat{\gamma}_0)^2 + \left.\frac{\partial^2 l(\theta,\gamma)}{\partial \theta \partial \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\theta_1 - \theta_0)(\hat{\gamma}_0 - \hat{\gamma}_1) \qquad (6)$$

Since θ_1 is assume to be the true value of θ, based on the large sample properties, $\hat{\gamma}_1 \to \gamma$ as n gets large. Consequently, we have $-\frac{1}{n}\left.\frac{\partial^2 l(\theta,\gamma)}{\partial^2 \theta}\right|_{(\theta_1,\hat{\gamma}_1)} \to I_{\theta_1\theta_1}$, $-\frac{1}{n}\left.\frac{\partial^2 l(\theta,\gamma)}{\partial^2 \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} \to I_{\gamma\gamma}$, and $-\frac{1}{n}\left.\frac{\partial^2 l(\theta,\gamma)}{\partial \theta \partial \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} \to I_{\theta_1\gamma}$. If we use the traditional notation (θ, γ) for true values, (6) can be written as:

$$l(\theta_1, \hat{\gamma}_1) - l(\theta_0, \hat{\gamma}_0) \approx \left.\frac{\partial l(\theta,\gamma)}{\partial \theta}\right|_{(\theta_1,\hat{\gamma}_1)} (\theta_1 - \theta_0) + \left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\hat{\gamma}_1 - \hat{\gamma}_0)$$
$$+ \frac{n}{2}\left\{I_{\theta\theta}(\theta_1 - \theta_0)^2 + I_{\gamma\gamma}(\hat{\gamma}_1 - \hat{\gamma}_0)^2 - 2I_{\theta\gamma}(\theta_1 - \theta_0)(\hat{\gamma}_0 - \hat{\gamma}_1)\right\} \qquad (7)$$

We next consider a first-order Taylor series expansion of $\frac{\partial l(\theta,\gamma)}{\partial \gamma}$ about $(\theta_1, \hat{\gamma}_1)$:

$$\frac{\partial l(\theta,\gamma)}{\partial \gamma} \approx \left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} + \left.\frac{\partial^2 l(\theta,\gamma)}{\partial \theta \partial \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\theta - \theta_1) + \left.\frac{\partial^2 l(\theta,\gamma)}{\partial^2 \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\gamma - \hat{\gamma}_1) \qquad (8)$$

Plug in (θ_0, γ) and (θ_1, γ) for (θ, γ) in (8) respectively and take the difference of the two terms, we have:

$$\left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_0,\gamma)} - \left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\gamma)} \approx \left.\frac{\partial^2 l(\theta,\gamma)}{\partial \theta \partial \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\theta_0 - \theta_1) \approx nI_{\theta\gamma}(\theta_1 - \theta_0) \qquad (9)$$

Another way to approximate the left hand side of (9) is to apply the first order Taylor series expansion to each of the following two terms:

$$\left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_0,\gamma)} - \left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_0,\hat{\gamma}_0)} \approx \left.\frac{\partial^2 l(\theta,\gamma)}{\partial^2 \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\gamma-\hat{\gamma}_0) \approx -nI_{\gamma\gamma}(\gamma-\hat{\gamma}_0) \tag{10}$$

$$\left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} - \left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\gamma)} \approx -\left.\frac{\partial^2 l(\theta,\gamma)}{\partial^2 \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\gamma-\hat{\gamma}_1) \approx nl_{\gamma\gamma}(\gamma-\hat{\gamma}_1) \tag{11}$$

Summing (10) and (11) yields the following result:

$$\left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_0,\gamma)} - \left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\gamma)} \approx nI_{\gamma\gamma}(\hat{\gamma}_0-\hat{\gamma}_1) \tag{12}$$

Combing (9) and (12) yields

$$(\hat{\gamma}_0-\hat{\gamma}_1) \approx (\theta_1-\theta_0)\frac{I_{\theta\gamma}}{I_{\gamma\gamma}} \tag{13}$$

Approximate $\hat{\gamma}_0-\hat{\gamma}_1$ by $(\theta_1-\theta_0)\frac{I_{\theta\gamma}}{I_{\gamma\gamma}}$ based on (13), (7) can be simplified into:

$$l(\theta_1,\hat{\gamma}_1)-l(\theta_0,\hat{\gamma}_0) \approx \left.\frac{\partial l(\theta,\gamma)}{\partial \theta}\right|_{(\theta_1,\hat{\gamma}_1)} (\theta_1-\theta_0) + \left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\hat{\gamma}_1-\hat{\gamma}_0)$$

$$+\frac{n}{2}(\theta_1-\theta_0)^2\left\{I_{\theta\theta}-\frac{I_{\theta\gamma}^2}{I_{\gamma\gamma}}\right\} \tag{14}$$

Think of the first two terms in (14) as a function of $\hat{\gamma}_1$, a first order Taylor series expansion about γ yields:

$$\left.\frac{\partial l(\theta,\gamma)}{\partial \theta}\right|_{(\theta_1,\hat{\gamma}_1)} (\theta_1-\theta_0) + \left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\hat{\gamma}_1)} (\hat{\gamma}_1-\hat{\gamma}_0)$$

$$\approx \left.\frac{\partial l(\theta,\gamma)}{\partial \theta}\right|_{(\theta_1,\gamma)}(\theta_1-\theta_0)+\left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\gamma)}(\hat{\gamma}_1-\hat{\gamma}_0)+(\theta_1-\theta_0)\left.\frac{\partial^2 l(\theta,\gamma)}{\partial\theta\partial\gamma}\right|_{(\theta_1,\gamma)}(\hat{\gamma}_1-\gamma)$$

$$+(\hat{\gamma}_1-\hat{\gamma}_0)\left.\frac{\partial^2 l(\theta,\gamma)}{\partial^2\gamma}\right|_{(\theta_1,\gamma)}(\hat{\gamma}_1-\gamma)$$

$$\approx \left.\frac{\partial l(\theta,\gamma)}{\partial \theta}\right|_{(\theta_1,\gamma)}(\theta_1-\theta_0)+\left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\gamma)}(\hat{\gamma}_1-\hat{\gamma}_0)-\frac{n}{2}(\hat{\gamma}_1-\gamma)\left\{(\theta_1-\theta_0)I_{\theta\gamma}+(\hat{\gamma}_1-\hat{\gamma}_0)I_{\gamma\gamma}\right\} \quad (15)$$

The last term in (15) equals to 0 as a consequence of (13), and (15) is simplified into:

$$\left.\frac{\partial l(\theta,\gamma)}{\partial \theta}\right|_{(\theta_1,\hat{\gamma}_1)}(\theta_1-\theta_0)+\left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\hat{\gamma}_1)}(\hat{\gamma}_1-\hat{\gamma}_0)$$

$$\approx \left.\frac{\partial l(\theta,\gamma)}{\partial \theta}\right|_{(\theta_1,\gamma)}(\theta_1-\theta_0)+\left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\gamma)}(\hat{\gamma}_1-\hat{\gamma}_0) \quad (16)$$

Based on the result of (16), we rewrite (14) into:

$$l(\theta_1,\hat{\gamma}_1)-l(\theta_0,\hat{\gamma}_0)\approx \left.\frac{\partial l(\theta,\gamma)}{\partial \theta}\right|_{(\theta_1,\gamma)}(\theta_1-\theta_0)+\left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\gamma)}(\hat{\gamma}_1-\hat{\gamma}_0)$$

$$+\frac{n}{2}(\theta_1-\theta_0)^2\left\{I_{\theta\theta}-\frac{I_{\theta\gamma}^2}{I_{\gamma\gamma}}\right\} \quad (17)$$

Reuse the result of (13), rewrite (17) into:

$$l(\theta_1,\hat{\gamma}_1)-l(\theta_0,\hat{\gamma}_0)\approx \left.\frac{\partial l(\theta,\gamma)}{\partial \theta}\right|_{(\theta_1,\gamma)}(\theta_1-\theta_0)-\left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\gamma)}(\theta_1-\theta_0)\frac{I_{\theta\gamma}}{I_{\gamma\gamma}}$$

$$+\frac{n}{2}(\theta_1-\theta_0)^2\left\{I_{\theta\theta}-\frac{I_{\theta\gamma}^2}{I_{\gamma\gamma}}\right\} \quad (18)$$

Recall that θ_1 is assumed to be the true value of θ, we can simply rewrite $\left.\frac{\partial l(\theta,\gamma)}{\partial \gamma}\right|_{(\theta_1,\gamma)}$ by $\frac{\partial l(\theta,\gamma)}{\partial \gamma}$. Therefore, (18) can be written as:

$$l(\theta_1,\hat{\gamma}_1)-l(\theta_0,\hat{\gamma}_0)\approx(\theta_1-\theta_0)\left\{\frac{\partial l(\theta,\gamma)}{\partial\theta}-\frac{\partial l(\theta,\gamma)}{\partial\gamma}\frac{I_{\theta\gamma}}{I_{\gamma\gamma}}+\frac{n}{2}(\theta_1-\theta_0)\left(I_{\theta\theta}-\frac{I_{\theta\gamma}^2}{I_{\gamma\gamma}}\right)\right\} \tag{19}$$

In the IID case, (19) can be further written as:

$$\begin{aligned} & l(\theta_1,\hat{\gamma}_1)-l(\theta_0,\hat{\gamma}_0) \\ & \approx(\theta_1-\theta_0)\left\{\sum_{i=1}^{n}\frac{\partial\log f(y_i;\theta,\gamma)}{\partial\theta}-\frac{I_{\theta\gamma}}{I_{\gamma\gamma}}\frac{\partial\log f(y_i;\theta,\gamma)}{\partial\gamma}+\frac{n}{2}(\theta_1-\theta_0)\left(I_{\theta\theta}-\frac{I_{\theta\gamma}^2}{I_{\gamma\gamma}}\right)\right\} \\ & \approx\sum_{i=1}^{n}(\theta_1-\theta_0)\left\{\frac{\partial\log f(y_i;\theta,\gamma)}{\partial\theta}-\frac{I_{\theta\gamma}}{I_{\gamma\gamma}}\frac{\partial\log f(y_i;\theta,\gamma)}{\partial\gamma}+\frac{n}{2}(\theta_1-\theta_0)\left(I_{\theta\theta}-\frac{I_{\theta\gamma}^2}{I_{\gamma\gamma}}\right)\right\} \\ & \approx\sum_{i=1}^{n}\tilde{Z}_i \end{aligned} \tag{20}$$

where $\tilde{Z}_i$ are IID random variables and $E(\tilde{Z}_i)=\frac{(\theta_1-\theta_0)^2(I_{\theta\theta}-I_{\theta_\gamma}^2/I_{\gamma\gamma})}{2}$ and $Var(\tilde{Z}_i)=(\theta_1-\theta_0)^2(I_{\theta\theta}-I_{\theta\gamma}^2/I_{\gamma\gamma})$.

The above derivations are all based on the assumption that θ_1 was the true value. If we repeat the arguments assuming θ_0 was the true value, and apply Taylor series expansion around $(\theta_0,\hat{\gamma}_0)$ rather than $(\theta_1,\hat{\gamma}_1)$, then we will have:

$$l(\theta_1, \hat{\gamma}_1) - l(\theta_0, \hat{\gamma}_0)$$

$$\approx \sum_{i=1}^{n} (\theta_1 - \theta_0) \left\{ \frac{\partial \log f(y_i; \theta, \gamma)}{\partial \theta} - \frac{I_{\theta\gamma}}{I_{\gamma\gamma}} \frac{\partial \log f(y_i; \theta, \gamma)}{\partial \gamma} - \frac{n}{2} (\theta_1 - \theta_0) \left(I_{\theta\theta} - \frac{I_{\theta\gamma}^2}{I_{\gamma\gamma}} \right) \right\}$$

$$\approx \sum_{i=1}^{n} \tilde{Z}_i \tag{21}$$

where $\tilde{Z}_i$ are IID random variables with $E(\tilde{Z}_i) = -\frac{(\theta_1 - \theta_0)^2 (I_{\theta\theta} - I_{\theta\gamma}^2 / I_{\gamma\gamma})}{2}$ and $Var(\tilde{Z}_i) = (\theta_1 - \theta_0)^2 (I_{\theta\theta} - I_{\theta\gamma}^2 / I_{\gamma\gamma})$. To express (20) and (21) in one equation with θ interpreted as the generic true value, we will have the following equation:

$$l(\theta_1, \hat{\gamma}_1) - l(\theta_0, \hat{\gamma}_0)$$

$$\approx \sum_{i=1}^{n} (\theta_1 - \theta_0) \left\{ \frac{\partial \log f(y_i; \theta, \gamma)}{\partial \theta} - \frac{I_{\theta\gamma}}{I_{\gamma\gamma}} \frac{\partial \log f(y_i; \theta, \gamma)}{\partial \gamma} + (\theta - \frac{\theta_1 - \theta_0}{2}) \left(I_{\theta\theta} - \frac{I_{\theta\gamma}^2}{I_{\gamma\gamma}} \right) \right\}$$

with $E(\tilde{Z}_i) = (\theta_1 - \theta_0) \left(\theta - \frac{\theta_0 + \theta_1}{2} \right) (I_{\theta\theta} - I_{\theta\gamma}^2 / I_{\gamma\gamma})$ and $Var(\tilde{Z}_i) = (\theta_1 - \theta_0)^2 (I_{\theta\theta} - I_{\theta\gamma}^2 / I_{\gamma\gamma})$.

Now we draw an analogy as follows: Let $Y_1, \cdots, Y_n$ be IID random variables from a normal distribution with known mean μ and variance σ^2. The Wald's SPRT of hypothesis $\mathrm{H}_0 : \mu = \mu_0$ versus $\mathrm{H}_1 : \mu = \mu_1$ has the following stopping boundary:

$$\sigma^2 B < (\mu_1 - \mu_0) \sum_{i=1}^{n} (Y_i - \frac{\mu_0 + \mu_1}{2}) < \sigma^2 A \text{ with } A = \ln \frac{1-\beta}{\alpha} \text{ and } B = \ln \frac{\beta}{1-\alpha}.$$

If we replace Y_i by $\tilde{Z}_i$, and $(\mu, \mu_0, \mu_1, \sigma^2)$ by

$$\left(\mu, -\frac{(\mu_1 - \mu_0)^2 (I_{\theta\theta} - I_{\theta\gamma}^2 / I_{\gamma\gamma})}{2}, \frac{(\mu_1 - \mu_0)^2 (I_{\theta\theta} - I_{\theta\gamma}^2 / I_{\gamma\gamma})}{2}, (\mu_1 - \mu_0)^2 (I_{\theta\theta} - I_{\theta\gamma}^2 / I_{\gamma\gamma}) \right),$$

the corresponding Wald's SPRT stopping boundary becomes:

$$(\theta_1-\theta_0)^2\left(I_{\theta\theta}-I_{\theta\gamma}^2/I_{\gamma\gamma}\right)B<(\theta_1-\theta_0)^2\left(I_{\theta\theta}-I_{\theta\gamma}^2/I_{\gamma\gamma}\right)\sum_{i=1}^{n}\tilde{Z}_i<(\theta_1-\theta_0)^2\left(I_{\theta\theta}-I_{\theta\gamma}^2/I_{\gamma\gamma}\right)A,$$

which can be simplified as $\sigma^2 B<\sum_{i=1}^{n}\tilde{Z}_i<\sigma^2 A$. Therefore the stop boundaries of Bartlett's SPRT are the same boundaries used by Wald's SPRT.

The derivations of Bartlett's SPRT discussed above assume that $|\theta_1-\theta|$ is sufficiently small (Govindarajulu, 2004). We investigated this assumption by letting θ_0 and θ_1 diverge from each other. We see that OC(θ_0) and OC(θ_1) turn close to 1 and 0, and the ASN values are small. Although the test seems to have a perfect discrimination between the two hypotheses about the parameter of interests, but the SPRT is still quite practical.

1.2.2. *Bartlett's SPRT for GLMs*

In a GLM context, the observations are independent but not identically distributed. The non-IID characteristic of the data is captured through the use of an observable $q\times 1$ covariate vector $\underset{\sim}{u}$. A known link function $g(\cdot)$ is assumed such that the primary parameter of interest in the model, θ, varies from observation to observation through $g(\theta)=\zeta_0+\sum_{i=1}^{q}\zeta_i u_i$, where $\underset{\sim}{\zeta}=(\zeta_0,\zeta_1,\ldots,\zeta_q)'$ are unknown parameters. We can then write the probability function as $f(y;\underset{\sim}{u},\underset{\sim}{\zeta},\underset{\sim}{\sigma}^2)$.

Let ρ denote a coefficient in the link function that is of interest and define $\underset{\sim}{\xi}$ to be the vector of all the remaining coefficients together with $\underset{\sim}{\sigma}^2$. Then the probability

function can be written as $f(y;\underset{\sim}{u},\rho,\underset{\sim}{\xi})$. Let $\hat{\underset{\sim}{\xi}}_n(\rho)$ denote the conditional MLE of $\underset{\sim}{\xi}$ based on the first *n* observations. Ignoring the non-IID framework, a naïve use of Bartlett's SPRT to test $\mathrm{H}_0:\rho=\rho_0$ vs. $\mathrm{H}_1:\rho=\rho_1$ would be based on the log likelihood ratio

$$\lambda_n^* = \ln\left(\prod_{i=1}^{n}\{f(y_i;\underset{\sim}{u}_i,\rho_1,\hat{\underset{\sim}{\xi}}_n(\rho_1))/f(y_i;\underset{\sim}{u}_i,\rho_0,\hat{\underset{\sim}{\xi}}_n(\rho_0))\}\right). \tag{22}$$

This naïve procedure can be formally justified as follows.

Theorem 1. Suppose the data follows the GLM defined in this section and that the distribution of the covariate $\underset{\sim}{U}$ does not depend on the parameters $(\rho,\underset{\sim}{\xi})$. Then an SPRT procedure based on (22) which accepts H_0 at the first *n* for which $\lambda_n^* \le A$, accepts H_1 at the first *n* for which $\lambda_n^* \ge B$, and continues sampling as long as $A<\lambda_n^*<B$, will have asymptotic type-1 and type-2 error rates equal to α and β, respectively.

Proof. First we note that $f(y;\underset{\sim}{u},\rho,\underset{\sim}{\xi})$ can be interpreted as a conditional probability function for Y, given $\underset{\sim}{U}=\underset{\sim}{u}$. Let $h(\underset{\sim}{u})$ denote the probability function of $\underset{\sim}{U}$. The joint probability function of $(Y,\underset{\sim}{U})$ is $f(y;\underset{\sim}{u},\rho,\underset{\sim}{\xi})h(\underset{\sim}{u})$. Since the observations $(Y_i,\underset{\sim}{U}_i)_{i=1}^n$ are independent and identically distributed, Bartlett's SPRT applies to this sequence using the log likelihood ratio $\ln\left(\prod_{i=1}^{n}\{f(y_i;\underset{\sim}{u}_i,\rho_1,\hat{\underset{\sim}{\xi}}_n(\rho_1))h(\underset{\sim}{u}_i)\}/\{f(y_i;\underset{\sim}{u}_i,\rho_0,\hat{\underset{\sim}{\xi}}_n(\rho_0))h(\underset{\sim}{u}_i)\}\right)$ and it is easy to see this is the same likelihood ratio given in (22).

CHAPTER 2: Design of Sequential Hypothesis Testing Methodology for Poisson GLMM

2.1. *Motivating Clinical Example*

In this chapter, we consider the problem of sequentially testing a simple versus simple hypothesis of the form $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ for problems where the data mechanism is a Poisson GLMM. In Chapter 1, we showed that Bartlett SPRT can be applied to GLM contexts, where the observations are independent but not identically distributed. However, sequential testing procedures under GLMM contexts have not been discussed in the existing literature. The presence of nuisance parameters disqualifies use of Wald's SPRT and the departure from IID observations seemingly disqualifies the use of Bartlett's SPRT. Based on the finding that Bartlett's SPRT does indeed apply to GLM contexts, we propose an extended Bartlett SPRT in this Chapter for use in Poisson GLMM.

There has been increasing research interest in sequential analysis of clinical trials based on normal theory. For example, Lee and DeMets (1991), Wu and Lan (1992), Scharfstein et al. (1997), and Spiessens et al. (2000) developed group sequential tests based on a normal-theory mixed effects linear model. However, in clinical trial studies, the data are often binary, count or categorical. Coad and Ivanova (2005) and Baksh et al. (2005) investigated the use of the triangular test (Whitehead, 1997) for binary data based on generalized linear model. Spiessens et al. (2002, 2003) compared group sequential

methods for binary longitudinal data based on a generalized estimating equation approach and a random effects model approach. All the above mentioned literature focuses on group sequential methods. Little attention has been paid to single-individual sequential hypothesis tests based on non-IID responses. Single-individual sequential hypothesis tests can be just as easy to operate and offer some savings in the required sample size. Our goal in this chapter is to develop a single-individual sequential test procedure that applies to non-IID settings, and in particular that applies to a motivating multi-center clinical trial study that we next describe.

Surgical Site Infections (SSI) is a significant source of post-operative morbidity and mortality. They are the third most common hospital-acquired infection, and account for 14% to 16% of all infections. The associated increase in treatment cost is around $2000 to $4500 per case, and can extend the postoperative length of stay by 7 to 10 days (Paulsen et al., 1994). Our goal is to design a sequential multicenter clinical trial study to compare two methods of antibiotic delivery for the prevention of SSI. The current method for SSI prophylaxis is an intravenous (IV) infusion of the antibiotic cefazolin, delivered 30-60 minutes prior to incision. A new delivery method, known as tumescent cefazolin delivery, consists of a subcutaneous infiltration (injection) of a large volume of dilute cefazolin in a solution of dilute local anesthesia (lidocaine and epinephrine) directly into the subcutaneous tissue surrounding the incision site.

The standard IV infusion of cefazolin delivers the antibiotic to the entire body and is associated with increased risks of potentially lethal clostridium difficile colitis as well as the development of antibiotic bacterial resistance. The advantage of tumescent

cefazolin is that there is a much higher and more prolonged concentration of cefazolin precisely at the incision site with significantly lower systemic antibiotic concentrations and therefore smaller risks of adverse side effects. We refer to these two treatment options as 'treatment' (tumescent) and 'control' (IV), respectively. Our objective is to propose a sequential hypothesis test of the efficacy of treatment versus control, and compare it with a fixed sample hypothesis test.

We assume a-priori that the median length of post surgery hospitalization for patients in the control group is 12 days. A treatment effect of 3 days is considered to have practical significance, so we impose that whatever hypothesis testing procedure is used should reliably detect a decrease of 3 days in the median hospitalization time. Our response variable is the hospitalization time following surgery, measured in number of days. Thus, a discrete probability distribution is a natural model for the data. Patients are admitted to one of several centers for their surgery, and it is reasonable to expect a center effect to influence the hospitalization times of all patients within the same center. Thus, we were led to use a Poisson GLMM as a model for the data. In particular, let Y_{ijk} be the number of days of post surgery hospitalization for the *k*-th patient in the *i*-th center under the *j*-th treatment, where $i = 1, \ldots, m$, $j = 1, 2$ and $k = 1, \ldots, n_{ij}$.

Let S_i denote a random effect associated with the *i*-th center. We assume the post surgery hospitalization time follows the Poisson GLMM defined by:

$$\begin{aligned} &a.\ \ Y_{ijk} \,|\, S_i \overset{ind}{\sim} \text{Poisson}(\mu_{ijk}) \\ &b.\ \ \log(\mu_{ijk}) = \zeta_0 + \zeta_1 u_j + S_i \\ &c.\ \ S_i \overset{iid}{\sim} N(0, \sigma^2) \end{aligned} \tag{23}$$

where u_j is 0 for patients assigned to the control (IV) treatment and 1 for patients assigned to the tumescent treatment method. The random effect, S_i, causes responses from all the patients in the i-th center to be correlated. Under this model the median value of the response under the control and treatment is $\exp(\zeta_0)$ and $\exp(\zeta_0 + \zeta_1)$, respectively. We assume $\zeta_0 = 2.5$ is known, to yield the a-priori median response of approximately 12 days for the control group. The value $\zeta_1 = -0.3$ corresponds to a median value of approximately 9 days for the treatment group. Therefore, the hypothesis we wish to test is $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$, with σ^2 regarded as a nuisance parameter. We take both the type-1 and type-2 error rates to be $\alpha = \beta = 0.01$.

It is worth mentioning that one can also compare the control group and the treatment group by keeping track of the presence or absence of SSI for each patient. In this way, the response variable will be binary and a logistic regression GLMM model can be fit to the data.

2.2. *Fixed Sample Size Test*

For the purpose of deriving a fixed sample size test of H_0 vs H_1, we suppose ζ_0 and σ^2 are known. Sometimes this can be a reasonable assumption, since those two parameters fully describe the distribution of hospitalization times for patients in the control group. If the control group is a legacy treatment, then it is feasible their values could be known from prior experience. Suppose there are four participating centers and $(n_{i1}, n_{i2})_{i=1}^4$ denote the number of patients in each of the two groups at each center. We

assume that we will use a balanced design where both the control and treatment group have r (to-be determined) patients. Then, using the usual notation for summing over subscripts $y_{i1\cdot} = \sum_{k=1}^{n_{i1}} y_{i1k}$, $y_{i2\cdot} = \sum_{k=1}^{n_{i2}} y_{i2k}$ and $n_{\cdot 1} = n_{\cdot 2} = r$. The way this clinical trial would actually be implemented is that as patients arrive at a center they are alternately assigned to the control or treatment group, and the data is analyzed when enough patients have arrived across the participating centers to make $n_{\cdot 1} = n_{\cdot 2} = r$. The total sample size is thus $2r$. The integrated likelihood function based on the observations is

$$L(\zeta_1) = \prod_{i=1}^{m} \int_{-\infty}^{\infty} \left(\frac{\exp(-n_{i1} e^{\zeta_0 + s_i}) \exp((\zeta_0 + s_i) y_{i1\cdot})}{\prod_{k=1}^{n_{i1}} y_{i1k}!} \right) \times \left(\frac{\exp(-n_{i2} e^{\zeta_0 + \zeta_1 + s_i}) \exp((\zeta_0 + \zeta_1 + s_i) y_{i2\cdot})}{\prod_{k=1}^{n_{i1}} y_{i2k}!} \right) \times \frac{\exp(-s_i^2 / 2\sigma^2)}{\sqrt{2\pi\sigma^2}} ds_i \tag{24}$$

For a fixed r, the most powerful size α test of $\mathrm{H}_0 : \zeta_1 = c_0$ vs $\mathrm{H}_1 : \zeta_1 = c_1$ is the Neyman-Pearson (NP) test that rejects H_0 if $\varphi = L(c_1)/L(c_0) > k_\alpha$ and accepts H_0 if $\varphi < k_\alpha$, where k_α is chosen to satisfy $P(\varphi > k_\alpha \mid \mathrm{H}_0) = \alpha$. In order to determine k_α , the null distribution of φ is needed. While an analytic form of the null distribution is not available, it is relatively easy to use Monte Carlo methods to simulate the null distribution. To this end, we could simulate 100,000 vectors of data, $\underset{\sim}{Y}$, under H_0 , and then extract the 99$^{\text{th}}$ percentile of the corresponding φ values as the estimate of k_α . We note, however, that for each data vector $\underset{\sim}{Y}$, evaluating the integrated likelihoods $L(c_1)$ and $L(c_0)$ required use of a numerical integration method. We used Gauss-Hermite

quadrature rule for that purpose. Gaussian quadrature is often precise with a remarkably small number of nodes, M (Givens and Hoeting 2005). We found from our application that $M = 50$ is sufficed for one dimensional integral.

For every r, the value of k_α guarantees the type-1 error of the NP test will be α. Our goal is to search for the value of r which will make the type-2 error equal to β. We do this by sequentially trying all values of $r \geq 2$ and using the Monte Carlo method above to evaluate the power of the corresponding size α NP test. That is, we simulate 100,000 data vectors $\underset{\sim}{Y}$ under H_1 and estimate the power by the fraction for which $\varphi = L(-0.3)/L(0) > k_\alpha$. The search stops at the first value of n for which the power is 0.99 or larger. Columns 4 and 8 of **Table 1** below show the required n for selected values of ζ_0 and σ. The second and the sixth columns of the table show the corresponding mean of an observation under H_0, while the third and the seventh columns of the table show the variance of an observation under H_0. Note that increasing σ has a small effect on the mean of the data, but a large effect on the variance. The increase in the required sample size is, however, quite modest.

ζ_0	$\sigma = 0.2$				$\sigma = 0.4$			
	$E(Y_{ijk} \mid \mathrm{H}_0)$	$Var(Y_{ijk} \mid \mathrm{H}_0)$	r	$k_{0.01}$	$E(Y_{ijk} \mid \mathrm{H}_0)$	$Var(Y_{ijk} \mid \mathrm{H}_0)$	r	$k_{0.01}$
2.2	9.21	12.67	55	0.9910	9.78	26.36	58	0.9909
2.3	10.18	14.40	50	0.9915	10.80	31.06	53	0.9907
2.4	11.25	16.41	46	0.9904	11.94	36.68	48	0.9906
2.5	12.43	18.73	41	0.9912	13.20	43.42	43	0.9907
2.6	13.74	21.44	37	0.9908	14.59	51.50	39	0.9903
2.7	15.18	24.58	34	0.9917	16.12	61.20	36	0.9907
2.8	16.78	28.26	31	0.9915	17.81	72.88	33	0.9912

Table 1. Required sample sizes *r* for each group, to provide type-1 and type-2 error rates of 0.01 when using the NP test of $\mathrm{H}_0 : \zeta_1 = 0$ vs $\mathrm{H}_1 : \zeta_1 = -0.3$

2.3. *Wald's SPRT*

Let n denote the current number of observation available to analyze and ζ_0 and σ^2 are known, a Wald SPRT of $\mathrm{H}_0 : \zeta_1 = c_0$ vs $\mathrm{H}_1 : \zeta_1 = c_1$ can be implemented by computing λ_n from the integrated likelihood shown in (24). As discussed in section 2.2, we used a M point Gaussian Hermite quadrature rule to evaluate the integrated likelihoods $L(c_1)$ and $L(c_0)$. Hence Wald's SPRT test statistics for the hypothesis of $\mathrm{H}_0 : \zeta_1 = c_0$ against $\mathrm{H}_1 : \zeta_1 = c_1$ is:

$$\lambda_n = \sum_{i=1}^{m} \log\left\{ \sum_{t=1}^{M} \exp\left(-n_{i1} e^{\zeta_0 + \sqrt{2}\sigma x_t} - n_{i2} e^{\zeta_0 + c_1 + \sqrt{2}\sigma x_t} + \sqrt{2}\sigma x_t y_{i1\cdot} + (c_1 + \sqrt{2}\sigma x_t) y_{i2\cdot} \right) \times w_t \right\}$$
$$- \sum_{i=1}^{m} \log\left\{ \sum_{t=1}^{M} \exp\left(-n_{i1} e^{\zeta_0 + \sqrt{2}\sigma x_t} - n_{i2} e^{\zeta_0 + c_0 + \sqrt{2}\sigma x_t} + \sqrt{2}\sigma x_t y_{i1\cdot} + (c_0 + \sqrt{2}\sigma x_t) y_{i2\cdot} \right) \times w_t \right\}$$

where x_t are the roots of Hermite polynomial and w_t are the associated weights. The rejection/acceptation rule for Wald's SPRT is:

i. Continue collecting samples while $\ln B < \lambda_n < \ln A$

ii. Stop and classify $\zeta_1 = c_0$ if $\lambda_n < \ln B$

iii. Stop and classify $\zeta_1 = c_1$ if $\lambda_n > \ln A$

where $A = \frac{1-\beta}{\alpha}$ and $B = \frac{\beta}{1-\alpha}$, and α and β are the proposed type I and type II error respectively.

The performance of a sequential hypothesis test procedure can be evaluated by examining the OC curve and the ASN curve, that give the probability of accepting the null hypothesis and the average number of samples required to reach a decision, respectively. The OC and ASN curves can be estimated by simulating a large number of sequential samples and recording the fraction of cases where the null hypothesis is accepted and correspondingly the average number of samples required to make the decision.

To illustrate the OC and ASN characteristics of Wald's SPRT, we conducted a simulation study as if the data is from a multicenter randomized clinical trials study with four participating centers. Each patient arrives at a participating center is assigned a number representing their order of arrival. Odd numbered patients were then randomly assigned to a treatment by the flip of a coin; the next even numbered patient was then assigned the alternate treatment to maintain a balance between the total numbers of patients assigned to each group at that center.

The OC and ASN curves of the Wald's SPRT depend on the values of ζ_0 and σ. What is true in all cases, however, is that the type-1 and type-2 error will not exceed the nominal values of α and β, respectively. To compare the fixed sample size test and the

Wald SPRT test of $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$, **Table 2** shows simulated OC and ASN curves (as a function of the true ζ_1) for the Wald's SPRT for the two cases $(\zeta_0, \sigma) = (2.5, 0.2)$ and $(\zeta_0, \sigma) = (2.5, 0.4)$. The simulation results are based on 5000 sample paths of Wald's SPRT where the nominal type-1 and type-2 errors are set to 0.01.

We can see that the simulated type-1 and type-2 error rates for the two cases in **Table 2** are $(0.006, 0.007)$ and $(0.005, 0.007)$, respectively, and are comfortably close to the nominal .01 values. Apart from the ASN values, **Table 2** also shows percentiles of the sample size required with the Wald SPRT. As was the case for the fixed sample size procedure, there is not a lot of difference in the OC and ASN characteristics for the two cases $\sigma = 0.2$ and $\sigma = 0.4$.

ζ_1	$(\zeta_0, \sigma) = (2.5, 0.2)$							$(\zeta_0, \sigma) = (2.5, 0.4)$						
	OC	Sample Size Distribution						OC	Sample Size Distribution					
		ASN	10th	25th	50th	75th	90th		ASN	10th	25th	50th	75th	90th
-0.45	0.000	18.73	11	13	17	23	29	0.000	21.16	12	15	19	25	33
-0.4	0.001	22.41	12	15	20	27	36	0.000	24.86	13	17	22	30	40
-0.35	0.001	27.74	13	17	24	34	48	0.001	30.43	14	19	26	37	52
-0.3	0.007	34.59	14	20	29	43	63	0.007	39.37	17	23	33	48	71
-0.25	0.032	49.78	17	25	40	63	96	0.036	53.61	19	27	43	69	103
-0.2	0.158	73.40	20	31	56	95	150	0.164	77.00	22	35	60	101	156
-0.15	0.480	88.33	21	35	64	115	185	0.496	91.84	24	40	70	121	189
-0.1	0.833	70.14	18	29	54	93	143	0.832	73.05	21	33	57	95	147
-0.05	0.962	47.08	14	21	36	61	95	0.965	50.85	17	25	41	65	97
0	0.994	32.94	12	17	27	42	63	0.995	35.63	14	20	29	45	65
0.05	0.998	24.04	10	14	20	30	43	0.998	25.91	12	16	22	32	44
0.1	1.000	18.54	9	11	16	23	32	0.999	20.31	10	13	18	25	33
0.15	1.000	15.38	8	10	14	19	25	1.000	16.79	9	12	15	20	26

Table 2. OC and ASN properties of Wald SPRT for $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ when nominal type-1 and type-2 error rates are 0.01

Figure 3 depicts the OC characteristics for the two cases $\sigma = 0.2$ and $\sigma = 0.4$. Similar with the case for the fixed sample size procedure, there is not a lot of difference. **Figure 4** includes an illustrative comparison of the savings in sampling using the Wald's SPRT as compared to the fixed sample size NP test for the case of $\sigma = 0.2$. The comparison for the case of $\sigma = 0.4$ has similar result. It is seen that the largest ASN value occurs around the average value of 0 and -0.3, since the midpoint is the hardest to make a decision as to which of H_0 or H_1 is true. The ASN values gradually decreases as the value of ζ_1 moves further away from 0 and -0.3. What is also notable about the comparison is that relative to simpler models, the region where the fixed sample size plan does better than the SPRTs for our GLMM model is narrower.

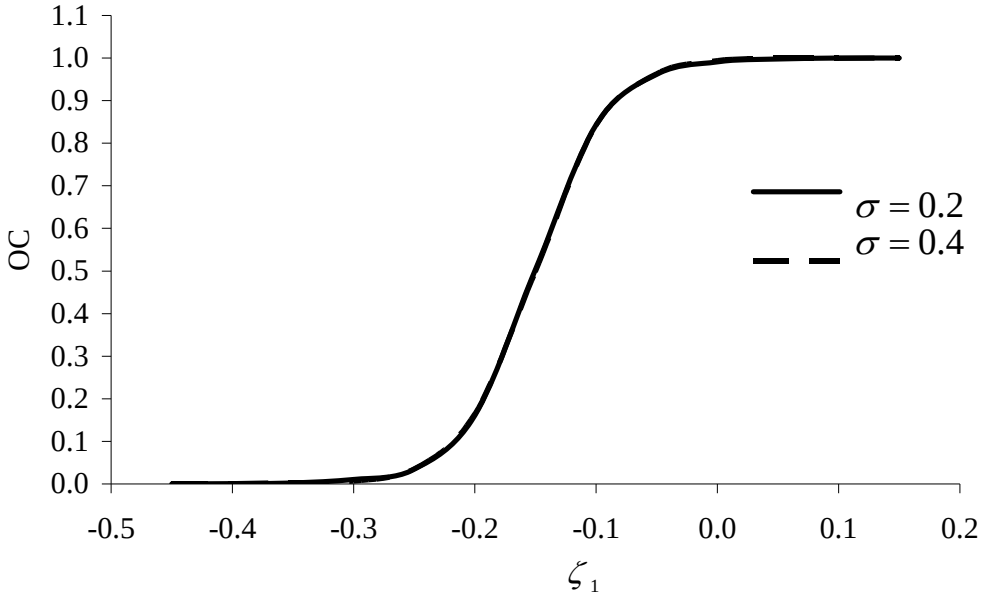

Figure 3. OC Curve for Wald SPRT Test of $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ with $(\zeta_0, \sigma) = (2.5, 0.2)$ and $(\zeta_0, \sigma) = (2.5, 0.4)$ and nominal type-1 and type-2 error rates of 0.001

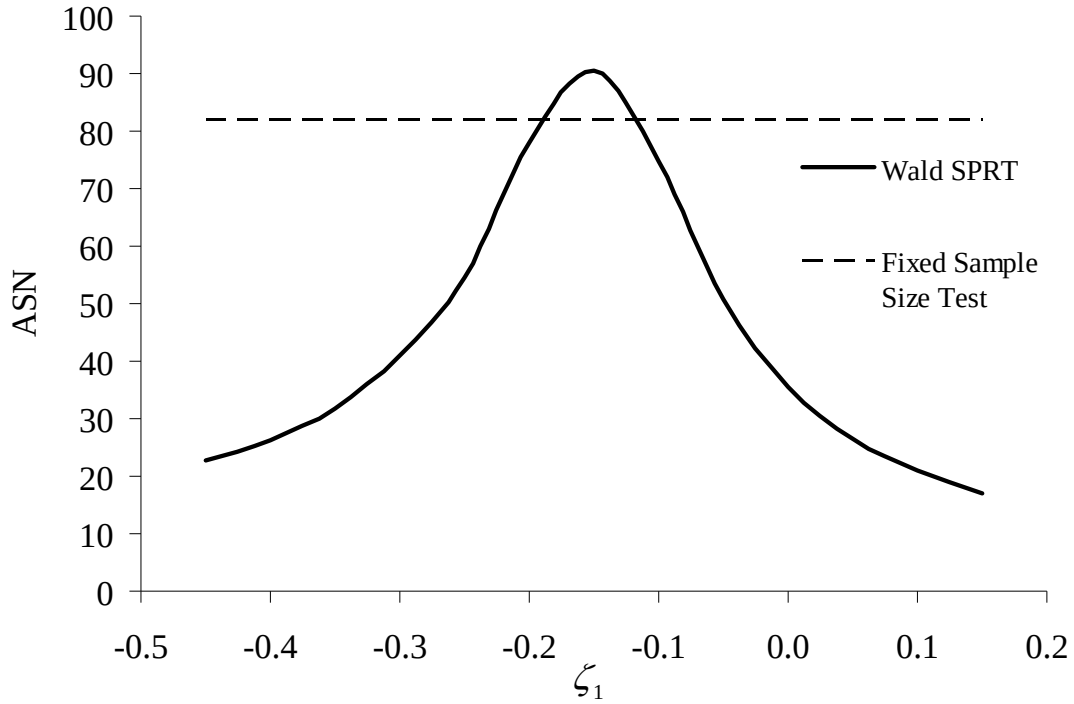

Figure 4. ASN Curve for Wald SPRT and Fixed Sample Size Test of $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ with $(\zeta_0, \sigma) = (2.5, 0.2)$ and nominal type-1 and type-2 error rates of 0.001

2.4. *Proposed SPRT for Poisson GLMM*

2.4.1. *GLMM-Bartlett SPRT*

For the GLMM setup (23), both the fixed sample size NP test and the Wald's SPRT procedures are based on the assumption that ζ_0 and σ^2 are known. While the case of known ζ_0 and σ^2 is plausible when there is historical information about the control group available, there are many applications where either no information on the control group is available, or when both groups represent new treatment options. In this

case, ζ_0 and σ^2 are unknown nuisance parameters and we need to derive an appropriate SPRT procedure.

Let *n* denote the current number of observation available to analyze, and as before let $(n_{i1}, n_{i2})_{i=1}^{m}$ denote the number of patients in each treatment group for the different centers. Conditional on $\underset{\sim}{S} = \underset{\sim}{s}$, the observations come from a GLM model. Thus, by Theorem 1 if $\underset{\sim}{S}$ were observable Bartlett's SPRT to test $H_0 : \zeta_1 = c_0$ vs. $H_1 : \zeta_1 = c_1$ could be used with the likelihood ratio

$$\lambda_n^* = \frac{\prod_{i=1}^{m} \left\{ \exp\left(-e^{s_i}(n_{i1}e^{\hat{\zeta}_{0,n}(c_1)} + n_{i2}e^{\hat{\zeta}_{0,n}(c_1)+c_1}) + \hat{\zeta}_{0,n}(c_1) y_{i1\cdot} + (\hat{\zeta}_{0,n}(c_1) + c_1) y_{i2\cdot} \right) / \prod_{k=1}^{n_{i1}} y_{i1k}! \prod_{k=1}^{n_{i2}} y_{i2k}! \right\}}{\prod_{i=1}^{m} \left\{ \exp\left(-e^{s_i}(n_{i1}e^{\hat{\zeta}_{0,n}(c_0)} + n_{i2}e^{\hat{\zeta}_{0,n}(c_0)+c_0}) + \hat{\zeta}_{0,n}(c_0) y_{i1\cdot} + (\hat{\zeta}_{0,n}(c_0) + c_0) y_{i2\cdot} \right) / \prod_{k=1}^{n_{i1}} y_{i1k}! \prod_{k=1}^{n_{i2}} y_{i2k}! \right\}}$$

where $\hat{\zeta}_{0,n}(c_1)$ and $\hat{\zeta}_{0,n}(c_0)$ are the conditional MLEs of ζ_0, given $\zeta_1 = c_1$ and $\zeta_1 = c_0$, respectively. We can see that λ_n^* depends on the unrealized values of the random effects through the function $\tau_i = \exp(s_i)$. Since these quantities are not observable, we propose to replace them by a consistent predictor based on the observable data. Noting that the $\{S_i\}_{i=1}^{m}$ are independent, the best linear predictor (BLP) of τ_i only depends on the observable data through $\underset{\sim}{Y}_i = (Y_{i11}, \cdots, Y_{i1n_{i1}}, Y_{i21}, \cdots, Y_{i2n_{i2}})'$ and is equal to

$$\text{BLP}(\tau_i) = E(\tau_i) + Cov(\underset{\sim}{Y}_i, \tau_i)' \left[Cov(\underset{\sim}{Y}_i) \right]^{-1} (\underset{\sim}{Y}_i - E(\underset{\sim}{Y}_i)) \ . \qquad (25)$$

It is obvious that $E(\tau_i) = e^{\sigma^2/2}$. $Cov(\underset{\sim}{Y}_i, \tau_i)$ and $Cov(\underset{\sim}{Y}_i)$ can be calculated as follows:

$$Cov(\underset{\sim}{Y}_i, \tau_i) = Cov\left(E(\underset{\sim}{Y}_i \middle| S_i), E(\tau_i \middle| S_i)\right) + E\left(Cov(\underset{\sim}{Y}_i, \tau_i \middle| S_i)\right)$$

$$= Cov\left(E(\underset{\sim}{Y}_i \middle| S_i), E(\tau_i \middle| S_i)\right)$$

$$= Cov\left((e^{\zeta_0+S_i}, \cdots, e^{\zeta_0+S_i}, e^{\zeta_0+\zeta_1+S_i}, \cdots, e^{\zeta_0+\zeta_1+S_i})', e^{S_i}\right)$$

$$= e^{\sigma^2}(e^{\sigma^2}-1)(e^{\zeta_0}\underset{\sim}{1}'_{n_{i1}\times 1}, e^{\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1})' \qquad (26)$$

Let $V_1 = Var(Y_{i1k})$, where y_{i1k} is the observation for the *k*th patient in the *i*th center under the control IV treatment method, then:

$$V_1 = E\left(Var(Y_{i1k} \middle| S_i)\right) + Var\left(E(Y_{i1k} \middle| S_i)\right)$$

$$= e^{\zeta_0+\sigma^2/2} + e^{2\zeta_0+\sigma^2}(e^{\sigma^2}-1)$$

Similarly, let $V_2 = Var(Y_{i2k})$ where y_{i2k} is the observation for the *k*th patient in the *i*th center under the tumescent treatment method:

$$V_2 = E\left(Var(Y_{i2k} \middle| S_i)\right) + Var\left(E(Y_{i2k} \middle| S_i)\right)$$

$$= e^{\zeta_0+\zeta_1+\sigma^2/2} + e^{2\zeta_0+2\zeta_1+\sigma^2}(e^{\sigma^2}-1)$$

Let $C_1 = Cov(Y_{i1h}, Y_{i1l})$, where y_{i1h} and y_{i1l} represent two different patients from the same *i*th center under the same control IV method:

$$C_1 = Cov\left(E(Y_{i1h} \middle| S_i), E(Y_{i1l} \middle| S_i)\right) + E\left(Cov(Y_{i1h}, Y_{i1l} \middle| S_i)\right)$$

$$= Cov\left(E(Y_{i1h} \middle| S_i), E(Y_{i1l} \middle| S_i)\right)$$

$$= e^{2\zeta_0+\sigma^2}(e^{\sigma^2}-1), h \neq l$$

Similarly, let $C_2 = Cov(Y_{i2h}, Y_{i2l})$ where y_{i2h} and y_{i2l} represent two different patients from the same *i*th center and under the same tumescent treatment method:

$$C_2 = e^{2\zeta_0 + 2\zeta_1 + \sigma^2}(e^{\sigma^2} - 1), h \neq l$$

$C_3 = Cov(Y_{i1h}, Y_{i2l})$ with Y_{i1h} and Y_{i2l} represent two different patients in the same *i*th center but under the different treatment methods:

$$C_3 = Cov\left(E(Y_{i1h}|S_i), E(Y_{i2l}|S_i)\right) + E\left(Cov(Y_{i1h}, Y_{i2l}|S_i)\right)$$

$$= Cov\left(E(Y_{i1h}|S_i), E(Y_{i2l}|S_i)\right)$$

$$= e^{2\zeta_0 + \zeta_1 + \sigma^2}(e^{\sigma^2} - 1)$$

Now $Cov(\underset{\sim}{Y}_i)$ can be expressed in term of V_1, V_2, C_1, C_2, and C_3 as follows:

Let $A_{11} = \begin{pmatrix} V_1 & C_1 & \cdots & C_1 \\ & V_1 & \ddots & \vdots \\ & & \ddots & C_1 \\ & & & V_1 \end{pmatrix}$, $A_{12} = \begin{pmatrix} C_3 & \cdots & C_3 \\ & \ddots & \vdots \\ & & C_3 \end{pmatrix}$, $A_{22} = \begin{pmatrix} V_2 & C_2 & \cdots & C_2 \\ & V_2 & \ddots & \vdots \\ & & \ddots & C_2 \\ & & & V_2 \end{pmatrix}$,

$A_{21} = \begin{pmatrix} C_3 & \cdots & C_3 \\ & \ddots & \vdots \\ & & C_3 \end{pmatrix}$,

Therefore,

$$Cov(\underset{\sim}{Y}_i) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \text{ and } Cov(\underset{\sim}{Y}_i)^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \quad (27)$$

where,

$$B_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$$

$$= \frac{1}{e^{\zeta_0+\sigma^2/2}} I_{n_{i1}\times n_{i1}} - \frac{e^{\sigma^2}-1}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)} J_{n_{i1}\times n_{i1}},$$

$$B_{12} = -A_{11}^{-1}A_{12}B_{22}$$

$$= -\frac{e^{\sigma^2}-1}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)} J_{n_{i1}\times n_{i2}},$$

$$B_{21} = -A_{22}^{-1}A_{21}B_{11}$$

$$= -\frac{e^{\sigma^2}-1}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)} J_{n_{i2}\times n_{i1}}$$

$$B_{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$$

$$= \frac{1}{e^{\zeta_0+\zeta_1+\sigma^2/2}} I_{n_{i2}\times n_{i2}} - \frac{e^{\sigma^2}-1}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)} J_{n_{i2}\times n_{i2}}.$$

Finally, we have

$$E(\underset{\sim}{Y}_i) = E(E(\underset{\sim}{Y}_i | S_i))$$

$$= E(e^{\zeta_0+S_i}, \cdots, e^{\zeta_0+S_i}, e^{\zeta_0+\zeta_1+S_i}, \cdots, e^{\zeta_0+\zeta_1+S_i})'$$

$$= (e^{\zeta_0+\sigma^2/2}\underset{\sim}{1}'_{n_{i1}\times 1}, e^{\zeta_0+\zeta_1+\sigma^2/2}\underset{\sim}{1}'_{n_{i2}\times 1})'.$$

Now that (25) can be written as:

$$\begin{aligned}\mathrm{BLP}(\tau_i) = e^{\sigma^2/2} + e^{\sigma^2}(e^{\sigma^2}-1)&(e^{\zeta_0}\underset{\sim}{1}'_{n_{i1}\times 1}, e^{\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1})\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22}\end{pmatrix} \\ &\times\left(\underset{\sim}{y}_i - (e^{\zeta_0+\sigma^2/2}\underset{\sim}{1}'_{n_{i1}\times 1}, e^{\zeta_0+\zeta_1+\sigma^2/2}\underset{\sim}{1}'_{n_{i2}\times 1})'\right)\end{aligned}$$

$$= e^{\sigma^2/2}\left[\frac{1/n_{i\cdot} + \bar{Y}_{i\cdot\cdot}(e^{\sigma^2}-1)}{1/n_{i\cdot} + (w_{i1}e^{\zeta_0+\sigma^2/2} + w_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}\right] \qquad (28)$$

where $w_{ij} = n_{ij} / n_{i\cdot}$ ($j = 1, 2$). However, since the parameters $(\zeta_0, \zeta_1, \sigma^2)$ are unknown, we have to use the empirical best linear predictor (eBLP) of τ_i instead. Let $(\hat{\zeta}_{0,n}, \hat{\zeta}_{1,n}, \hat{\sigma}_n^2)$ denote the unconditional MLEs obtained by maximizing the full integrated likelihood

$$L_n(\zeta_0, \zeta_1, \sigma^2) = \prod_{i=1}^{m} \left\{ \int_{-\infty}^{\infty} \exp\left(-n_{i1} e^{\zeta_0 + s_i} - n_{i2} e^{\zeta_0 + \zeta_1 + s_i} + (\zeta_0 + s_i) y_{i1\cdot} + (\zeta_0 + \zeta_1 + s_i) y_{i2\cdot} \right) \right.$$
$$\left. \times \frac{1}{\prod_{k=1}^{n_{i1}} y_{i1k}! \prod_{k=1}^{n_{i2}} y_{i2k}!} \times \frac{\exp(-s_i^2 / 2\sigma^2)}{\sqrt{2\pi\sigma^2}} ds_i \right\} \tag{29}$$

which is the same as (24) except it is now viewed as function of all three parameters. Again we use M point Gaussian Hermite quadrature rule for the purpose of evaluating the integrated likelihood. If we let $Q_i = \sum_{t=1}^{M} R_{it} w_t$, where,

$$R_{it} = \exp\left(-n_{i1} e^{\zeta_0 + \sqrt{2}\sigma x_t} - n_{i2} e^{\zeta_0 + \zeta_1 + \sqrt{2}\sigma x_t} + (\zeta_0 + \sqrt{2}\sigma x_t) y_{i1\cdot} + (\zeta_0 + \zeta_1 + \sqrt{2}\sigma x_t) y_{i2\cdot} \right),$$

then the logarithm of the full integrated likelihood can be approximated by:

$$l(\zeta_0, \zeta_1, \sigma^2) \approx \sum_{i=1}^{m} \log Q_i - m \log(\sqrt{\pi}) - C$$

where x_t are the roots of Hermite polynomial, w_t are the associated weights, and C is a constant term in the log-likelihood function. For use with Newton's method to get the unconditional MLEs, the first order derivatives with respect to ζ_0 , ζ_1 and σ are calculated as follows:

$$\frac{\partial L(\zeta_0,\zeta_1,\sigma^2)}{\partial \zeta_0}=\sum_{i=1}^{m}\left\{\frac{1}{Q_i}\sum_{t=1}^{M}\left(-n_{i1}e^{\zeta_0+\sqrt{2}\sigma x_t}-n_{i2}e^{\zeta_0+\zeta_1+\sqrt{2}\sigma x_t}y_{i1\cdot}+y_{i2\cdot}\right)R_{it}w_t\right\}$$

$$\frac{\partial L(\zeta_0,\zeta_1,\sigma^2)}{\partial \zeta_1}=\sum_{i=1}^{m}\left\{\frac{1}{Q_i}\sum_{t=1}^{M}\left(-n_{i2}e^{\zeta_0+\zeta_1+\sqrt{2}\sigma x_t}+y_{i2\cdot}\right)R_{it}w_t\right\}$$

$$\frac{\partial L(\zeta_0,\zeta_1,\sigma^2)}{\partial \sigma}=\sum_{i=1}^{m}\left\{\frac{1}{Q_i}\sum_{t=1}^{M}\left(-n_{i1}e^{\zeta_0+\sqrt{2}\sigma x_t}-n_{i2}e^{\zeta_0+\zeta_1+\sqrt{2}\sigma x_t}+y_{i1\cdot}+y_{i2\cdot}\right)\times\sqrt{2}x_tR_{it}w_t\right\}$$

The eBLP of τ_i is then

$$\text{eBLP}(\tau_i)=e^{\hat{\sigma}_n^2/2}\left[\frac{1/n_{i\cdot}+\overline{Y}_{i\cdot\cdot}(e^{\hat{\sigma}_n^2}-1)}{1/n_{i\cdot}+(w_{i1}e^{\hat{\zeta}_{0,n}+\hat{\sigma}_n^2/2}+w_{i2}e^{\hat{\zeta}_{0,n}+\hat{\zeta}_{1,n}+\hat{\sigma}_n^2/2})(e^{\hat{\sigma}_n^2}-1)}\right] \tag{30}$$

Under the Poisson GLMM set up defined by (23), the eBLP(τ_i) defined by (30) is a consistent predictor of τ_i under certain conditions that we discuss in section 2.4.2. We note in passing that the eBLP of s_i is not consistent, so one would not want to overlook the fact that λ_n^* depends on s_i through the $\tau_i=\exp(s_i)$ quantities. Let $\hat{\tau}_{i,n}$ denote consistent predictors of $\tau_i=\exp(s_i)$. Our proposed sequential test, which we refer to as the GLMM-Bartlett SPRT, for testing $\text{H}_0:\zeta_1=c_0$ vs $\text{H}_1:\zeta_1=c_1$ is based on the log likelihood ratio

$$\Lambda_n^*=$$

$$\ln\left\{\frac{\prod_{i=1}^{m}\left\{\exp\left(-\hat{\tau}_{i,n}(n_{i1}e^{\hat{\zeta}_{0,n}(c_1)}+n_{i2}e^{\hat{\zeta}_{0,n}(c_1)+c_1})+\hat{\zeta}_{0,n}(c_1)y_{i1\cdot}+(\hat{\zeta}_{0,n}(c_1)+c_1)y_{i2\cdot}\right)/\prod_{k=1}^{n_{i1}}y_{i1k}!\prod_{k=1}^{n_{i2}}y_{i2k}!\right\}}{\prod_{i=1}^{m}\left\{\exp\left(-\hat{\tau}_{i,n}(n_{i1}e^{\hat{\zeta}_{0,n}(c_0)}+n_{i2}e^{\hat{\zeta}_{0,n}(c_0)+c_0})+\hat{\zeta}_{0,n}(c_0)y_{i1\cdot}+(\hat{\zeta}_{0,n}(c_0)+c_0)y_{i2\cdot}\right)/\prod_{k=1}^{n_{i1}}y_{i1k}!\prod_{k=1}^{n_{i2}}y_{i2k}!\right\}}\right\} \tag{31}$$

with the usual decision criteria which accepts H_0 at the first n for which $\Lambda_n^* \le A$, accepts H_1 at the first n for which $\Lambda_n^* \ge B$, and continues by sampling another observation if $A < \Lambda_n^* < B$. $\hat{\zeta}_{0,n}(c_1)$ and $\hat{\zeta}_{0,n}(c_0)$ are the conditional MLEs of ζ_0, given $\zeta_1 = c_1$ and $\zeta_0 = c_0$, respectively. Although the constant term $\prod_{k=1}^{n_{i1}} y_{i1k}! \prod_{k=1}^{n_{i2}} y_{i2k}!$ could be canceled out in the right hand side of (31) we found that keeping it leads to a more computationally stable form for Λ_n^*.

2.4.2. *The Consistency of* eBLP(τ_i)

In this section, we discuss the consistency of eBLP(τ_i) which our proposed GLMM-Bartlett SPRT procedure depends heavily on. We prove a theorem on the consistency of eBLP(τ_i) under certain conditions.

Theorem 2. Under the Poisson GLMM defined by (23), the eBLP(τ_i) defined by (30) is a consistent predictor of τ_i [i.e., $\text{eBLP}(\tau_i) \xrightarrow{p} \tau_i$] as $w_{i1} = n_{i1} / n_{i\cdot} \to q_i$ provided that $(\hat{\zeta}_{0,n}, \hat{\zeta}_{1,n}, \hat{\sigma}_n^2)$ is a consistent estimator of $(\zeta_0, \zeta_1, \sigma^2)$.

Proof. First we write the prediction error associated with eBLP(τ_i) as the sum of two terms:

$$\text{eBLP}(\tau_i) - \tau_i = [\text{eBLP}(\tau_i) - \text{BLP}(\tau_i)] + [\text{BLP}(\tau_i) - \tau_i].$$

Considering the second term first, the Mean Square Error (MSE) of τ_i is equal to

$$\text{MSE}(\text{BLP}(\tau_i)) = Var(\tau_i) - Cov(\underset{\sim}{Y}_i, \tau_i)'\left[Cov(\underset{\sim}{Y}_i)\right]^{-1} Cov(\underset{\sim}{Y}_i, \tau_i)\,. \tag{32}$$

It can be shown that

$$Var(\tau_i) = E\left(Var(\tau_i|S_i)\right) + Var\left(E(\tau_i|S_i)\right)$$

$$= e^{\sigma^2}(e^{\sigma^2} - 1)\,.$$

Recall from (26) and (27) that

$$Cov(\underset{\sim}{Y}_i, \tau_i) = e^{\sigma^2}(e^{\sigma^2} - 1)(e^{\zeta_0}\underset{\sim}{1}'_{n_{i1}\times 1}, e^{\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1})'\,,$$

and $Cov(\underset{\sim}{Y}_i)^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$.

Now we have all the individual parts of (32) in tractable forms. Replace those parts in (32) we have

$$\begin{aligned}\text{MSE}(\text{BLP}(\tau_i)) &= e^{\sigma^2}(e^{\sigma^2} - 1) - e^{2\sigma^2}(e^{\sigma^2} - 1)^2(e^{\zeta_0}\underset{\sim}{1}'_{n_{i1}\times 1}, e^{\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1})\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}(e^{\zeta_0}\underset{\sim}{1}'_{n_{i1}\times 1}, e^{\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1})' \\ &= e^{\sigma^2}(e^{\sigma^2} - 1) - e^{2\sigma^2}(e^{\sigma^2} - 1)^2 \times \Big(e^{2\zeta_0}\underset{\sim}{1}'_{n_{i1}\times 1}B_{11}\underset{\sim}{1}_{n_{i1}\times 1} + e^{2\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1}B_{21}\underset{\sim}{1}_{n_{i1}\times 1} \\ &\quad + e^{2\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i1}\times 1}B_{12}\underset{\sim}{1}_{n_{i2}\times 1} + e^{2\zeta_0+2\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1}B_{22}\underset{\sim}{1}_{n_{i2}\times 1}\Big)\end{aligned}$$

Now let

$$P_1 = e^{2\zeta_0}\underset{\sim}{1}'_{n_{i1}\times 1}B_{11}\underset{\sim}{1}_{n_{i1}\times 1},$$

$$= e^{2\zeta_0}\underset{\sim}{1}'_{n_{i1}\times 1}\left[\frac{1}{e^{\zeta_0+\sigma^2/2}}I_{n_{i1}\times n_{i1}} - \frac{e^{\sigma^2}-1}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}J_{n_{i1}\times n_{i1}}\right]\underset{\sim}{1}_{n_{i1}\times 1}$$

$$= n_{i1}e^{2\zeta_0}\left[e^{-(\zeta_0+\sigma^2/2)} - \frac{n_{i1}(e^{\sigma^2}-1)}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}\right]$$

$$P_2 = e^{2\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1} B_{21}\underset{\sim}{1}_{n_{i1}\times 1},$$

$$= e^{2\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1}\frac{-(e^{\sigma^2}-1)}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}J_{n_{i2}\times n_{i1}}\underset{\sim}{1}_{n_{i1}\times 1}$$

$$= \frac{-n_{i1}n_{i2}e^{2\zeta_0+\zeta_1}(e^{\sigma^2}-1)}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}$$

$$P_3 = e^{2\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i1}\times 1} B_{12}\underset{\sim}{1}_{n_{i2}\times 1},$$

$$= e^{2\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i1}\times 1}\frac{-(e^{\sigma^2}-1)}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}J_{n_{i1}\times n_{i2}}\underset{\sim}{1}_{n_{i2}\times 1}$$

$$= \frac{-n_{i1}n_{i2}e^{2\zeta_0+\zeta_1}(e^{\sigma^2}-1)}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}$$

$$P_4 = e^{2\zeta_0+2\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1} B_{22}\underset{\sim}{1}_{n_{i2}\times 1},$$

$$= e^{2\zeta_0+2\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1}\left[\frac{1}{e^{\zeta_0+\zeta_1+\sigma^2/2}}I_{n_{i2}\times n_{i2}} - \frac{e^{\sigma^2}-1}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}J_{n_{i2}\times n_{i2}}\right]\underset{\sim}{1}_{n_{i2}\times 1}$$

$$= n_{i2}e^{2\zeta_0+2\zeta_1}\left[e^{-(\zeta_0+\zeta_1+\sigma^2/2)} - \frac{n_{i2}(e^{\sigma^2}-1)}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}\right]$$

$$P_1+P_2+P_3+P_4 = n_{i1}e^{2\zeta_0}\left[e^{-(\zeta_0+\sigma^2/2)} - \frac{n_{i1}(e^{\sigma^2}-1)}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}\right]$$

$$+n_{i2}e^{2\zeta_0+2\zeta_1}\left[e^{-(\zeta_0+\zeta_1+\sigma^2/2)} - \frac{n_{i2}(e^{\sigma^2}-1)}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}\right]$$

$$-\frac{2n_{i1}n_{i2}e^{2\zeta_0+\zeta_1}(e^{\sigma^2}-1)}{1+(n_{i1}e^{\zeta_0+\sigma^2/2}+n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}$$

$$= \frac{n_{i1}e^{\zeta_0} + n_{i2}e^{\zeta_0+\zeta_1}}{e^{\sigma^2/2}\left(1 + n_{i1}e^{\zeta_0+\sigma^2/2}(e^{\sigma^2}-1) + n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2}(e^{\sigma^2}-1)\right)} \tag{33}$$

It follows that

$$\text{MSE}(\text{BLP}(\tau_i)) = e^{\sigma^2}(e^{\sigma^2}-1) - e^{2\sigma^2}(e^{\sigma^2}-1)^2 \times \left(P_1 + P_2 + P_3 + P_4\right)$$

$$= e^{\sigma^2}(e^{\sigma^2}-1) - \frac{e^{2\sigma^2}(e^{\sigma^2}-1)^2(n_{i1}e^{\zeta_0} + n_{i2}e^{\zeta_0+\zeta_1})}{e^{\sigma^2/2}\left(1 + n_{i1}e^{\zeta_0+\sigma^2/2}(e^{\sigma^2}-1) + n_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2}(e^{\sigma^2}-1)\right)},$$

let $w_{ij} = n_{ij}/n_{i\cdot}$ $(j = 1, 2)$, it can be shown that the mean squared error (MSE) of $\text{BLP}(\tau_i)$ is

$$\text{MSE}[\text{BLP}(\tau_i)] = \frac{e^{\sigma^2}(e^{\sigma^2}-1)/n_{i\cdot}}{1/n_{i\cdot} + (w_{i1}e^{\zeta_0+\sigma^2/2} + w_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}.$$

Since $P\left(\left|\text{BLP}(\tau_i) - \tau_i\right| > \varepsilon\right) < \text{MSE}[\text{BLP}(\tau_i)]/\varepsilon^2$, it follows that $\text{BLP}(\tau_i) \xrightarrow{p} \tau_i$ as $n_{i\cdot} \to \infty$.

It suffices to show the first term similarly converges in probability to zero. It is easily verified that

$$\text{eBLP}(\tau_i) - \text{BLP}(\tau_i) = \left\{\left[\frac{e^{\hat{\sigma}_n^2/2}/n_{i\cdot}}{1/n_{i\cdot} + (w_{i1}e^{\hat{\zeta}_{0,n}+\hat{\sigma}_n^2/2} + w_{i2}e^{\hat{\zeta}_{0,n}+\hat{\zeta}_{1,n}+\hat{\sigma}_n^2/2})(e^{\hat{\sigma}_n^2}-1)}\right]\right.$$

$$\left. - \left[\frac{e^{\sigma^2/2}/n_{i\cdot}}{1/n_{i\cdot} + (w_{i1}e^{\zeta_0+\sigma^2/2} + w_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}\right]\right\}$$

$$+ \left\{\left[\frac{e^{\hat{\sigma}_n^2/2}(e^{\hat{\sigma}_n^2}-1)}{1/n_{i\cdot} + (w_{i1}e^{\hat{\zeta}_{0,n}+\hat{\sigma}_n^2/2} + w_{i2}e^{\hat{\zeta}_{0,n}+\hat{\zeta}_{1,n}+\hat{\sigma}_n^2/2})(e^{\hat{\sigma}_n^2}-1)}\right]\right.$$

$$\left. - \left[\frac{e^{\sigma^2/2}(e^{\sigma^2}-1)}{1/n_{i\cdot} + (w_{i1}e^{\zeta_0+\sigma^2/2} + w_{i2}e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2}-1)}\right]\right\}\bar{Y}_{i\cdot\cdot}$$

Noting that $\bar{Y}_{i\cdot\cdot} = w_{i1}\bar{Y}_{i1\cdot} + w_{i2}\bar{Y}_{i2\cdot}$, and let $q_i = \frac{n_{i1}}{n_{i\cdot}}$, it follows from the strong law of large numbers that $\bar{Y}_{i\cdot\cdot}\big|S_i = s_i \xrightarrow{a.s.} q_i e^{\zeta_0+s_i} + (1-q_i)e^{\zeta_0+\zeta_1+s_i}$. It follows that unconditionally $\bar{Y}_{i\cdot\cdot} \xrightarrow{a.s.} q_i e^{\zeta_0+S_i} + (1-q_i)e^{\zeta_0+\zeta_1+S_i}$. Provided that $(\hat{\zeta}_{0,n}, \hat{\zeta}_{1,n}, \hat{\sigma}_n^2)$ is a consistent estimator of $(\zeta_0, \zeta_1, \sigma^2)$, by Slutzky's theorem the two terms in the braces on the right-hand side converge to zero in probability, and thus $\text{eBLP}(\tau_i) \xrightarrow{p} \text{BLP}(\tau_i)$ as was claimed.

In the proof above, a very important assumption is that $(\hat{\zeta}_{0,n}, \hat{\zeta}_{1,n}, \hat{\sigma}_n^2)$ is a consistent estimator of $(\zeta_0, \zeta_1, \sigma^2)$. The necessary condition for this consistency is that the number of centers has the control group patients in it goes to infinity, and the number of centers has the treatment group patients in it goes to infinity.

We discussed the fact that λ_n^* depends on s_i through the $\tau_i = \exp(s_i)$ quantities and the inconsistency of $\text{eBLP}(s_i)$ briefly in section 2.4.1. Now we also verify the above statement by calculating the MSE($\text{BLP}(s_i)$).

$$\text{MSE}(\text{BLP}(s_i)) = Var(S_i) - Cov(\underset{\sim}{Y}_i, S_i)'\left[Cov(\underset{\sim}{Y}_i)\right]^{-1} Cov(\underset{\sim}{Y}_i, S_i)$$

$$Cov(\underset{\sim}{Y}_i, S_i) = Cov\left(E(\underset{\sim}{Y}_i\big|S_i), E(S_i\big|S_i)\right) + E\left(Cov(\underset{\sim}{Y}_i, S_i\big|S_i)\right)$$

$$= Cov\left(E(\underset{\sim}{Y}_i\big|S_i), E(S_i\big|S_i)\right)$$

$$= Cov\left((e^{\zeta_0+S_i}, \cdots, e^{\zeta_0+S_i}, e^{\zeta_0+\zeta_1+S_i}, \cdots, e^{\zeta_0+\zeta_1+S_i})', S_i\right)$$

$$= E(e^{S_i}S_i) \times (e^{\zeta_0}\underset{\sim}{1}'_{n_{i1}\times 1}, e^{\zeta_0+\zeta_1}\underset{\sim}{1}'_{n_{i2}\times 1})'$$

Since $E(e^{S_i}S_i) = \int e^{s_i} s_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{s_i^2}{2\sigma^2}} ds_i$

$$= \int s_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{s_i^2 - 2\sigma^2 s_i + \sigma^4}{2\sigma^2}} e^{\frac{\sigma^4}{2\sigma^2}} ds_i$$

$$= e^{\frac{\sigma^4}{2\sigma^2}} \int s_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s_i - \sigma^2)^2}{2\sigma^2}} ds_i$$

$$= \sigma^2 e^{\frac{\sigma^2}{2}},$$

$Cov(\underset{\sim}{Y}_i, S_i) = \sigma^2 e^{\frac{\sigma^2}{2}} (e^{\zeta_0} \underset{\sim}{1}'_{n_{i1}\times 1}, e^{\zeta_0+\zeta_1} \underset{\sim}{1}'_{n_{i2}\times 1})'$.

Once again using equation (27) and (32), we have:

$$\mathrm{MSE}\left(\mathrm{BLP}(s_i)\right) = \sigma^2 - \sigma^4 e^{\sigma^2} (e^{\zeta_0} \underset{\sim}{1}'_{n_{i1}\times 1}, e^{\zeta_0+\zeta_1} \underset{\sim}{1}'_{n_{i2}\times 1}) \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} (e^{\zeta_0} \underset{\sim}{1}'_{n_{i1}\times 1}, e^{\zeta_0+\zeta_1} \underset{\sim}{1}'_{n_{i2}\times 1})'$$

$$= \sigma^2 - \sigma^4 e^{\sigma^2} \times \Big(e^{2\zeta_0} \underset{\sim}{1}'_{n_{i1}\times 1} B_{11} \underset{\sim}{1}_{n_{i1}\times 1} + e^{2\zeta_0+\zeta_1} \underset{\sim}{1}'_{n_{i2}\times 1} B_{21} \underset{\sim}{1}_{n_{i1}\times 1} + e^{2\zeta_0+\zeta_1} \underset{\sim}{1}'_{n_{i1}\times 1} B_{12} \underset{\sim}{1}_{n_{i2}\times 1}$$
$$+ e^{2\zeta_0+2\zeta_1} \underset{\sim}{1}'_{n_{i2}\times 1} B_{22} \underset{\sim}{1}_{n_{i2}\times 1} \Big).$$

Based on (33), we simplify $\mathrm{MSE}\left(\mathrm{BLP}(s_i)\right)$ into:

$$\mathrm{MSE}\left(\mathrm{BLP}(s_i)\right) = \sigma^2 - \frac{\sigma^4 e^{\sigma^2} (w_{i1} e^{\zeta_0} + w_{i2} e^{\zeta_0+\zeta_1})}{1/n_{i\cdot} + (w_{i1} e^{\zeta_0+\sigma^2/2} + w_{i2} e^{\zeta_0+\zeta_1+\sigma^2/2})(e^{\sigma^2} - 1)}$$

with $w_{ij} = n_{ij} / n_{i\cdot}$ $(j = 1, 2)$. It follows that $\mathrm{BLP}(s_i) \xrightarrow{p} \sigma^2 (1 - \frac{\sigma^2 e^{\sigma^2/2}}{e^{\sigma^2} - 1})$ as $n_{i\cdot} \to \infty$. The MSE of $\mathrm{BLP}(s_i)$ does not converge in probability to zero and $\mathrm{BLP}(s_i)$ is not a consistent predictor of s_i.

2.4.3. *Performance of GLMM-Bartlett SPRT*

We use OC curve and ASN curve to evaluate the performance of our proposed GLMM-Bartlett SPRT. Our simulation was conducted by assigning a number to each arriving patient within each of the four participating centers, representing their order of arrival. Odd numbered patients were then randomly assigned to a treatment by the flip of a coin; the next even numbered patient was then assigned the alternate treatment to maintain a balance between the total numbers of patients assigned to each group at that center.

When ζ_0 and σ^2 are unknown, our proposed GLMM-Bartlett SPRT is implemented using (31). The OC and ASN of the proposed procedure were investigated with the same simulation setup used when evaluating the Wald SPRT and the results are shown in **Table 3**.

ζ_1	$(\zeta_0, \sigma) = (2.5, 0.2)$							$(\zeta_0, \sigma) = (2.5, 0.4)$						
		Sample Size Distribution							Sample Size Distribution					
	OC	ASN	10th	25th	50th	75th	90th	OC	ASN	10th	25th	50th	75th	90th
-0.4	0	26.33	13	17	23	32	43	0	25.52	12	16	22	31	42
-0.35	0.003	31.82	15	20	28	39	54	0.003	30.40	13	18	26	38	53
-0.3	0.010	40.95	17	23	34	52	74	0.007	39.10	15	22	32	49	71
-0.25	0.035	54.39	19	28	44	70	104	0.036	54.00	18	26	42	69	105
-0.2	0.161	77.96	23	36	60	101	160	0.164	73.49	20	33	57	96	148
-0.15	0.503	90.49	24	39	69	120	189	0.498	83.40	22	37	65	109	169
-0.1	0.843	74.65	21	34	58	98	152	0.841	70.91	20	32	54	91	146
-0.05	0.961	50.87	17	26	40	65	97	0.964	48.48	15	23	38	62	94
0	0.991	35.62	14	20	30	45	64	0.994	34.02	13	19	28	43	62
0.05	0.998	26.52	12	16	23	33	46	0.999	26.04	11	15	23	33	46
0.1	1	21.01	10	14	19	26	34	0.999	20.40	9	13	18	25	35
0.15	1	16.94	9	12	15	21	27	1	16.76	8	11	15	21	27

Table 3. OC and ASN properties of Proposed GLMM-Bartlett's SPRT for $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ when nominal type-1 and type-2 error rates are 0.01

The realized type-1 and type-2 error rates for the GLMM-Bartlett's SPRT are satisfactorily close to the nominal rates. Just as we saw for both the fixed sample size and Wald SPRT procedures, the difference in the OC and ASN characteristics for the two cases $\sigma = 0.2$ and $\sigma = 0.4$ is not large. However, we do see that a slightly higher sample size is required for the case $\sigma = 0.2$ and this result is reflecting the fact that the variance of the MLE of ζ_1 is a decreasing function of σ. Comparing **Table 2** and **Table 3**, we see that the sample size required using the GLMM-Bartlett's SPRT is slightly larger than the sample size required with Wald's SPRT. This is expected, since it is natural there would be a penalty (taking the form of a larger required sample size) for having to estimate (ζ_0, σ).

Figure 5 includes an illustrative comparison of the savings in sampling using the GLMM-Bartlett SPRT, Wald SPRT as compared to the fixed sample size NP test. The ASN curves vividly portray the penalty of larger sample size required for GLMM-Bartlett SPRT than Wald's SPRT. What is also notable about the comparison is that relative to simpler models, the region where the fixed sample size plan does better than the SPRTs for our GLMM model is narrower.

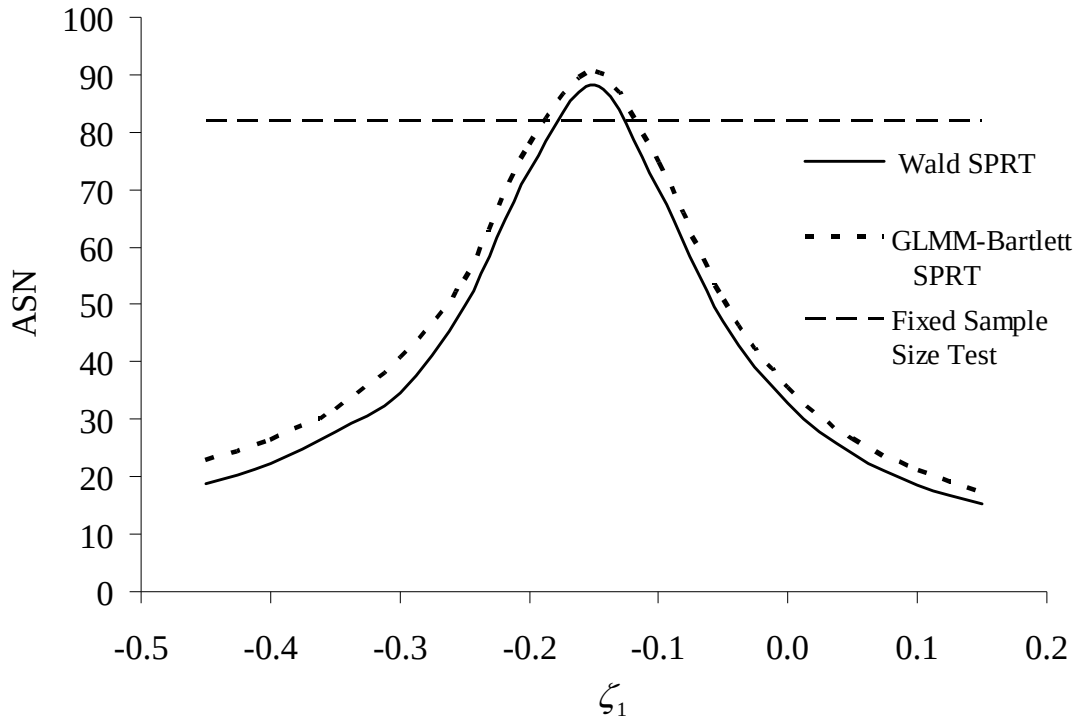

Figure 5. ASN Curve for GLMM-Bartlett SPRT, Wald's SPRT and Fixed Sample Size Test of $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ with $(\zeta_0, \sigma) = (2.5, 0.2)$ and nominal type-1 and type-2 error rates of 0.01

2.4.4. *Computational Approximation*

We suggested using a Gauss-Hermite quadrature rule to evaluate the full integrated likelihood (29) for the GLMM-Bartlett Procedure. To simplify the computation, we also propose a Laplace-like approximation to the integrated likelihood function. Let

$$f(s_i, \zeta_0, \zeta_1) = -n_{i1} e^{\zeta_0 + s_i} - n_{i2} e^{\zeta_0 + \zeta_1 + s_i} + (\zeta_0 + s_i) y_{i1\cdot} + (\zeta_0 + \zeta_1 + s_i) y_{i2\cdot} - \log\left(\prod_{k=1}^{n_{i1}} y_{i1k}!\right) - \log\left(\prod_{k=1}^{n_{i2}} y_{i2k}!\right).$$

The integrated likelihood function (29) then can be written as

$$L(\zeta_0,\zeta_1,\sigma^2)=\prod_{i=1}^{m}\left\{\int_{-\infty}^{\infty}e^{f(s_i,\zeta_0,\zeta_1)}\times\frac{\exp(-s_i^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}ds_i\right\}.$$

Denote the value of s_i that maximizes $f(s_i,\zeta_0,\zeta_1)$ by s_{i0}:

$$s_{i0}=\log\left(\frac{y_{i\cdot\cdot}}{n_{i1}e^{\zeta_0}+n_{i2}e^{\zeta_0+\zeta_1}}\right).$$

A second order Taylor series expansion of $f(s_i,\zeta_0,\zeta_1)$ about the point $s_i=s_{i0}$ gives

$$f(s_i,\zeta_0,\zeta_1)\approx f(s_{i0},\zeta_0,\zeta_1)+\left.\frac{\partial f(s_i,\zeta_0,\zeta_1)}{s_i}\right|_{s_{i0}}(s_i-s_{i0})+\frac{1}{2}\left.\frac{\partial^2 f(s_i,\zeta_0,\zeta_1)}{\partial^2 s_i}\right|_{s_{i0}}(s_i-s_{i0})^2$$

$$\left.\frac{\partial f(s_i,\zeta_0,\zeta_1)}{s_i}\right|_{s_{i0}}=y_{i\cdot\cdot}-n_{i1}e^{\zeta_0+s_{i0}}-n_{i2}e^{\zeta_0+\zeta_1+s_{i0}}=0 \tag{34}$$

$$\left.\frac{\partial^2 f(s_i,\zeta_0,\zeta_1)}{\partial^2 s_i}\right|_{s_{i0}}=-n_{i1}e^{\zeta_0+s_{i0}}-n_{i2}e^{\zeta_0+\zeta_1+s_{i0}}=-y_{i\cdot\cdot} \tag{35}$$

By utilizing (34) and (35), we can compute the following result:

$$\exp\{f(s_i,\zeta_0,\zeta_1)\}\approx\exp\left\{f(s_{i0},\zeta_0,\zeta_1)-\frac{1}{2}y_{i\cdot\cdot}(s_i-s_{i0})^2\right\} \tag{36}$$

Since s_{i0} does not depend on s_i, substituting (36) we have the approximated likelihood as

$$L_n(\zeta_0,\zeta_1,\sigma^2)\approx\prod_{i=1}^{m}\left\{\int_{-\infty}^{\infty}e^{f(s_{i0},\zeta_0,\zeta_1)-y_{i\cdot\cdot}(s_i-s_{i0})^2/2}\times\frac{\exp(-s_i^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}ds_i\right\}.$$

$$=\prod_{i=1}^{m}\left\{\frac{e^{f(s_{i0},\zeta_0,\zeta_1)}}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}e^{-\left(\frac{1}{2\sigma^2}+\frac{y_{i\cdot\cdot}}{2}\right)\left(s_i-\frac{y_{i\cdot\cdot}s_0}{1/\sigma^2+y_{i\cdot\cdot}}\right)^2}\times e^{\left(\frac{1}{2\sigma^2}+\frac{y_{i\cdot\cdot}}{2}\right)\left(\frac{y_{i\cdot\cdot}s_0}{1/\sigma^2+y_{i\cdot\cdot}}\right)^2-\frac{y_{i\cdot\cdot}s_0^2}{2}}ds_i\right\}$$

$$= \prod_{i=1}^{m} \frac{e^{f(s_{i0},\zeta_0,\zeta_1)}}{\sqrt{1+\sigma^2 y_{i\cdot\cdot}}} \times \exp\left(-\frac{s_{i0}^2 y_{i\cdot\cdot}}{2(1+\sigma^2 y_{i\cdot\cdot})} \right) \tag{37}$$

where,

$$e^{f(s_{i0},\zeta_0,\zeta_1)} = \exp\Big\{ -n_{i1} e^{\zeta_0 + s_{i0}} - n_{i2} e^{\zeta_0+\zeta_1+s_{i0}} + (\zeta_0 + s_{i0}) y_{i1\cdot} + (\zeta_0 + \zeta_1 + s_{i0}) y_{i2\cdot}$$

$$- \log(\prod_{k=1}^{n_{i1}} y_{i1k}!) - \log(\prod_{k=1}^{n_{i2}} y_{i2k}!) \Big\}$$

Consequently, (37) can be simplified into

$$L_n(\zeta_0,\zeta_1,\sigma^2) \approx \prod_{i=1}^{m} \left\{ \frac{e^{-y_{i\cdot\cdot}} y_{i\cdot\cdot}^{y_{i\cdot\cdot}}}{\prod_{k=1}^{n_{i1}} y_{i1k}! \prod_{k=1}^{n_{i1}} y_{i2k}!} \frac{e^{\zeta_0 y_{i1\cdot} + (\zeta_0+\zeta_1) y_{i2\cdot}}}{(n_{i1\cdot} e^{\zeta_0} + n_{i2\cdot} e^{\zeta_0+\zeta_1})^{y_{i\cdot\cdot}}} \frac{e^{-\frac{1}{2}\frac{y_{i\cdot\cdot} s_{0i}^2(\zeta_0,\zeta_1)}{(1+y_{i\cdot\cdot}\sigma^2)}}}{\sqrt{1+y_{i\cdot\cdot}\sigma^2}} \right\}, \tag{38}$$

which, compared to (29), does not require numerical integration. The GLMM-Bartlett SPRT for testing $H_0 : \zeta_1 = c_0$ vs $H_1 : \zeta_1 = c_1$ can thus alternatively be based on the approximation in (38). The log likelihood function for $(\zeta_0, \zeta_1, \sigma^2)$ is then

$$l_n(\zeta_0,\zeta_1,\sigma^2) \propto \sum_{i=1}^{m} \left\{ \zeta_1 y_{i1\cdot} - y_{i\cdot\cdot} \log\left(n_{i1} + n_{i2} e^{\zeta_1}\right) - \frac{1}{2}\log\left(1+\sigma^2 y_{i\cdot\cdot}\right) - \frac{y_{i\cdot\cdot} s_{i0}^2}{2(1+\sigma^2 y_{i\cdot\cdot})} \right\}$$

The first order derivatives with respect to ζ_0, ζ_1 and σ, for use with Newton's method, are given by:

$$\frac{\partial l_n(\zeta_0,\zeta_1,\sigma^2)}{\partial \zeta_0} = \sum_{i=1}^{m} \frac{y_{i\cdot\cdot} \log\left(y_{i\cdot\cdot} / (n_{i1} e^{\zeta_0} + n_{i2} e^{\zeta_0+\zeta_1}) \right)}{1+\sigma^2 y_{i\cdot\cdot}}$$

$$\frac{\partial l_n(\zeta_0,\zeta_1,\sigma^2)}{\partial \zeta_1} = \sum_{i=1}^{m} \left\{ y_{i2\cdot} - \frac{y_{i\cdot\cdot} n_{i2} e^{\zeta_0}}{n_{i1} + n_{i2} e^{\zeta_1}} \left[1 + \frac{\log\left(y_{i\cdot\cdot} / (n_{i1} e^{\zeta_0} + n_{i2} e^{\zeta_0+\zeta_1}) \right)}{1+\sigma^2 y_{i\cdot\cdot}} \right] \right\}$$

$$\frac{\partial l_n(\zeta_0,\zeta_1,\sigma^2)}{\partial \sigma^2} = \frac{1}{2}\sum_{i=1}^{m} -\frac{y_{i\cdot\cdot}}{(1+\sigma^2 y_{i\cdot\cdot})} + \frac{y_{i\cdot\cdot}^2 s_{i0}^2}{(1+\sigma^2 y_{i\cdot\cdot})^2}$$

Figure 6 characterizes a comparison between the exact likelihood of $\exp\{f(s_i,\zeta_0,\zeta_1)\}$ and the proposed approximation made by (36). The figure shows the example of $m=4$, $\zeta_0=2.5$, $\zeta_1=-0.3$, and $\sigma=0.2$, as those parameter values have been used in the previous discussion. Without loss of generality, we also set $n_{i1}=n_{i2}=4$. The exact likelihood and the approximated likelihood are compared as a function of the random effect s_i. As can been seen from figure 6, there is barely a noticeable difference between the two curves. It is worth pointing out that we have also tried several other parameter values for the comparison, and we see the same results.

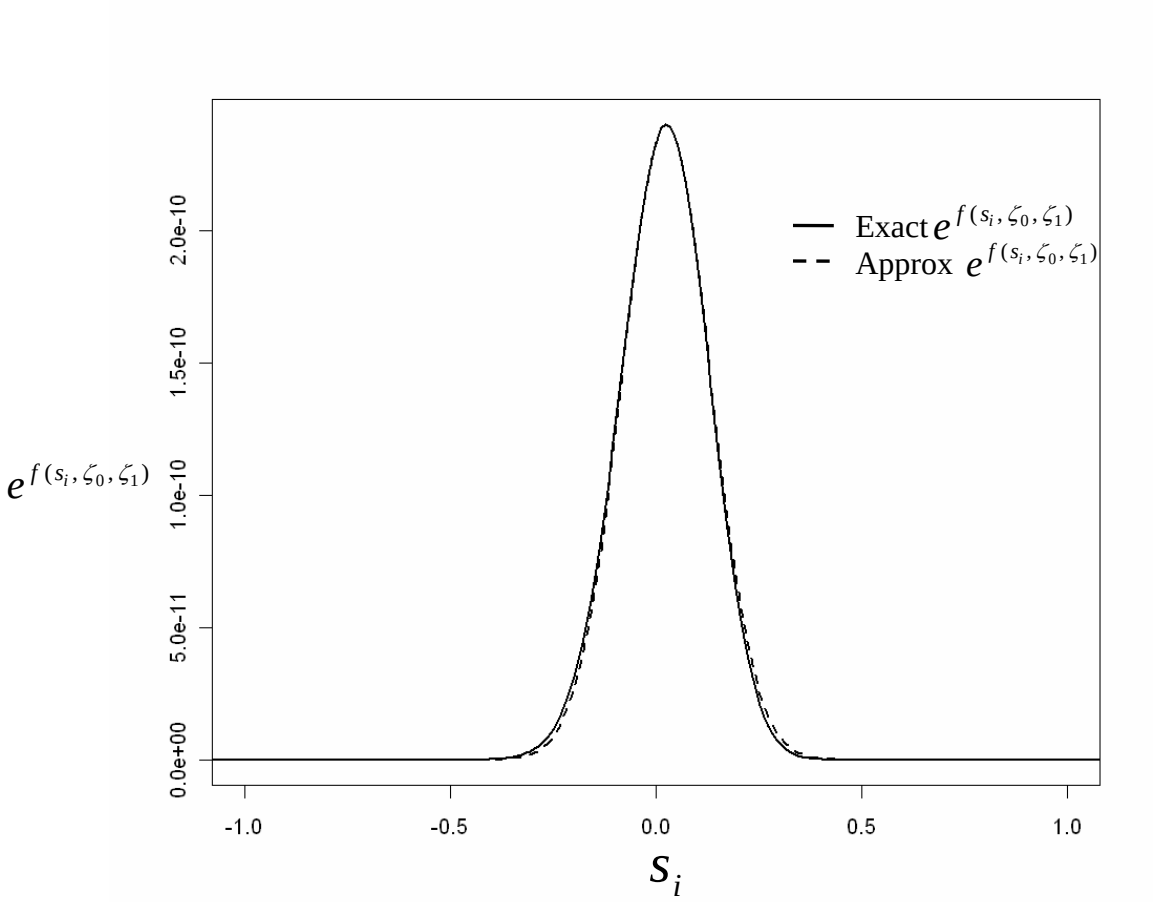

Figure 6. Comparison of Exact and Approximate Likelihood

Table 4 shows a comparison of the results between the GLMM-Bartlett SPRT based on the approximation of the likelihood (38) and the full integrated likelihood (29) with Gauss-Hermite quadrature rule. The OC and ASN of the two procedures were investigated with the same simulation setup used in table 3 for the case where (ζ_0, σ) = (2.5, 0.2). We can see from Table 4 that there is barely a noticeable difference in the OC and ASN characteristics for the two procedures. The simulated type-1 and type-2 error rates for the GLMM-Bartlett procedure based on the approximate likelihood function is (0.008, 0.009), which is comfortably close to the nominal 0.01 values.

ζ_1	GLMM-Bartlett Quadrature							GLMM-Bartlett Approx						
	OC	Sample Size Distribution						OC	Sample Size Distribution					
		ASN	10th	25th	50th	75th	90th		ASN	10th	25th	50th	75th	90th
-0.45	0	22.84	12	15	21	28	36	0	22.48	12	15	20	27	36
-0.4	0	26.33	13	17	23	32	43	0.001	26.51	13	17	23	33	44
-0.35	0.003	31.82	15	20	28	39	54	0.002	31.65	14	20	28	39	54
-0.3	0.010	40.95	17	23	34	52	74	0.008	40.63	17	23	34	51	74
-0.25	0.035	54.39	19	28	44	70	104	0.038	55.84	19	28	45	72	106
-0.2	0.161	77.96	23	36	60	101	160	0.160	78.84	22	35	61	103	158
-0.15	0.503	90.49	24	39	69	120	189	0.489	94.8	25	41.75	75	130	198
-0.1	0.843	74.65	21	34	58	98	152	0.832	76.6	21	34	59	100	154
-0.05	0.961	50.87	17	26	40	65	97	0.965	50.97	17	26	41	66	97
0	0.991	35.62	14	20	30	45	64	0.991	35.02	14	20	29	44	64
0.05	0.998	26.52	12	16	23	33	46	0.999	26.41	12	16	23	33	46
0.1	1	21.01	10	14	19	26	34	1	20.76	10	14	19	25	34
0.15	1	16.94	9	12	15	21	27	1	17.2	9	12	16	21	28

Table 4. OC and ASN properties comparison of GLMM-Bartlett's SPRT with Quadrature rule and approximate likelihood for $H_0 : \zeta_1 = 0$ vs $H_1 : \zeta_1 = -0.3$ when nominal type-1 and type-2 error rates are 0.01 and $(\zeta_0, \sigma) = (2.5, 0.2)$.

CHAPTER 3: Design of Sequential Hypothesis Testing Methodology for Spatial GLMM

3.1. *Motivating Entomology Example*

Generally, pest assessment protocols are required to be economical. When deciding to implement pest control measures, the cost of acquiring the information necessary to make a prudent decision must not outweigh the cost of ignoring a potential pest problem. Often, sequential sampling methods are utilized as a cost effective approach to pest assessment [Fowler and Lynch (1987), Mulekar et. al. (1993), Binns et. al. (2000), Young and Young (1998) and Shah et. al. (2009)]. As discussed by Wald (1947), compared to the most efficient fixed sample size procedures, sequential procedures often require only half as many sample observations.

Our motivating application of this chapter is the assessment of an orchard of fruit-bearing trees for a potential pest problem. Orchards of trees are typically organized into small sized blocks that can be individually treated for pests as necessary. Correctly understanding the spatial distribution of the pest is very important when deciding which blocks to treat. Neglecting the spatial dependencies could result in improper characterization of the pest density and lead to poor decisions about applying a treatment. Geo-statistical analyses have been previously used to describe the spatial distribution of a diverse group of insects of agricultural importance such as *Lygus hesperus* (Hempitera: Miridae) on lentils (Schotzko and O'Keeffe 1989), *Limonius californicus* (Coleoptera: Elateridae) in a corn-alfalfa crop rotation system (Williams et al. 1992), *Helicoverpa*

armigera (Lepidoptera: Noctuidae) on cotton (Gozé et al. 2003), *Lobesia botrana* (Lepidoptera: Tortricidae) (Ifoulis and Savopoulou-Soultani 2006) and *Jacobiasca lybica* (Hemiptera: Cicadellidae) on grapes (Ramírez-Dávila and Porcayo-Camargo 2008). However, in all of the above studies, the spatial analyses are conducted using transformed count data in an attempt to approximately satisfy normality assumptions. The transformation approach has limitations, particularly for count data, and should really only be considered an option for data on a continuous scale measurement (Gotway and Stroup 1997).

GLMM is a class of statistical models that are particularly useful for modeling discrete data (e. g., counts) that may exhibit correlation (Breslow and Clayton 1993). GLMM's have been used across multiple scientific disciplines, including ecology of insect populations (Candy 2000, Elston et. al. 2001, Barchia et. al. 2003, Paterson and Lello 2003, Elias et. al. 2006, Bennett et. al. 2008, Bianchi et. al. 2008, Takakura 2009). Specifically, Elias et. al. 2006 fit a negative binomial GLMM with spatial exponential structure to counts of blacklegged ticks. In this chapter, we integrate the same type of GLMM into a sequential probability ratio test that can be used to efficiently test for pest problems.

The rest of this paper is organized as follows. We illustrate the existence of spatial correlations by analyzing the spatial structure of count data measured on a foliar-feeding mite pest. We propose a sampling methodology for contexts where periodic assessments are to be made to determine if treatment is necessary. Our proposed sampling plan consists of a first occasion fixed sample followed by sequential samples on

each subsequent occasion, and in each case we appropriately address the spatial distribution of pest counts. We summarize the proposed sampling plan by giving a detailed illustration of how it will work from the perspective of a practitioner.

3.1.1. *O. Perseae Data*

In this section, we analyze *Oligonychus perseae* pest count data with a spatial GLMM. Our analysis will illustrate the existence of spatial correlation with respect to this pest. *O. perseae* is an invasive pest that is the most serious foliar pest of avocados in California. Our data was collected during the summer of 2009 on two commercial 'Hass' avocado orchards (A and B). Our two orchards were located in Carpinteria, California, USA and their trees were planted on relatively flat terrain according to a grid system consisting of rows and columns. At each orchard, selected trees in a block were flagged with marking tape; block A from orchard A and block B from orchard B. A map of the trees to be sampled was drawn, and the distance between the bases of tree trunks of each pair of trees was measured in meters. Eight leaves were randomly collected from each tree with 2 leaves taken from each cardinal direction (N, E, S, W). The cardinal directions were determined using a Brunton compact baseplate compass. A total of 30 trees on a 5 x 6 grid were sampled from orchard A and a total of 60 trees on a 5 x 12 grid were sampled from orchard B. Sampled leaves were placed in labeled paper bags and transported in a cooler to a laboratory. Samples were stored in a refrigerator at an average temperature of 3.5 °C until they could be examined to determine the number of *O. perseae* on them. All sampled leaves were examined under a stereoscopic microscope

and all *O. perseae* life stages (except eggs) on the leaf undersurface were counted. Mite counts for each orchard were completed within one week of the sampling date.

3.1.2. *Analysis of O. Perseae Data with Spatial GLMM*

The information compiled from the samples includes the grid location of the tree and the quadrant where the sample leaves were collected. Using PROC GLIMMIX in the SAS software (SAS Institute Inc., 2008) we fit the data to a negative binomial GLMM with a spatial exponential covariance structure. We chose a negative binomial GLMM due to its flexibility in handling over dispersed data and the fact that it includes a Poisson model as a limiting case. Since we expect that the count observations are likely to be correlated spatially depending on the distance between trees, we included a spatially correlated random tree effect in the GLMM link function.

Let Y_{ijk} be the *O. perseae* count of the *k*-th sampled leaf collected from the *j*-th quadrant of the *i*-th tree, with $i = 1, \cdots, n$ (where *n* is 30 for the block on orchard A and 60 for the block on orchard B), $j = 1, \cdots, 4$ and $k = 1, 2$. Let γ_j denote the fixed quadrant effect, S_i denote the spatially correlated random tree effect, and ε_{ij} denote the random interaction effect associated with the *i*-th tree and the *j*-th quadrant. Our GLMM can be summarized as:

$$\begin{aligned} &a. \quad Y_{ijk} \,|\, S_i, \varepsilon_{ij} \overset{ind}{\sim} \text{NegativeBinomial}\ (\mu_{ijk}, \kappa) \\ &b. \quad \log(\mu_{ijk}) = \gamma_j + \varepsilon_{ij} + S_i \\ &c. \quad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2) \\ &d. \quad S_i \sim MVN(0, \Sigma(\sigma_S^2, \theta)) \end{aligned} \tag{39}$$

where κ is a (positive) dispersion parameter. Smaller values of κ correspond to more pronounced clustering of the mites and larger values lead to more Poisson-like data. $\Sigma(\sigma_S^2, \theta)$ is the spatial exponential covariance structure whose (i, i') element is $\sigma_s^2 \exp(-d_{i,i'}/\theta)$, where θ is a scale parameter that dictates the strength of the spatial correlation, and $d_{i,i'}$ is the Euclidean distance between the i-th and i'-th tree.

The estimates of the variance-covariance parameters are shown in **Table 5**. The dispersion parameter estimates clearly suggest there exists over dispersion in the data. The scale parameter estimates indicate significant spatial correlation for the blocks on both orchards. For example, they imply the correlation between adjacent trees is 0.52 in block A, and 0.22 in block B. **Table 5** also shows the estimates of the variance parameter values in the spatial covariance structure and the variance of the tree by quadrant interaction random effect, from which it can be seen that there is more variability in the block A data than in the block B data.

Covariance Parameter	Orchard			
	Block B		Block A	
	Estimates	Standard Error	Estimates	Standard Error
Dispersion Parameter of Negative Binomial Distribution (κ)	2.19	0.04	1.00	0.15
Variance of Spatial Covariance Structure (σ_S^2)	1.42	0.31	2.55	1.72
Scale Parameter in Spatial Covariance Structure (θ)	0.67	0.24	1.55	1.52
Variance of Tree by Quadrant Interaction (σ_ε^2)	0.14	0.05	1.46	0.32

Table 5. Variance Covariance Parameter Estimates for *O. Perseae* Count Data.

An F-test for equal quadrant effects in block A and block B data yield P-values of 0.0024 and 0.0001, respectively, suggesting that the mite counts vary significantly between quadrants in each block. **Table 6** details the estimates of the fixed quadrant effects. For block A, pairwise contrast comparisons show that the South and West quadrants are significantly different from the East and North quadrant, but not significantly different from each other. For block B, the East and South quadrants are significantly different from the North and West quadrant, but not significantly different from each other.

	Orchard	
	Block B	**Block A**
Quadrant Effect	Estimate (log-scale)	Estimate (log-scale)
East	5.32	1.21
North	4.98	1.16
South	5.17	2.22
West	4.88	2.03

Table 6. Analysis of Fixed Effects for *O. Perseae* Count Data

Figure 7 shows 90% tolerance intervals for the estimated conditional distributions of count data based on model (39). The dots are the observed *O. perseae* counts plotted against the fitted conditional mean counts where the random effects have been replaced by their empirical best linear unbiased predictor (which results from using the pseudo-likelihood option with PROC GLIMMIX). The solid lines are the 90% tolerance intervals computed for each observation using the fitted conditional means and the

estimated parameter κ together with the negative binomial distribution. The good capture rate by the tolerance bands suggests that the data fit model (39) very well.

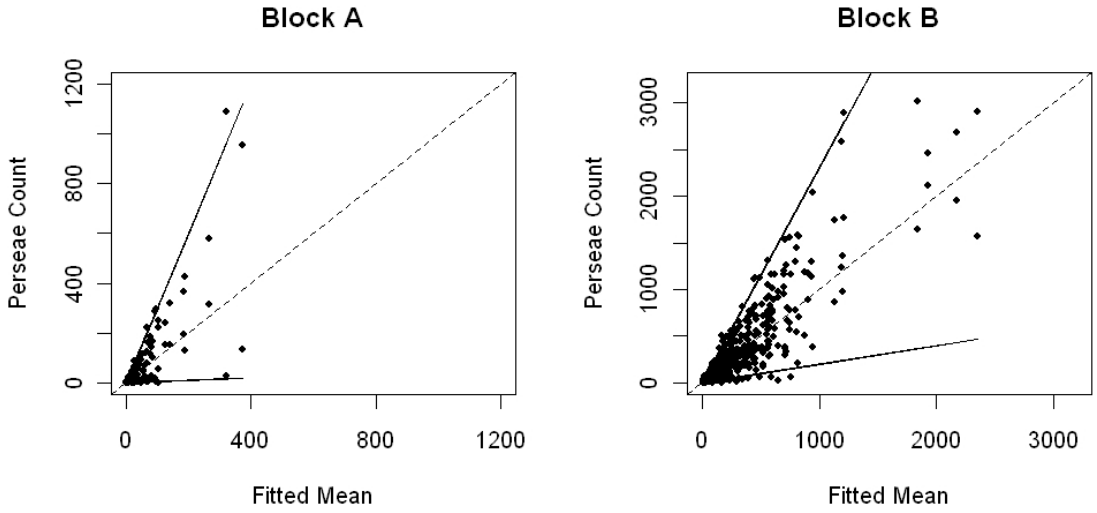

Figure 7. 90% Tolerance Bands (Estimated) for Conditional Negative Binomial Models

3.2. *Proposed Sampling Methodology*

In pest management applications, sequential hypothesis testing has typically been carried out using Wald's SPRT. The goal of Wald's SPRT is to sequentially discriminate between two simple hypotheses about the parameter of interests:

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu = \mu_1 \quad (\mu_1 > \mu_0).$$

Here, μ_0 would represent an acceptable pest density, below which calls for no treatment and μ_1 represents an unacceptable treatment above which calls for treatment. As is discussed in the previous chapters, while providing fixed type-1 and type-2 error rates, Wald's SPRT procedure requires there be no unknown nuisance parameters. When there

are unknown nuisance parameters, such as the dispersion parameter κ of the Negative Binomial distribution, Bartlett proved in the independent and identically distributed (IID) case that a simple modification in the likelihood ratio is sufficient to preserve the asymptotic type-1 and type-2 error properties of the sequential test procedure (1946). Shah et. al. (2009) recently discussed the use of Bartlett's sequential procedure in the context of IID pest count applications. In this section, we propose a periodic sampling methodology that addresses spatially correlated data comprised of a first occasion fixed size sample and subsequent occasion sequential samples. In each instance, H_0 versus H_1 will be tested.

Our proposed procedure begins by collecting a one-time first occasion sample. For the first occasion sample, we suggest the practitioner to sample 2 leaves per tree on all four quadrants from either adjacent trees or every other tree in the middle of the block. The number of trees to be sampled could be 5 to 10 percent of the block size. Information collected from this first occasion sample will suffice to make the first decision about treatment, and will further serve as a pilot sample for setting up the subsequent sequential samples.

Each of the subsequent sequential samples will be used with an SPRT procedure to determine if pest treatment is needed. Based on the analysis of the data in the first occasion sample, a spatial range can be determined so that trees sampled further apart are practically uncorrelated. As a rule of thumb, the so-called practical sampling range is the distance at which the correlation between counts on two trees is reduced to 0.05 (Schabenberger and Gotway 2005). When trees are sampled in such a way that all

pairwise distances exceed the practical sampling range, the sampling is described as "out of range" sampling. Out of range sampling will intuitively lead to less sampling and it also enables a more transparent use of Bartlett's SPRT. We thus propose that for each subsequent sequential sample, we sample one leaf per tree from trees that are out of range relative to each other. The analysis of the data from the first occasion sample will also point to which quadrant is the most infested, and correspondingly we also propose to only sample from that quadrant in the subsequent sequential samples. We suggest starting each sequential sample in one corner of the block and sampling along the diagonal line of the block, repeating if necessary, with the other diagonal. If at the time that both diagonals have been sampled and the sequential procedure still needs to continue, sampling can continue along randomly picked rows until the sequential procedure stops.

Our sampling strategy, as applied to the blocks on orchard A and orchard B would say the following: for block A, sample leaves on either the south quadrant or the west quadrant from every sixth tree; and for block B, sample leaves on either the east quadrant or the south quadrant from every second tree.

Referring to (39), the data collected for subsequent sequential samples follow the simplified GLMM:

$$\begin{aligned} &a. \quad Y_i \,|\, S_i \overset{ind}{\sim} \text{NegativeBinomial}\ (\mu_i, \kappa) \\ &b. \quad \log(\mu_i) = \mu + S_i \\ &c. \quad S_i \overset{iid}{\sim} N(0, \sigma^2) \end{aligned} \tag{40}$$

where Y_i is the pest count recorded from the leaf sampled from the i-th tree. μ_i is the conditional mean of the Y_i, κ is the scale parameter of the negative binomial distribution,

and S_i are independent random tree effects with σ^2 equals to their variance. Note the S_i in (40) are taken as independent since on subsequent sampling occasions, the trees are sampled "out of range". Also we note that the quadrant effect in model (39) is absorbed into μ in (40) since only the most infested quadrant is sampled in subsequent samples, and that σ^2 in (40) is $\sigma_\varepsilon^2 + \sigma_S^2$ in (39).

In (40), the conditional means μ_i are random variables. It is easy to show that $\exp(\mu)$ is the median of the μ_i quantities. Hence, a test of $\mathrm{H}_0 : \mu = \mu_0$ versus $\mathrm{H}_1 : \mu = \mu_1$ $(\mu_1 > \mu_0)$ is equivalent to a test about the median of the conditional mean number of pest counts on each leaf. Applying Bartlett's SPRT to the IID Y_i, the sequential test is based on the log likelihood ratio

$$\lambda_n = \log\{ f(\underset{\sim}{y}; \mu_1, \hat{\sigma}_n^2(\mu_1), \hat{\kappa}_n(\mu_1)) / f(\underset{\sim}{y}; \mu_0, \hat{\sigma}_n^2(\mu_0), \hat{\kappa}_n(\mu_0))\} \qquad (41)$$

Where n denotes the current number of observations in the sample, $\underset{\sim}{y} = (y_1, \cdots, y_n)'$ and

$$f(\underset{\sim}{y}; \mu, \sigma^2, \kappa) = \prod_{i=1}^{n} \int_{s_i} \frac{\Gamma(y_i + \kappa)}{\Gamma(y_i + 1)\Gamma(\kappa)} \left(\frac{e^{\mu + s_i}}{e^{\mu + s_i} + \kappa} \right)^{y_i} \left(\frac{\kappa}{e^{\mu + s_i} + \kappa} \right)^{\kappa} \frac{1}{2\pi\sqrt{\sigma^2}} e^{-s_i^2 / 2\sigma^2} ds_i \qquad (42)$$

Define $A = \ln(\beta/(1-\alpha))$ and $B = \ln((1-\beta)/\alpha)$. The stop boundaries of Bartlett's SPRT is to accept H_0 at the first *n* for which $\lambda_n \le A$, to accept H_1 at the first *n* for which $\lambda_n \ge B$, and to continue by sampling another tree if $A < \lambda_n < B$. The asymptotic type-1 and type-2 error rates satisfy $Pr(\text{Reject } \mathrm{H}_0 \mid \mathrm{H}_0) \le \alpha$ and $Pr(\text{Reject } \mathrm{H}_1 \mid \mathrm{H}_1) \le \beta$, respectively.

Equation (42) is the marginal distribution of IID Y_i, integrating out the common random effect S_i. We note that evaluating the integrated likelihood in (42), for use in (41), requires use of a numerical integration method. As is discussed in chapter 2, we use $M = 50$ point Gauss-Hermite quadrature rule:

$$f(y_i ; \mu, \sigma^2, \kappa) \approx \frac{1}{\sqrt{\pi}} \sum_{t=1}^{M} R_{it} w_t$$

where $R_{it} = \frac{\Gamma(y_i + \kappa)}{\Gamma(y_i + 1)\Gamma(\kappa)} \left(\frac{e^{\mu + \sqrt{2}\sigma x_t}}{e^{\mu + \sqrt{2}\sigma x_t} + \kappa} \right)^{y_i} \left(\frac{\kappa}{e^{\mu + \sqrt{2}\sigma x_t} + \kappa} \right)^{\kappa}$, x_t are the roots of Hermite polynomial, w_t are the associated weights.

The values $\hat{\sigma}_n^2(\mu)$ and $\hat{\kappa}_n(\mu)$ denote the conditional MLE of the unknown nuisance parameters σ^2 and κ, given a value of μ. These values are obtained by setting μ to a fixed value in (42) and then maximizing the right hand side with respect to (σ^2, κ). The positive property of both σ and κ requires a maximization with constraints. We reparameterize $\kappa = e^a$ and $\sigma = e^b$ and maximize the log likelihood with respect to a and b without constraint. For use with Newton's method to get the conditional MLEs, the first order derivatives and the hessian matrix with respect to a and b are calculated:

$$\frac{\partial \log f(\underset{\sim}{y} ; \mu, \sigma^2, \kappa)}{\partial a} \approx \frac{\partial \log \prod_{i=1}^{n} \sum_{t=1}^{M} R_{it} w_t}{\partial a} = \sum_{i=1}^{n} \left(\sum_{t=1}^{M} \frac{\partial R_{it}}{\partial a} w_t \Big/ \sum_{t=1}^{M} R_{it} w_t \right)$$

with,

$$\frac{\partial R_{it}}{\partial a} = \frac{e^{(\mu+\sqrt{2}\sigma x_t)y_i}}{\Gamma(y_i+1)} \frac{\partial}{\partial a}\left\{ \frac{\Gamma(y_i+e^a)}{\Gamma(e^a)} \times \frac{e^{ae^a}}{(e^{\mu+\sqrt{2}\sigma x_t}+e^a)^{y_i+e^a}} \right\}$$

$$= \frac{e^{(\mu+\sqrt{2}\sigma x_t)y_i}}{\Gamma(y_i+1)} \left\{ \frac{\partial}{\partial a}\left(\frac{e^{ae^a}}{(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)^{y_i+e^a}} \right) \times \frac{\Gamma(y_i+e^a)}{\Gamma(e^a)} + \frac{\partial}{\partial a}\left(\frac{\Gamma(y_i+e^a)}{\Gamma(e^a)} \right) \times \frac{e^{ae^a}}{(e^{\mu+\sqrt{2}\sigma x_t}+e^a)^{y_i+e^a}} \right\}$$

where

$$\frac{\partial}{\partial a}\left(\frac{e^{ae^a}}{(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)^{y_i+e^a}} \right)$$

$$= \left(\frac{e^a}{e^{\mu+\sqrt{2}\sigma x_t}+e^a} \right)^{e^a} \times \frac{\partial}{\partial a}\left[e^a \log\left(\frac{e^a}{e^{\mu+\sqrt{2}\sigma x_t}+e^a} \right) \right] \times \left(\frac{1}{e^{\mu+\sqrt{2}\sigma x_t}+e^a} \right)^{y_i}$$

$$- \frac{e^a}{(e^{\mu+\sqrt{2}\sigma x_t}+e^a)^{y_i+1}} \times \frac{e^{ae^a} y_i}{(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)^{e^a}}$$

$$= \frac{\kappa^\kappa}{(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)^{\kappa+y_i}} \left(\kappa\log(\kappa) - \kappa\log(e^{\mu+\sqrt{2}\sigma x_t}+\kappa) + \frac{e^{\mu+\sqrt{2}\sigma x_t}\kappa}{e^{\mu+\sqrt{2}\sigma x_t}+\kappa} \right) - \frac{\kappa^{\kappa+1} y_i}{(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)^{y_i+\kappa+1}}$$

$$\frac{\partial}{\partial a}\left(\frac{\Gamma(y_i+e^a)}{\Gamma(e^a)} \right) = \kappa \times \frac{\Gamma'(y_i+\kappa)\Gamma(\kappa) - \Gamma'(\kappa)\Gamma(y_i+\kappa)}{\Gamma^2(\kappa)}$$

Where Γ' denotes the derivative of Gamma function.

Therefore, $\frac{\partial R_{it}}{\partial a}$ can be simplified as:

$$\frac{\partial R_{it}}{\partial a} = \frac{\Gamma'(y_i+\kappa)\Gamma(\kappa) - \Gamma'(\kappa)\Gamma(y_i+\kappa)}{\Gamma^2(\kappa)} \times \frac{\kappa^{\kappa+1} e^{(\mu+\sqrt{2}\sigma x_t)y_i}}{\left(e^{(\mu+\sqrt{2}\sigma x_t)}+\kappa\right)^{y_i+\kappa}\Gamma(y_i+1)}$$

$$+ \frac{e^{(\mu+\sqrt{2}\sigma x_t)y_i}\kappa^{\kappa+1}}{\Gamma(y_i+1)(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)^{y_i+\kappa}} \left(a - \log(e^{\mu+\sqrt{2}\sigma x_t}+\kappa) + \frac{e^{\mu+\sqrt{2}\sigma x_t}-y_i}{e^{\mu+\sqrt{2}\sigma x_t}+\kappa} \right)$$

$$= \frac{\kappa}{\Gamma(y_i+1)}\left(\frac{e^{\mu+\sqrt{2}\sigma x_t}}{e^{\mu+\sqrt{2}\sigma x_t}+\kappa}\right)^{y_i}\left(\frac{\kappa}{e^{\mu+\sqrt{2}\sigma x_t}+\kappa}\right)^{\kappa}$$

$$\times\left(\frac{\Gamma'(y_i+\kappa)\Gamma(\kappa)-\Gamma'(\kappa)\Gamma(y_i+\kappa)}{\Gamma^2(\kappa)}+a-\log(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)+\frac{e^{\mu+\sqrt{2}\sigma x_t}-y_i}{e^{\mu+\sqrt{2}\sigma x_t}+\kappa}\right)$$

$$\frac{\partial \log f(\underset{\sim}{y};\mu,\sigma^2,\kappa)}{\partial^2 a} \approx \sum_{i=1}^{n}\frac{\left(\sum_{t=1}^{M}\frac{\partial R_{it}}{\partial^2 a}w_t\right)\sum_{t=1}^{M}R_{it}w_t-\left(\sum_{t=1}^{M}\frac{\partial R_{it}}{\partial a}w_t\right)^2}{\left(\sum_{t=1}^{M}R_{it}w_t\right)^2}$$

With,

$$\frac{\partial R_{it}}{\partial^2 a}=\left\{\frac{\Gamma''(y_i+\kappa)\Gamma^3(\kappa)-\Gamma''(\kappa)\Gamma(y_i+\kappa)\Gamma^2(\kappa)+2\Gamma'(y_i+\kappa)\Gamma'^2(\kappa)-2\Gamma'(\kappa)\Gamma(\kappa)\Gamma'(y_i+k)}{\Gamma^4(\kappa)}\right.$$

$$\left.+1-\frac{\kappa}{e^{\mu+\sqrt{2}\sigma x_t}+\kappa}-\frac{(e^{\mu+\sqrt{2}\sigma x_t}-y_i)\kappa}{(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)^2}\right\}\times\frac{e^{(\mu+\sqrt{2}\sigma x_t)y_i}\kappa^{\kappa+2}}{(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)^{y_i+\kappa}\Gamma(y_i+1)}$$

$$\frac{\partial \log f(\underset{\sim}{y};\mu,\sigma^2,\kappa)}{\partial b} \approx \frac{\partial \log\prod_{i=1}^{n}\sum_{t=1}^{M}R_{it}w_t}{\partial b}\sum_{i=1}^{n}\left(\sum_{t=1}^{M}\frac{\partial R_{it}}{\partial b}w_t \Big/ \sum_{t=1}^{M}R_{it}w_t\right)$$

with,

$$\frac{\partial R_{it}}{\partial b}=\frac{\Gamma(y_i+\kappa)}{\Gamma(y_i+1)\Gamma(\kappa)}\frac{\partial}{\partial b}\left\{\left(1-\frac{\kappa}{e^{\mu+\sqrt{2}e^b x_t}+\kappa}\right)^{y_i}\left(\frac{\kappa}{e^{\mu+\sqrt{2}e^b x_t}+\kappa}\right)^{\kappa}\right\}$$

$$=\frac{\Gamma(y_i+\kappa)}{\Gamma(y_i+1)\Gamma(\kappa)}\left\{\frac{\sqrt{2}\sigma x_t\kappa e^{\mu+\sqrt{2}\sigma x_t}}{(e^{\mu+\sqrt{2}e^b x_t}+\kappa)^2}\left(1-\frac{\kappa}{e^{\mu+\sqrt{2}e^b x_t}+\kappa}\right)^{y_i-1}\left(\frac{\kappa}{e^{\mu+\sqrt{2}e^b x_t}+\kappa}\right)^{\kappa}\right.$$

$$\left.-\left(1-\frac{\kappa}{e^{\mu+\sqrt{2}e^b x_t}+\kappa}\right)^{y_i}\left(\frac{\kappa}{e^{\mu+\sqrt{2}e^b x_t}+\kappa}\right)^{\kappa-1}\frac{\sqrt{2}\kappa^2 x_t e^{\mu+\sqrt{2}e^b x_t}}{(e^{\mu+\sqrt{2}e^b x_t}+\kappa)^2}\right\}$$

$$=\frac{\Gamma(y_i+\kappa)}{\Gamma(y_i+1)\Gamma(\kappa)}\frac{\sqrt{2}\sigma x_t(\kappa-e^{\mu+\sqrt{2}\sigma x_t})}{(\kappa+e^{\mu+\sqrt{2}\sigma x_t})}\left(\frac{e^{\mu+\sqrt{2}\sigma x_t}}{e^{\mu+\sqrt{2}\sigma x_t}+\kappa}\right)^{y_i}\left(\frac{\kappa}{e^{\mu+\sqrt{2}\sigma x_t}+\kappa}\right)^{\kappa}$$

$$\frac{\partial \log f(\underset{\sim}{y};\mu,\sigma^2,\kappa)}{\partial^2 b} \approx \sum_{i=1}^{n} \frac{\left(\sum_{t=1}^{M} \frac{\partial R_{it}}{\partial^2 b} w_t\right)\sum_{t=1}^{M} R_{it} w_t - \left(\sum_{t=1}^{M} \frac{\partial R_{it}}{\partial b} w_t\right)^2}{\left(\sum_{t=1}^{M} R_{it} w_t\right)^2}$$

with,

$$\frac{\partial R_{it}}{\partial^2 b} = \frac{2\sigma^2 x_t^2 e^{(\mu+\sqrt{2}\sigma x_t)y_i} \kappa^{\kappa+1} \Gamma(y_i+\kappa)}{(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)^{y_i+\kappa+1}\Gamma(y_i+1)\Gamma(\kappa)}$$

$$\times\left(\frac{\sqrt{2}\sigma x_t e^{\mu+\sqrt{2}\sigma x_t}(y_i+\kappa+1)(e^{\mu+\sqrt{2}\sigma x_t}-1)}{e^{\mu+\sqrt{2}\sigma x_t}+\kappa} + \sqrt{2}\sigma x_t(y_i - e^{\mu+\sqrt{2}\sigma x_t} y_i - e^{\mu+\sqrt{2}\sigma x_t}) - e^{\mu+\sqrt{2}\sigma x_t}\right)$$

$$\frac{\partial \log f(\underset{\sim}{y};\mu,\sigma^2,\kappa)}{\partial a \partial b} \approx \sum_{i=1}^{n}\left\{\sum_{t=1}^{M}\frac{\partial R_{it}}{\partial a \partial b} w_t \Big/ \sum_{t=1}^{M} R_{it} w_t - \left(\sum_{t=1}^{M}\frac{\partial R_{it}}{\partial b}\right)\times w_t \left(\sum_{t=1}^{M}\frac{\partial R_{it}}{\partial a \partial b} w_t\right) \Big/ \left(\sum_{t=1}^{M} R_{it} w_t\right)^2\right\}$$

With,

$$\frac{\partial R_{it}}{\partial a \partial b} = \frac{\sqrt{2}\sigma x_t e^{(\mu+\sqrt{2}\sigma x_t)y_i}}{\Gamma(y_i+1)(e^{(\mu+\sqrt{2}\sigma x_t)y_i}+\kappa)^{y_i+1}}\left\{\frac{\Gamma'(y_i+\kappa)\Gamma(\kappa)-\Gamma'(\kappa)\Gamma(y_i+\kappa)}{\Gamma^2(\kappa)(\kappa+e^{(\mu+\sqrt{2}\sigma x_t)})^{\kappa}}\kappa^{\kappa+1}\left(\kappa - e^{(\mu+\sqrt{2}\sigma x_t)}\right)\right.$$

$$+\frac{\Gamma(y_i+\kappa)}{\Gamma(\kappa)(e^{(\mu+\sqrt{2}\sigma x_t)}+\kappa)^{\kappa}}\times\left(\frac{\kappa}{e^{\mu+\sqrt{2}\sigma x_t}+\kappa}\right)^{\kappa}\kappa^{\kappa+1}(a\kappa+\kappa+1)$$

$$\left.-(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)\log(e^{\mu+\sqrt{2}\sigma x_t}+\kappa)+\frac{e^{2\mu+2\sqrt{2}\sigma x_t}-\kappa^2}{e^{\mu+\sqrt{2}\sigma x_t}+\kappa}+ae^{\mu+\sqrt{2}\sigma x_t}\right\}$$

Due to the fact that Bartlett's SPRT utilizes conditional estimates of the nuisance parameters, collecting an adequate number of initial samples before implementing Bartlett's SPRT is necessary. We suggest using the first occasion sample data to decide the starting sample size required by Bartlett's SPRT as follows:

i. Simulate pest count data with respect to (40) using the first occasion sample estimates $\hat{\kappa}$, $\hat{\mu}$ ($\hat{\mu}$ is equal to the $\hat{\gamma}_j$ corresponding to the

quadrant determined to be sampled based on the initial first occasion sample analysis) and $\hat{\sigma}^2$ ($\hat{\sigma}^2 = \hat{\sigma}_\varepsilon^2 + \hat{\sigma}_S^2$).

ii. Perform the "out of range" sequential hypothesis test on the simulated data beginning with an initial starting sample size of 3 trees, which is equal the number of unknown parameters in the model (40).

iii. Repeat this procedure a large number of times (for example, 3000 times) and estimate OC(μ_0) and OC(μ_1) by the fraction of cases where H_0 and H_1 are accepted.

iv. If the estimated OC(μ_0) and OC(μ_1) values are not satisfactorily close to their nominal values $1-\alpha$ and β, respectively, increase the starting sample size for trees by 1 and repeat (i) – (iv) with the larger starting sample size.

Eventually the estimated OC(μ_0) and OC(μ_1) values will be satisfactorily close to their nominal values and the iterative procedure will stop. The starting sample size at the end of the iterations can be used when implementing Bartlett's SPRT for the block under study for the sequential samples at future occasions.

3.3. *Illustration Example on Leafhopper Count Data*

3.3.1. *First Occasion Sampling*

We have illustrated the existence of spatial correlations in pest counts by the analysis of *O. peasea* data. Now we want to illustrate how our proposed sequential hypothesis test plan could be applied to spatially correlated pest count. We use the leafhopper pest to illustrate our proposed sampling methodology because it is more practical to obtain sequential pest counts with this pest than it is for the *O. perseae* whose higher densities and microscopic size make it difficult to count.

Wine grape leafhoppers are plant feeders, and entomologists suggest a treatment of the vineyard when 1st generation leafhopper nymphs (i.e., the immature form of the insect) number 20 or more per leaf. However, when there are 10 or less per leaf, no significant damage will be caused and there is no need for such treatment (UC IPM 1992). Based on (40), and recalling the median of the conditional means is $\exp(\mu)$, we wish to test the hypothesis $H_0 : \mu = 1.9$ versus $H_1 : \mu = 2.3$ on leafhopper pest count.

Table 7 shows a hypothetical first occasion sample data similar to what a practitioner would begin with. The data is comprised of 288 leafhopper counts, collected from 36 trees on a six by six grid from the middle of the block, where eight leaves are collected from sampling 2 leaves sampled from each cardinal direction. The pest counts range from 0 to 45 with a sample mean of 8.48 and a sample variance of 41.94. It is worth of mentioning again that a practitioner might also choose to sample every other tree on a twelve by twelve block to expand out the geographic size of this first occasion sample.

As discussed in the previous section, the first occasion sample is used to make the fist pest assessment decision, as well as provide the necessary information to set up the subsequent sequential samples. **Table 8** lists the estimates of the fixed quadrant effects from fitting model (39) to the first occasion sample data. The P-value of the F test for equal quadrant effects is less than 0.0001, and pairwise contrast tests show that East quadrant is significantly higher than all the other three quadrants. A 90% confidence interval for the most infested East quadrant fixed effect is (2.20, 2.52). Since $H_1 : \mu = 2.3$ is in the 90% confidence interval range, we would therefore conclude that the leafhopper per leaf are more than 20 on the sampled block and there is a need for first occasion treatment of the block.

Quadrant Effect	Estimate (log-scale)	Standard Error
East	2.36	0.10
North	2.16	0.11
South	1.91	0.11
West	1.71	0.11

Table 8. Analysis of Fixed Effects for Hypothetical Leafhopper Count Data

Tree	Coordinate		East Quadrant		North Quadrant		South Quadrant		West Quadrant	
	x	y	Leaf #1 Count	Leaf #2 Count	Leaf #1 Count	Leaf #2 Count	Leaf #1 Count	Leaf #2 Count	Leaf #1 Count	Leaf #2 Count
1	1	1	5	5	7	4	3	15	20	17
2	2	1	5	7	10	10	10	12	19	33
3	3	1	6	4	5	2	11	11	18	9
4	4	1	3	1	14	14	9	7	7	27
5	5	1	7	1	12	1	5	9	9	11
6	6	1	4	6	5	2	9	4	9	5
7	1	2	8	5	3	3	0	9	2	9
8	2	2	6	2	7	5	3	13	4	9
9	3	2	0	11	2	5	17	9	8	4
10	4	2	11	10	6	8	5	11	11	14
11	5	2	0	2	7	9	4	2	10	16
12	6	2	2	3	1	1	6	1	7	1
13	1	3	6	5	25	7	22	8	23	10
14	2	3	1	6	8	9	11	10	6	4
15	3	3	7	5	4	7	7	9	5	5
16	4	3	4	6	5	11	8	20	9	7
17	5	3	4	3	10	3	10	7	8	15
18	6	3	8	9	3	16	38	7	12	27
19	1	4	5	3	3	6	10	6	9	11
20	2	4	9	6	9	4	12	15	9	19
21	3	4	7	7	12	11	24	6	5	14
22	4	4	3	10	1	0	6	5	3	3
23	5	4	3	5	11	11	12	0	13	13
24	6	4	7	24	16	11	15	9	45	37
25	1	5	3	5	3	3	7	3	1	8
26	2	5	8	1	6	4	6	14	10	8
27	3	5	4	3	13	1	5	3	9	10
28	4	5	7	16	4	8	7	10	5	6
29	5	5	11	7	13	8	2	2	7	6
30	6	5	21	9	12	18	21	12	8	12
31	1	6	6	5	8	7	13	9	10	10
32	2	6	6	6	2	4	4	5	15	12
33	3	6	3	5	1	3	6	9	7	7
34	4	6	1	4	5	7	17	20	26	18
35	5	6	4	12	8	9	4	11	12	27
36	6	6	2	5	22	6	4	8	5	6

Table 7. Hypothetical First Occasion Initial Sample Leafhopper Counts Data

Table 9 lists the estimated variance covariance parameters from fitting model (39) to the first occasion sample data, based on which we can define the practical range for the subsequent occasion sequential samples. As is shown in **Table 4**, the scale parameter in spatial covariance structure is estimated to be $\hat{\theta} = 0.58$. The practical range for uncorrelated observations is calculated by $-\hat{\theta}\log(0.05) \approx 2$, implying sampling every 3rd tree will approximately yield uncorrelated data. Together with information on the fixed quadrant effects, we conclude from the first occasion sample that future subsequent sequential procedures in this block should sample one leaf per tree on the East quadrant from every third tree.

Covariance Parameter	Estimated Values	Standard Error of Estimation
Dispersion Parameter of Negative Binomial Distribution (κ)	6.04	0.03
Variance of Spatial Covariance Structure (σ_S^2)	0.11	0.04
Scale Parameter in Spatial Covariance Structure (θ)	0.58	0.43
Variance of Tree by Quadrant Interaction (σ_ε^2)	0.04	0.03

Table 9. Variance-Covariance Parameter Values for Leafhopper Initial Sample Data

Model (40), using $\kappa = 6.04$ and $\sigma^2 = 0.15$ from **Table 9** (the sum of 0.11 and 0.04), can be used with the iterative procedure described in the previous section to find a suitable starting sample size for the proposed subsequent occasion "out of range"

sequential sampling procedure. Taking the nominal type-1 and type-2 errors to both equal 0.1, we therefore desire OC(μ_0) to be close to 0.9 and OC(μ_1) to be close to 0.1. **Table 10** lists the OC(μ_0) and OC(μ_1) based on 3,000 simulated sample paths for alternative starting sample sizes. It is noted that as the starting sample size increases from 3 trees to 9 trees, the corresponding OC(μ_0) values get closer to the nominal value while OC(μ_1) remains in the desirable value range. When the starting sample size is 9 trees, the OC value is 0.90 at H_0, and 0.05 at H_1. We would therefore select 9 trees as the starting sample size for the proposed subsequent occasion sequential sampling procedure.

Starting Sample Size	3	4	5	6	7	8	9
OC (μ_0)	0.87	0.88	0.88	0.89	0.89	0.89	0.90
OC (μ_1)	0.07	0.08	0.07	0.06	0.06	0.05	0.05

Table 10. Calculation of Starting Sample Size for Subsequent Occasion Sequential Hypothesis Test for Leafhopper Pest Counts

3.3.2. *Subsequent Occasion Sequential Sampling*

For the subsequent occasion sequential samples, it was determined above that a practitioner will sample from the East quadrant of every 3rd tree. After collecting the first 9 trees, Bartlett's SPRT can be initiated as defined via (41) and (42). Based on nominal type-1 and type-2 errors equal to 0.1, λ_n is compared to the two stop boundaries

$A = -2.2$ and $B = 2.2$. The sequential sample stops at the first n for which either $\lambda_n \le A$ or $\lambda_n \ge B$, and continues by sampling another tree if $A < \lambda_n < B$.

Figure 8 depicts a hypothetical sample path for a subsequent sequential sample that a practitioner might experience. The observed leafhopper pest counts along the sample path are: 6, 12, 11, 12, 5, 13, 2, 17, 6, 0, 8, 4, 9, 6, 6, 1, 1, 5, 5. For the example shown in **Figure 8**, the sequential testing procedure starts with 9 trees and stops at 19 trees. Since $\lambda_{19} = -2.345$ is less than the lower stopping boundary -2.2, the sequential procedure is terminated and the null hypothesis is accepted. It would be concluded that on average, the leafhoppers per leaf are less than 10 on the sampled block, and there is no need to treat the block.

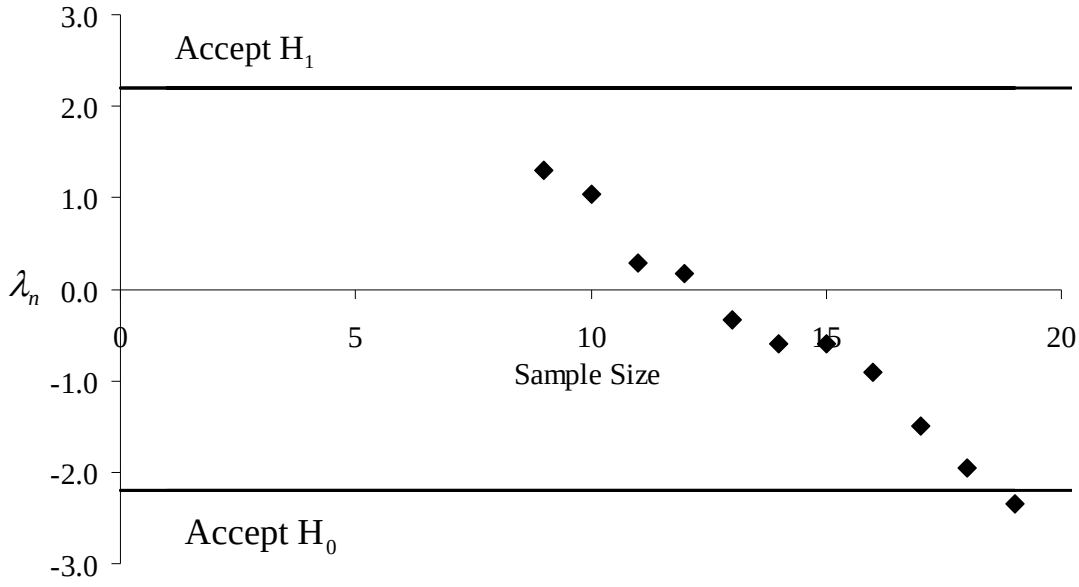

Figure 8. Sample Boundaries and Hypothetical Sequential Sample Path for Leafhopper Pest Counts

Table 11 shows simulated OC and ASN curves as a function of the true μ for our proposed subsequent occasion "out of range" sequential hypothesis test, using 9 trees as the starting sample size. The simulation results are based on 3,000 sample paths where the nominal type-1 and type-2 errors are set to 0.1. We can see that the realized type-1 and type-2 error rates are (0.10, 0.05) respectively, which are desirable values.

μ	OC	Sample Size (# trees) Distribution					
		ASN	10th	25th	50th	75th	90th
1.7	0.99	12.23	9	9	11	14	18
1.8	0.98	14.31	9	9	12	17	23
1.9	0.90	17.60	9	9	13	21	33
2	0.70	20.90	9	10	15	26	42
2.1	0.40	20.99	9	10	15	25	42
2.2	0.14	16.88	9	9	12	20	30
2.3	0.05	13.14	9	9	10	15	21
2.4	0.01	10.89	9	9	9	11	16
2.5	0.01	9.87	9	9	9	9	12

Table 11. OC and ASN properties for $H_0 : \mu_0 = 1.9$ vs $H_1 : \mu_1 = 2.3$ with $\kappa = 6.04$ and nominal type-1 and type-2 error rates are 0.1.

3.3.3. *Fixed Sample Size Test*

For the purpose of an illustrative comparison of the savings in sampling using sequential procedure as compared to the fixed sample size test, we suppose κ and σ^2 are known. This can be a reasonable assumption, since we can use the estimated parameter values obtained from the first occasion initial sample data.

Similar as chapter 2, NP test of $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1$ is used to calculate the fixed sample size required. A Size α NP test rejects H_0 if $\varphi_n = f_n(\mu_1) / f_n(\mu_0) > k_\alpha$ and accepts H_0 if $\varphi_n < k_\alpha$, where k_α is chosen to satisfy $P(\varphi_n > k_\alpha \mid H_0) = \alpha$. In order to determine k_α, the null distribution of φ_n is needed. We use Monte Carlo methods to simulate the null distribution. We simulate 100,000 vectors of data, $\underset{\sim}{Y}$ with respect to (40) under H_0, and then extract the 99th percentile as the estimate of k_α. For every *n*, the value of k_α guarantees the type-1 error of the NP test will be α. Try all values of $n \geq 2$ and use the Monte Carlo method above to evaluate the power of the corresponding size α NP test sequentially until at the first value of *n* for which the power is $1-\beta$ or larger. That is, we simulate 100,000 data vectors $\underset{\sim}{Y}$ with respect to (40) under H_1 and estimate the power by the fraction for which $\varphi_n = L_n(\mu_1) / L_n(\mu_0) > k_\alpha$.

Based on the estimates from the hypothetical first occasion leafhopper data introduced in section 3.3.1, using $\kappa = 6$ and $\sigma^2 = 0.15$, our simulation of a 0.1 size NP test of testing $H_0 : \mu = 1.9$ vs $H_1 : \mu = 2.3$ yields a fixed sample size of 19. **Figure 9** depicts the comparison of the savings in sampling using our proposed subsequent occasion sequential procedure and the NP test.

In this chapter, we have discussed the importance of incorporating spatial properties into a model when setting up a sequential hypothesis test procedure. In addition, we proposed a sampling methodology that is comprised of first occasion fixed size sample and subsequent occasion "out of range" sequential samples that utilize a

negative binomial spatial GLMM setting. It is worth mentioning that the proposed sequential procedure can also be extended to Poisson spatial GLMM setting.

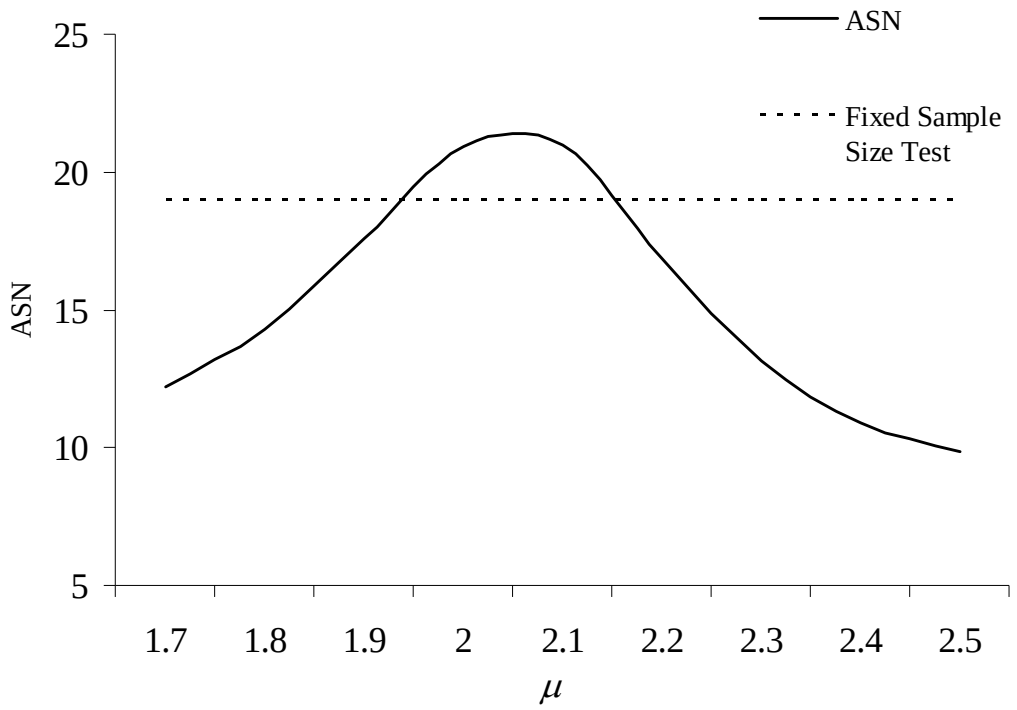

Figure 9. ASN Curve for Proposed Subsequent Occasion Sequential Procedure and Fixed Sample Size Test of $H_0 : \mu = 1.9$ vs $H_1 : \mu = 2.3$ with nominal type-1 and type-2 errors rates of 0.1 on Illustrated Leafhopper Count Data

CHAPTER 4: Design of Sequential Hypothesis Testing Methodology for Generative Spatial Model

4.1. *Poisson Generative Spatial Model*

In chapter 3, we extended Bartlett's method to contexts where spatial GLMMs are used to take into account the spatial distribution of pest counts. As an alternative to spatial GLMMs, we propose a generative spatial model in this chapter. We then apply and extend sequential probability ratio test to our generative spatial model.

4.1.1. *Model Formulation*

In order to characterize the spatial distribution of the pest for the purpose of determining whether intervention via pesticides is needed or not during the sequential procedures, we propose a generative spatial model. We explain our proposed model based on a hypothetical 8 by 8 grid shown in **Figure 10**. Let Y_i be the pest count observed on sample leaf collected on tree i. Our proposed generative spatial model describes Y_i as the sum of its adjacent latent random variables, which can be formulated as bellow:

$$Y_i = X_{i-1} + X_i + X_{i+1} + X_{i-9} + X_{i-8} + X_{i-7} + X_{i+7} + X_{i+8} + X_{i+9} \tag{43}$$

where $X_i \overset{iid}{\sim} \text{Poisson}(\mu/9)$ represent latent random variables corresponding to the trees in the grid. There are 36 additional latent variables outside the block along the edges in order to be able to define Y_i for all 64 trees. As an example, Y_{36} will be the sum of its

adjacent latent X_i random variables corresponding to tree 27, 28, 29, 35, 36, 37, 43, 44, and 45 which are circled by the dashed line in **Figure 10**.

Based on the convolution property of Poisson random variables, $\underset{\sim}{Y}$ follows Poisson distribution with mean $\mu\underset{\sim}{1}$ and variance covariance structure $\Sigma(\mu)$. Let $d_{i,i'}$ be the Euclidean distance between the i-th and i'-th trees, $\Sigma(\mu)$ is a spatial covariance structure modeled as a function of $d_{i,i'}$ as follows:

$$\Sigma(\mu) = \begin{cases} 2/3\mu, & \text{if } d_{i,i'} = 1 \\ 4/9\mu, & \text{if } d_{i,i'} = \sqrt{2} \\ 1/3\mu, & \text{if } d_{i,i'} = 2 \\ 2/9\mu, & \text{if } d_{i,i'} = \sqrt{5} \\ 1/9\mu, & \text{if } d_{i,i'} = \sqrt{8} \\ 0, & \text{if } d_{i,i'} \geq 3 \end{cases} \tag{44}$$

From (44), we see that observations that are 3 or more unit distances apart are uncorrelated with each other. Comparing to the exponential spatial covariance structure we have discussed in chapter 3, the computational advantage of (44) is that we only have one unknown parameter which is the parameter of interest.

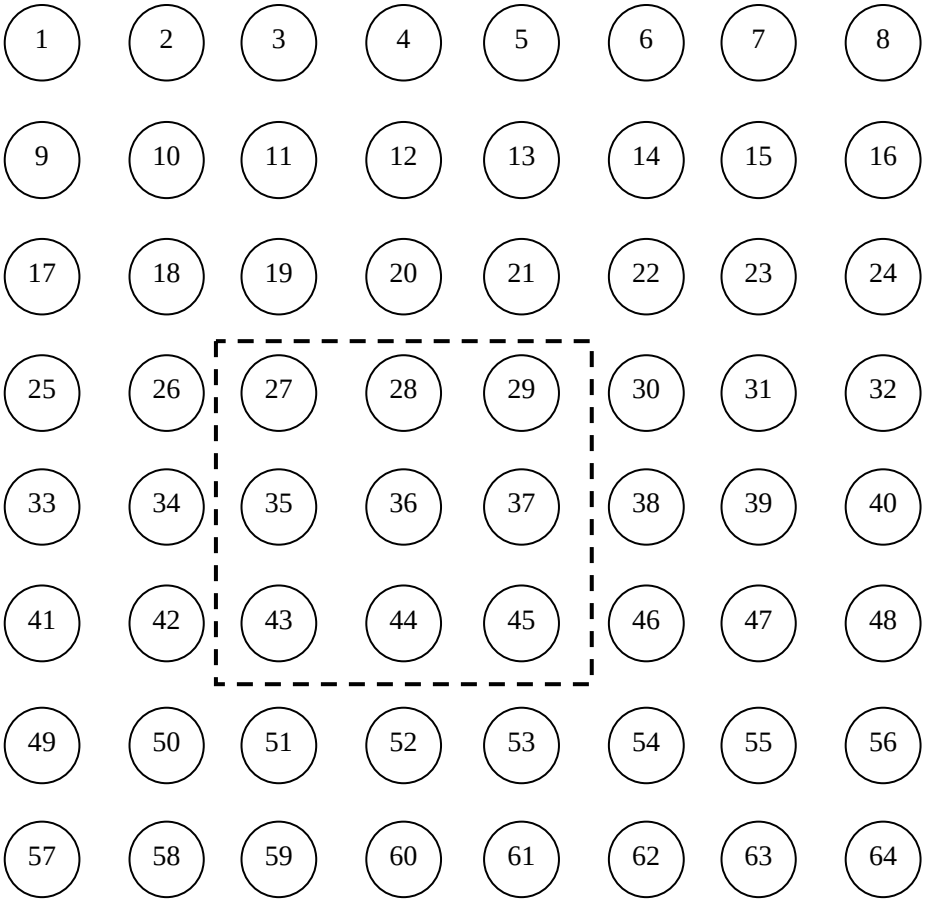

Figure 10. Hypothetical Block of 64 Trees

Figure 11 is a heat map generated based on hypothetical pest count data with respect to (43) on a 8 by 8 grid with μ=2. Larger values are represented by darker color squares and smaller values by lighter squares. It can be seen that the model is able to generate "hot spots" that are reflective of spatial correlation.

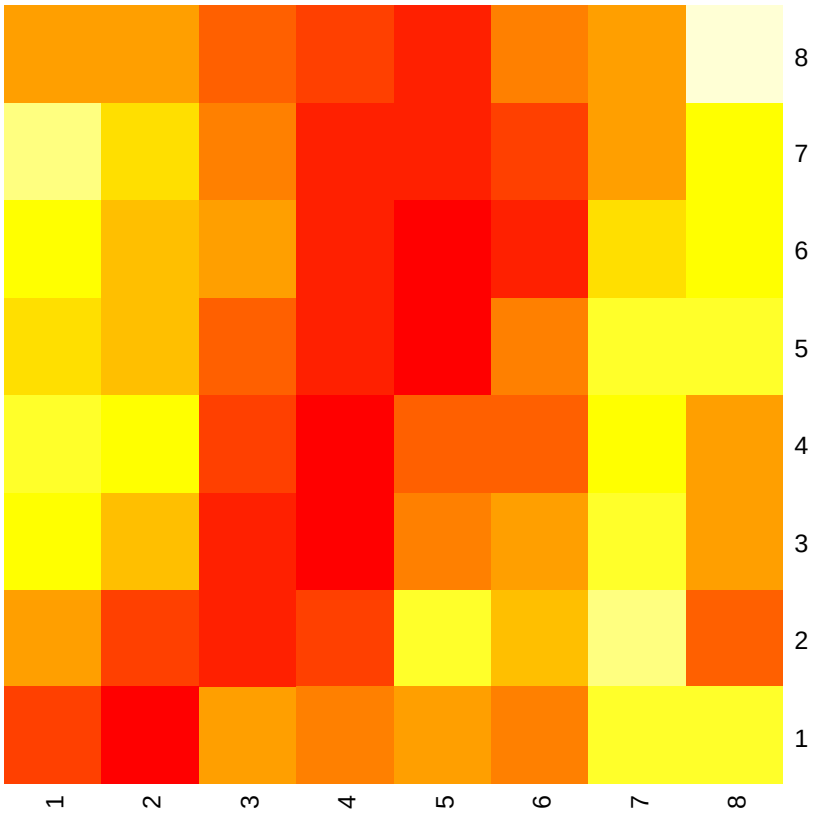

Figure 11. Heat Map of Hypothetical Pest Count Data based on Poisson Generative Spatial Model

4.1.2. *Wald's SPRT for Poisson Generative Spatial Model*

Our goal of pest management is to sequentially discriminate between two simple hypotheses about the mean number of pest on each leaf:

$$\mathrm{H}_0 : \mu = \mu_0 \qquad \mathrm{H}_1 : \mu = \mu_1 \quad (\mu_1 > \mu_0). \tag{45}$$

Since there are no unknown nuisance parameters, we can simply apply Wald's SPRT procedure to our proposed Poisson generative spatial model. However, it is not easy to write out the joint likelihood function of $\underset{\sim}{Y}$. Based on the practical consequence of the

central limit theorem, we can alternatively approximate the joint likelihood function of $\underset{\sim}{Y}$ by a multivariate normal likelihood function:

$$\frac{1}{(2\pi)^{n/2}\left|\Sigma(\mu)\right|^{1/2}} e^{-\frac{1}{2}(\underset{\sim}{y}-\mu\underset{\sim}{1})'\Sigma(\mu)^{-1}(\underset{\sim}{y}-\mu\underset{\sim}{1})} \tag{46}$$

with $\mathrm{E}(Y_i)=\mu$ and $\Sigma(\mu)$ defined by (44).

Now we comparison the exact likelihood and the approximated likelihood based on the example shown in figure 10. Let us rewrite Y_{27} and Y_{28} into:

$$Y_{27} = X_{26} + X_{27} + X_{28} + X_{17} + X_{18} + X_{19} + X_{34} + X_{35} + X_{36} = Z_1 + Z_3$$

$$Y_{28} = X_{27} + X_{28} + X_{29} + X_{18} + X_{19} + X_{20} + X_{35} + X_{36} + X_{37} = Z_2 + Z_3$$

where

$Z_1, Z_2 \sim Poisson(\mu/3)$, and $Z_3 \sim Poisson(2\mu/3)$. Z_1, Z_2, Z_3 are independent of each other. The joint likelihood function of Y_{27} and Y_{28} can be written as:

$$f_{Y_{27},Y_{28}}(y_{27}, y_{28}) = \sum_{z_3=0}^{\min(y_{27},y_{28})} \frac{e^{-\mu/3}(\mu/3)^{y_{27}-z_3}}{(y_{27}-z_3)!} \times \frac{e^{-\mu/3}(\mu/3)^{y_{28}-z_3}}{(y_{28}-z_3)!} \times \frac{e^{-2\mu/3}(2\mu/3)^{z_3}}{(z_3)!}$$

$$= \sum_{z_3=0}^{\min(y_{27},y_{28})} \frac{e^{-4\mu/3}(\mu/3)^{y_{27}+y_{28}-2z_3}(2\mu/3)^{z_3}}{(y_{27}-z_3)!(y_{28}-z_3)!(z_3)!}$$

The approximated likelihood function based on (46) will be:

$$f_{Y_{27},Y_{28}}(y_{27}, y_{28}) \approx \frac{1}{(2\pi)^{2/2}\left|\Sigma(\mu)\right|^{1/2}} e^{-\frac{1}{2}(\underset{\sim}{y}-\mu\underset{\sim}{1})'\Sigma(\mu)^{-1}(\underset{\sim}{y}-\mu\underset{\sim}{1})}$$

with $\Sigma(\mu) = \begin{pmatrix} \mu & 2\mu/3 \\ 2\mu/3 & \mu \end{pmatrix}$, and $\underset{\sim}{y}' = (y_{27}, y_{28})$.

Figure 12 characterizes the exact $f_{Y_{27},Y_{28}}(y_{27}, y_{28})$ and the approximated $f_{Y_{27},Y_{28}}(y_{27}, y_{28})$ as a function of μ when $y_{27} = 9$ and $y_{28} = 5$. It can be seen that the approximated likelihood function is very close to the exact likelihood. We have also tried different values of (y_{27}, y_{28}) and similar plots are obtained. From this comparison, we see the complexity of deriving the exact joint likelihood function and the appealing advantage of the simplicity of multivariate normal approximation.

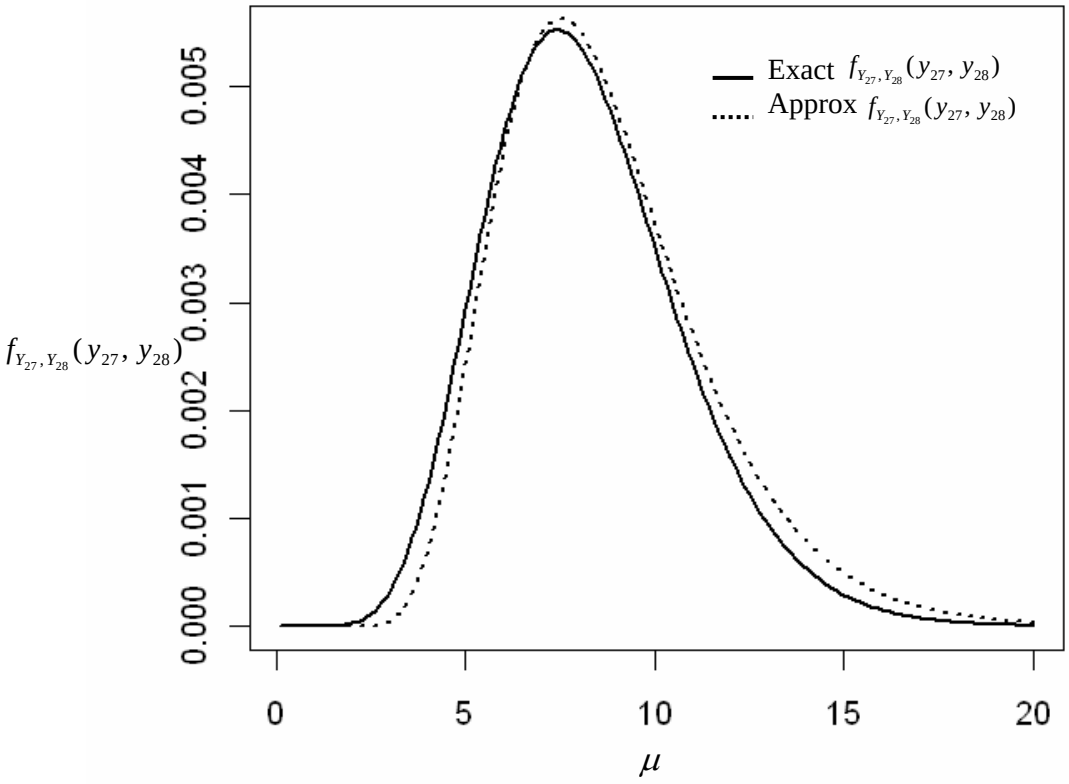

Figure 12. Comparison of Exact and Approximated Likelihood for Poisson Generative Spatial Model

Wald's SPRT procedure for testing hypotheses (45) is then based on the log likelihood ratio

$$\Lambda_n = \frac{1}{2}\log\left(\frac{|\Sigma(\mu_0)|}{|\Sigma(\mu_1)|}\right) + \frac{1}{2}(\underset{\sim}{y} - \mu_0\underset{\sim}{1})'\Sigma(\mu_0)^{-1}(\underset{\sim}{y} - \mu_0\underset{\sim}{1}) - \frac{1}{2}(\underset{\sim}{y} - \mu_1\underset{\sim}{1})'\Sigma(\mu_1)^{-1}(\underset{\sim}{y} - \mu_1\underset{\sim}{1})$$

with acceptance of H_0 at the first n for which $\Lambda_n \le \log\left(\frac{\beta}{1-\alpha}\right)$, and acceptance of H_1 at the first n for which $\Lambda_n \ge \log\left(\frac{1-\beta}{\alpha}\right)$. It is important to point out that we only consider the situation of sampling one leaf per tree in this chapter.

4.1.3. *Spatial Sampling Plan*

A sampling plan is an important issue to be considered when utilizing a sequential procedure for spatially correlated data. In this section, we propose two different sampling plans: Simple Random Sampling (SRS) and Cluster Sampling (CS).

With SRS plan the tree to be sampled is chosen randomly and entirely by chance without replacement. It is debatable as to whether SRS plan is a practical sampling plan. We include SRS plan in our discussion as a bench mark comparing to CS plan which is more practical.

CS plan is carried out by sampling trees in clusters. It can be described as follows. Firstly, appoint a randomly chosen tree as the center of a cluster, which includes all the adjacent trees that are 1 unit distance apart from the center tree. A cluster usually includes 5 trees if the center tree is not on the edge of the block under study. In case the

center tree locates on the edge of the block, the number of trees in the cluster varies from 3 to 4 depending on the location of the center tree. Meanwhile, although the appointment of the center tree is random, adjustments might be needed in order to avoid the situation of sampling the same tree twice.

CS plan allows the sequential procedure to be implemented on each cluster of trees sampled, instead of on each tree sampled. Intuitively, the observations collected based on the SRS plan will be less correlated than the observations sampled based on the CS plan. As a result, the SRS plan will lead to less sampling than the CS plan. On the other hand, the CS plan is arguably more practical. We compare the two sampling plans with simulation study in our next section.

4.1.4. *OC and ASN Curves for Poisson Generative Spatial Model*

We simulate a large number of sequential samples and record the fraction of cases where the null hypothesis is accepted and correspondingly the average number of samples required to make a decision. We pick 2 as the initial starting sample size for Wald SPRT with SRS plan, and 5 as the initial sample size for CS plan. It is worth mentioning that we have incorporated the proper correlation between the data points that were near enough to each other to be correlated during the SRS sampling. **Table 12** shows simulated OC and ASN curves (as a function of the true μ) for Wald's SPRT for the two sampling plan. The simulation results are based on 3,000 sample paths of SPRT of $H_0 : \mu = 7$ versus $H_1 : \mu = 10$. The nominal type-1 and type-2 errors are set to 0.1.

We can see that the realized type-1 and type-2 error rates for the two sampling plans CS and SRS in **Table 12** are (0.07, 0.09) and (0.07, 0.07) respectively, and are comfortably close to the nominal 0.1 values. Apart from the ASN values, **Table 12** also shows percentiles of the sample size required with the Wald's SPRT procedure. The difference in the OC characteristics for the two sampling plans is not large. **Figure 12** includes an illustrative comparison of the savings in sampling using CS plan as compared to SRS plan. The starting sample size for CS plan is larger than SRS plan by 3 trees, however we see more difference in the sample sizes required between the two sampling plans. This might due to the fact discussed early that CS plan samples more correlated data than the SRS plan. The ASN curves in **Figure 13** vividly portray the penalty of larger sample size required for CS plan than the SRS plan.

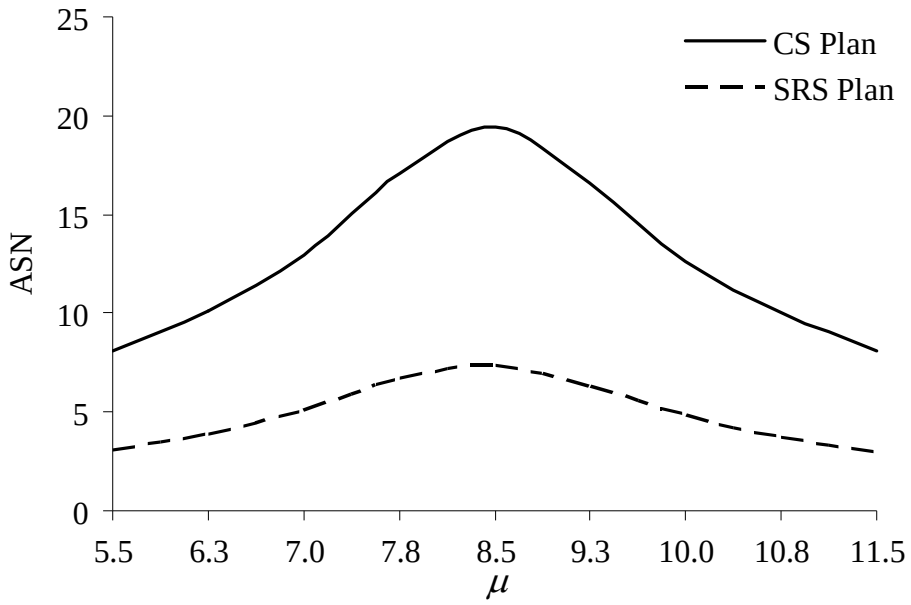

Figure 13. ASN Curves for Wald's SPRT for Poisson Generative Spatial Model ($H_0 : \mu = 7$ versus $H_1 : \mu = 10$, $\alpha = \beta = 0.1$)

μ	CS Plan							SRS Plan						
	OC	Sample Size Distribution						OC	Sample Size Distribution					
		ASN	10th	25th	50th	75th	90th		ASN	10th	25th	50th	75th	90th
6.25	0.99	10.09	5	5	10	10	15	0.99	3.86	2	2	3	5	6
7.00	0.93	12.97	5	9	10	15	24	0.93	5.10	2	3	4	6	9
7.75	0.79	17.10	5	10	14	22	34	0.78	6.69	2	3	5	8	13
8.50	0.49	19.42	5	10	15	25	40	0.48	7.38	2	3	6	10	15
9.25	0.21	16.56	5	9	14	20	34	0.19	6.27	2	3	5	8	12
10.00	0.09	12.59	5	5	10	15	25	0.07	4.83	2	2	4	6	9
10.75	0.03	10.06	5	5	9	13	19	0.02	3.69	2	2	3	4	7
11.50	0.01	8.13	5	5	5	10	15	0.01	2.99	2	2	2	3	5

Table 12. OC and ASN properties for Poisson Generative Spatial Model Testing $H_0 : \mu = 7$ versus $H_1 : \mu = 10$ when $\alpha = \beta = 0.1$.

4.2. *Negative Binomial Generative Spatial Model*

Negative binomial distribution is very flexible in handling over dispersed data, and it also includes Poisson random variables as a limiting case. In this section, we study the properties of our proposed negative binomial generative spatial model.

4.2.1. *Model Formulation*

Negative binomial generative spatial model is set up in a similar way as Poisson generative spatial model. We explain our proposed negative binomial generative spatial model based on the same hypothetical 8 by 8 grid in **Figure 8**. Let X_i denotes the latent random variable corresponds to tree i, where $X_i \overset{iid}{\sim} \text{NegativeBinomial}(\mu/9, \kappa/9)$, with mean equals $\mu/9$ and dispersion parameter equals $\kappa/9$. Let Y_i be the observed pest count from tree i. Our proposed generative spatial model describes Y_i as the sum of its adjacent latent random variables, which is formulated the same way as (43).

Based on the convolution property of negative binomial random variable, $\underset{\sim}{Y}$ follows Negative Binomial distribution with mean $\mu\underset{\sim}{1}$ and variance covariance structure $\Sigma(\mu, \kappa)$. Let $d_{i,i'}$ be the Euclidean distance between the i-th and i'-th trees, $\Sigma(\mu, \kappa)$ is a spatial variance covariance structure modeled as a function of $d_{i,i'}$ as follows:

$$\Sigma(\mu, \kappa) = \begin{cases} 2/3\mu(1+\mu/\kappa), & \text{if } d_{i,i'} = 1 \\ 4/9\mu(1+\mu/\kappa), & \text{if } d_{i,i'} = \sqrt{2} \\ 1/3\mu(1+\mu/\kappa), & \text{if } d_{i,i'} = 2 \\ 2/9\mu(1+\mu/\kappa), & \text{if } d_{i,i'} = \sqrt{5} \\ 1/9\mu(1+\mu/\kappa), & \text{if } d_{i,i'} = \sqrt{8} \\ 0, & \text{if } d_{i,i'} \geq 3 \end{cases} \tag{47}$$

Based on (47), we see that observations further apart than 3 unit distance are independent of each other and closer than 3 unit distance are correlated. Comparing to the exponential spatial covariance structure we have discussed in chapter 3, (47) is computationally simpler. However, comparing to (44), we have one unknown nuisance parameter κ.

Figure 14 is a heat map generated from a hypothetical pest count data on a 8 by 8 grid with μ=2 and κ=1. Same as **Figure 11**, larger values were represented by darker color squares and smaller values by lighter squares in **Figure 14**. Again it can be seen that the model can generate "hot spots" reflective of spatial correlation.

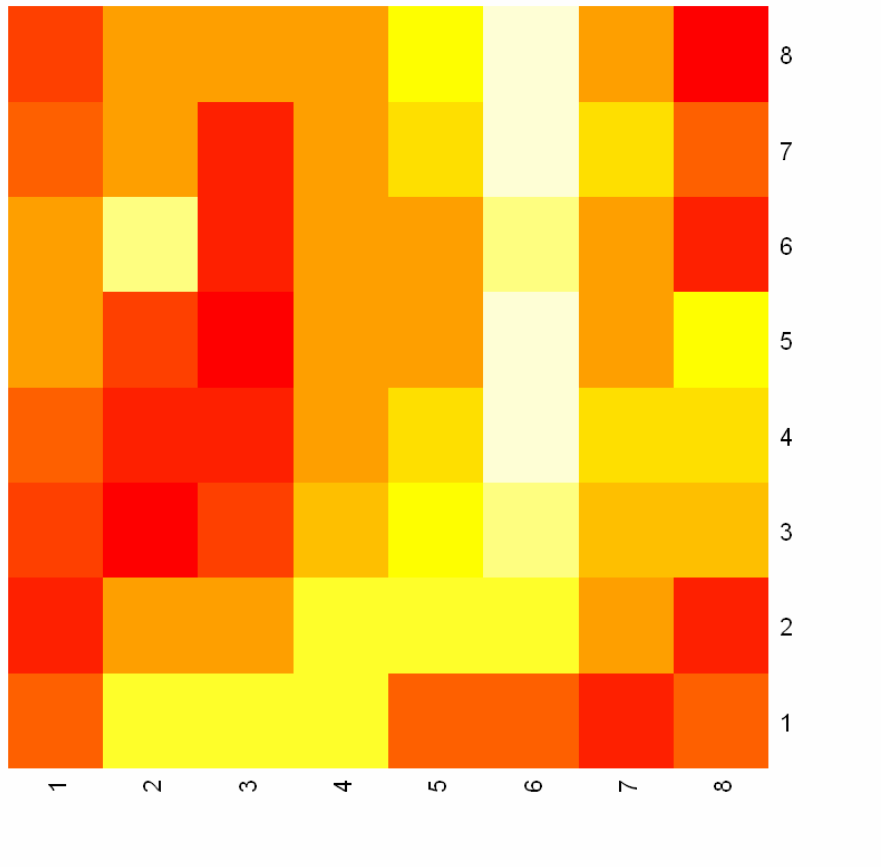

Figure 14. Heat Map of Hypothetical Pest Count Data based on Negative Binomial Generative Spatial Model

4.2.2. *Transformed Bartlett's SPRT for Negative Binomial Generative Spatial Model*

As is introduced in (45) the goal of pest management is to sequentially discriminate between two simple hypotheses about the mean number of pest on each leaf. Since there is an unknown nuisance parameter, we can not simply apply Wald's SPRT procedure to our proposed negative binomial generative spatial model. Based on the model formulation of our negative binomial generative spatial model, we proposed a transformed Bartlett's SPRT based on similar multivariate normal approximation that has been used in section 4.1.2.

Due to the complicity of the joint likelihood function of $\underset{\sim}{Y}$, we approximate it by a multivariate normal likelihood function given by (46) where $\mathrm{E}(Y_i) = \mu$ and $\Sigma(\mu)$ defined by (47). Denote $\Sigma(\mu)$ by $\mu(1+\mu/\kappa)\Omega$ where Ω is the known constant correlation matrix of $\underset{\sim}{Y}$. Let Γ be the Cholesky factorization of Ω which is $\Gamma'\Gamma = \Omega$. We then transform $\underset{\sim}{Y}$ into $\underset{\sim}{Z}(\mu)$ by $\underset{\sim}{Z}(\mu) = \Gamma'^{-1}(\underset{\sim}{Y} - \mu\underset{\sim}{1})$. It follows that $Z_i(\mu)$ will be approximately IID normal random variables with mean 0 and variance $\mu(1+\mu/\kappa)$ and we can apply Bartlett's SPRT to the $Z_i(\mu)$.

The log-likelihood function of $\underset{\sim}{Z}(\mu)$ is approximated as:

$$l(\mu, \kappa) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\mu - \frac{n}{2}\log(1+\frac{\mu}{\kappa}) - \sum_{i=1}^{n} z_i(\mu)/2\mu(1+\frac{\mu}{\kappa}).$$

The transformed Bartlett's SPRT test statistics for the negative binomial generative spatial model is then given by:

$$\Lambda_n^* = l(\mu_1, \hat{\kappa}(\mu_1)) - l(\mu_0, \hat{\kappa}(\mu_0))$$

$$= \frac{n}{2}\log\frac{\mu_0}{\mu_1} + \frac{n}{2}\log\left(1+\frac{\mu_0}{\hat{\kappa}(\mu_0)}\right) - \frac{n}{2}\log\left(1+\frac{\mu_1}{\hat{\kappa}(\mu_1)}\right) + \frac{\hat{\kappa}(\mu_0)\sum_{i=1}^{n} z_i^2(\mu_0)}{\mu_0+\mu_0^2} - \frac{\hat{\kappa}(\mu_1)\sum_{i=1}^{n} z_i^2(\mu_1)}{\mu_1+\mu_1^2}$$

with acceptance of H_0 at the first n for which $\Lambda_n^* \le \log\left(\frac{\beta}{1-\alpha}\right)$, and acceptance of H_1 at the first n for which $\Lambda_n^* \ge \log\left(\frac{1-\beta}{\alpha}\right)$.

$\hat{\kappa}(\mu_0)$ and $\hat{\kappa}(\mu_1)$ denote the conditional MLE of the unknown nuisance parameter κ given a value of μ. We compute the conditional MLE of nuisance parameter κ under H_0 and H_1 by maximizing $l(\mu,\kappa)$ with respect to κ. The first derivative with respect to κ is calculated to get the conditional MLE $\hat{\kappa}(\mu_0)$ and $\hat{\kappa}(\mu_1)$:

$$\frac{\partial l(\mu,\kappa)}{\partial \kappa} = -\frac{n\mu}{2\kappa(\kappa+\mu)} - \frac{\sum_{i=1}^{n} z_i(\mu)}{2(\kappa+\mu)} \tag{48}$$

Let the right hand side of (48) equal 0 and we have the conditional MLE of κ:

$$\hat{\kappa}(\mu) = \frac{n\mu^2}{\sum_{i=1}^{n} z_i(\mu) - n\mu}.$$

A positive second derivative shows that $\hat{\kappa}(\mu)$ is the MLE of κ.

4.2.3. *OC and ASN Curves for Negative Binomial Generative Spatial Model*

Again, we simulate a large number of sequential samples and record the fraction of cases where the null hypothesis is accepted and correspondingly the average number

of samples required to make a decision. The starting sample size for Bartlett SPRT could always be decided by the first occasion initial sample study as is discussed in chapter 3. We follow the same rules to find the starting sample size as is described in chapter 3 here and decide to use 3 as the starting sample size for SRS plan and 5 for CS plan. **Table 13** shows simulated OC and ASN curves (as a function of the true μ) for the transformed Bartlett's SPRT for the two sampling plan: CS and SRS. The simulation results are based on 3,000 sample paths of transformed Bartlett's SPRT of $H_0 : \mu = 7$ versus $H_1 : \mu = 10$. The true value of κ for simulation is 3. The nominal type-1 and type-2 errors are set to 0.1.

From **Table 13** we see that the realized type-1 and type-2 error rates for the two sampling plans CS and SRS are (0.05, 0.10) and (0.04, 0.11) respectively. They are close to the nominal 0.1 values. The difference in the OC characteristics for the two sampling plans is not large. Same as table 10, **Table 13** also shows percentiles of the sample size required with the proposed sequential procedure. Again, we do see that a higher sample size is required for CS sampling plan than SRS plan. The starting sample size for CS plan is 3 trees larger than SRS plan, however we see a difference of more than 3 trees in the ASN values between the two sampling plans. **Figure 15** includes an illustrative comparison of the ASN curves that portray the penalty of larger sample size required for CS plan than the SRS plan.

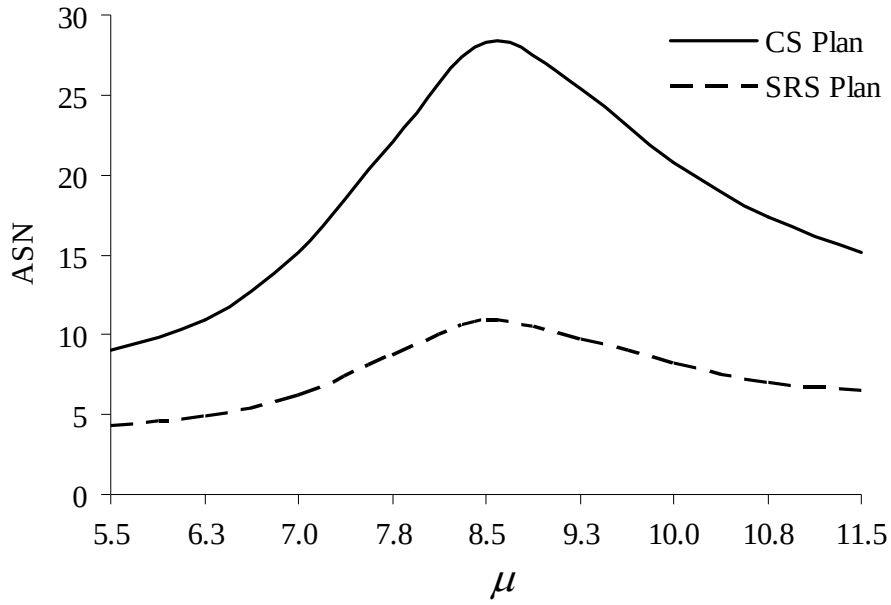

Figure 15. ANS Curves for Transformed Bartlett's SPRT with Negative Binomial Generative Spatial Model ($H_0 : \mu = 7$ versus $H_1 : \mu = 10$, $\kappa = 3$ and $\alpha = \beta = 0.1$)

In conclusion of this chapter, we want to make a comparison between our proposed generative spatial models and spatial GLMMs which have been discussed in chapter 3. As an alternative to spatial GLMMs, the simplicity of the concept and only one unknown parameter in the variance covariance structure are the generative model's appealing features. However the variance covariance structure of the generative model is fixed. The fixed covariance structure doesn't even allow the data to be uncorrelated. On the other hand, spatial GLMMs are more flexible in modeling different spatial distributions and characterizing the spatial dependencies. The catch is its computational complexity. Both of the two models could be implemented based on different practice and needs.

μ	CS Plan							SRS Plan						
	OC	Sample Size Distribution						OC	Sample Size Distribution					
		ASN	10th	25th	50th	75th	90th		ASN	10th	25th	50th	75th	90th
5.50	0.99	9.00	5	5	10	10	15	1.00	4.29	3	3	4	5	6
6.25	0.98	10.96	5	5	10	14	20	0.99	4.89	3	3	4	6	8
7.00	0.95	15.19	5	9	10	20	29	0.96	6.22	3	3	5	7	11
7.75	0.84	22.11	5	10	15	29	48	0.84	8.77	3	3	6	10	18
8.50	0.56	28.28	5	10	20	37	60	0.55	10.97	3	4	7	14	24
9.25	0.26	25.36	5	10	20	34	50	0.28	9.70	3	4	7	12	20
10.00	0.10	20.75	5	10	15	25	39	0.11	8.19	3	4	7	10	15
10.75	0.04	17.36	9	10	15	20	30	0.05	7.07	3	4	6	9	12
11.50	0.02	15.17	9	10	15	19	24	0.01	6.49	3	4	6	8	10

Table 13. OC and ASN properties for NB Generative Spatial Model Testing $H_0 : \mu = 7$ versus $H_1 : \mu = 10$ when $\kappa = 3$ and $\alpha = \beta = 0.1$.

CHAPTER 5: Future Work

5.1. *Extension of Sequential GLMM to Other Distributions*

In chapter 2, we proposed an extended Bartlett's SPRT that covers cases of Poisson GLMM with an application of designing a multi-center clinical trials study. In chapter 3, we extended Bartlett's method again to contexts where negative binomial spatial GLMM are appropriate for the pest count data. However, the application of sequential hypothesis tests to GLMM is not restricted to Poisson GLMM or negative binomial GLMM. Applications of sequential procedure to other GLMMs can be set up the same way as discussed in chapter 2 and 3, and the verification of the extension so far has to be done on an individual basis.

For example, in the needs of designing a presence-absence binomial sampling plan for our pest management application discussed in chapter 3, a logistic regression GLMM can be integrated with our proposed "out of range" sequential procedure. In that case, the response variable is of the form of presence or absence, and the link function will be the logit link of the probability of presence.

It is of interest to verify that both of our GLMM-Bartlett's procedure and the proposed subsequent occasion sequential procedure apply to GLMMs with distributions outside Poisson and negative binomial.

5.2. *Application of GLMM-Bartlett SPRT to Spatial GLMM*

In chapter 2, we propose an GLMM-Bartlett's SPRT to Poisson GLMM with independent random effect and we verify the application by simulation study. However, the extension of GLMM-Bartlett to spatial GLMM with correlated random effect is not easy. Spatial GLMM model with correlated random effects poses special computational challenges because of the high-dimensional integration required to evaluate the likelihood. Unlike the GLMM with independent random effects, Gaussian Hermite Quadrature approximation doesn't seem to be practical feasible. Other approaches such as pseudo likelihood method and MCEM algorithm have some potential, but initial efforts with each of these approximations failed. Further investigation of their applicability in the context of sequential hypothesis testing is needed. Each of these approaches is sketched in the two subsections that follow.

5.2.1. *Pseudo Likelihood Approach*

Pseudo likelihood estimation procedure has been developed to fit GLMMs with flexible covariance structures. The basic idea behind Pseudo likelihood estimation procedure is to linearize the problem so that the approach in Linear Mixed Model (LMM) could be used. Pseudo Likelihood estimation can be explained as follows (Wolfinger and O'Connell 1993):

We follow the context of GLMM described in Wolfinger and O'Connell's paper, although it is an awkward characterization of GLMM and different from our setup in the previous discussion. A data vector $\underset{\sim}{y}$ of length n satisfy:

$$\underset{\sim}{y} = \underset{\sim}{\mu} + \underset{\sim}{e}$$

with a link function $g(\cdot)$ that $g\left(\underset{\sim}{\mu}\right) = X\underset{\sim}{\alpha} + Z\underset{\sim}{\beta}$, where $\underset{\sim}{\alpha}$ is a vector of unknown fixed effects and $\underset{\sim}{\beta}$ is a vector of unobserved random effects. Assume $E(\underset{\sim}{\beta}) = \underset{\sim}{0}$ and $Cov(\underset{\sim}{\beta}) = D$, where D is unknown. Also $\underset{\sim}{e}$ is a vector of unobserved errors with $E(\underset{\sim}{e}|\underset{\sim}{\mu}) = 0$ and $Cov(\underset{\sim}{e}|\underset{\sim}{\mu}) = R_{\mu}^{1/2} R R_{\mu}^{1/2}$ where $R_{\mu}^{1/2}$ is a known diagonal matrix which contains the covariance function of the model and R is the unknown correlation matrix.

There are three approximations involved in Pseudo likelihood approach which include two analytic approximations and one probabilistic. For the first analytic approximation, let $\hat{\underset{\sim}{\alpha}}$ and $\hat{\underset{\sim}{\beta}}$ be the known estimates of $\underset{\sim}{\alpha}$ and $\underset{\sim}{\beta}$ and define

$$\hat{\underset{\sim}{\mu}} = g^{-1}(X\hat{\underset{\sim}{\alpha}} + Z\hat{\underset{\sim}{\beta}})$$

rewrite $\underset{\sim}{e}$ as

$$\underset{\sim}{e} = \underset{\sim}{y} - g^{-1}(X\underset{\sim}{\alpha} + Z\underset{\sim}{\beta}) \tag{49}$$

A first order Taylor series approximation of (49) about $(X\hat{\underset{\sim}{\alpha}} + Z\hat{\underset{\sim}{\beta}})$ yields an approximation of $\underset{\sim}{e}$ denoted by $\hat{\underset{\sim}{e}}$:

$$\hat{\underset{\sim}{e}} = \underset{\sim}{y} - g^{-1}(X\hat{\underset{\sim}{\alpha}} + Z\hat{\underset{\sim}{\beta}}) - (g^{-1})'(X\hat{\underset{\sim}{\alpha}} + Z\hat{\underset{\sim}{\beta}})(X\underset{\sim}{\alpha} + Z\underset{\sim}{\beta} - X\hat{\underset{\sim}{\alpha}} - Z\hat{\underset{\sim}{\beta}}) \tag{50}$$

Recalling that $(g^{-1})'(X\hat{\underset{\sim}{\alpha}} + Z\hat{\underset{\sim}{\beta}}) = \dfrac{1}{g'\left(g^{-1}(X\hat{\underset{\sim}{\alpha}} + Z\hat{\underset{\sim}{\beta}})\right)} = \dfrac{1}{g'(\hat{\underset{\sim}{\mu}})}$, (50) can therefore be simplified into:

$$\hat{\underset{\sim}{e}} = \underset{\sim}{y} - \hat{\underset{\sim}{\mu}} - \frac{1}{g'(\hat{\underset{\sim}{\mu}})}(X\underset{\sim}{\alpha} + Z\underset{\sim}{\beta} - X\hat{\underset{\sim}{\alpha}} - Z\hat{\underset{\sim}{\beta}}) \tag{51}$$

In the second probabilistic approximation, the conditional distribution of $\hat{\underset{\sim}{e}}$ given $\underset{\sim}{\alpha}$ and $\underset{\sim}{\beta}$ is approximated with a Gaussian distribution having the same first two moments as the conditional distribution of $\underset{\sim}{e}$, which can be expressed as

$$\hat{\underset{\sim}{e}} \big| \underset{\sim}{\beta} \sim MVN(0, R_{\mu}^{1/2} R R_{\mu}^{1/2}) \tag{52}$$

The third analytic approximation is substituting $\hat{\underset{\sim}{\mu}}$ for $\underset{\sim}{\mu}$ in the variance matrix. Now conditional on $(\underset{\sim}{\alpha}, \underset{\sim}{\beta})$ and combine (51) and (52), we have

$$\underset{\sim}{y} - \hat{\underset{\sim}{\mu}} - \frac{1}{g'(\hat{\underset{\sim}{\mu}})}(X\underset{\sim}{\alpha} + Z\underset{\sim}{\beta} - X\hat{\underset{\sim}{\alpha}} - Z\hat{\underset{\sim}{\beta}}) \sim MVN(0, R_{\hat{\mu}}^{1/2} R R_{\hat{\mu}}^{1/2}) \tag{53}$$

multiply the left side of (53) by $g'(\hat{\underset{\sim}{\mu}})$, conditional on $(\underset{\sim}{\alpha}, \underset{\sim}{\beta})$ we have

$$g'(\hat{\underset{\sim}{\mu}})(\underset{\sim}{y} - \hat{\underset{\sim}{\mu}}) + g(\hat{\underset{\sim}{\mu}}) \sim MVN(X\underset{\sim}{\alpha} + Z\underset{\sim}{\beta},\ g'(\hat{\underset{\sim}{\mu}}) R_{\hat{\mu}}^{1/2} R R_{\hat{\mu}}^{1/2} g'(\hat{\underset{\sim}{\mu}}))$$

If we define

$$v = g'(\hat{\underset{\sim}{\mu}})(\underset{\sim}{y} - \hat{\underset{\sim}{\mu}}) + g(\hat{\underset{\sim}{\mu}})$$

unconditionally, it follows that

$$v \sim MVN(X\underset{\sim}{\alpha},\ W^{-1/2} R W^{-1/2} + ZDZ')$$

where $W = R_{\hat{\mu}}^{-1}\left(g'(\hat{\underset{\sim}{\mu}})\right)^{-2}$. Now the problem has been simplified into classic LMM context and iteration algorithm can be applied to obtain the estimates of the unknown parameters.

Pseudo likelihood method has been applied to spatial GLMM (Schabenberger and Gotway 2005). However due to the complexity of the approximations involved in the

approach, the properties of estimations by pseudo likelihood method yet remain to be established.

5.2.2. *Monte Carlo EM Algorithm*

As another computational alternative besides pseudo likelihood approach, we discuss the application of Monte Carlo EM (MCEM) algorithm in this section. We start introducing MCEM algorithm by EM algorithm.

EM algorithm is an iterative algorithm for calculating ML (or REML) estimates, and its name standing for expectation maximization: it alternates between calculating conditional expected values and maximizing simplified likelihoods. The EM algorithm is designed for situations where the recognition or invention of "missing" data simplifies the maximum likelihood calculations. In mixed models the random effects that introduce correlation in the model are typically assumed to be the missing data. Filling in missing data by the EM algorithm and treating them as fixed known values often simplify the problem (McCulloch and Searle 2001).

Consider the most general GLMM model where $Y_i | \underset{\sim}{\beta}$ are independent but not identically distributed. A known link function $g(\cdot)$ is assumed that $g\left(\underset{\sim}{\mu}\right) = X\underset{\sim}{\alpha} + Z\underset{\sim}{\beta}$, where $\underset{\sim}{\alpha}$ is a vector of unknown fixed effects and $\underset{\sim}{\beta}$ is a vector of unobserved random effects. Assume $E(\underset{\sim}{\beta}) = \underset{\sim}{0}$ and $Cov(\underset{\sim}{\beta}) = D$, where D is unknown. Hence, EM algorithm Assume the random effect $\underset{\sim}{\beta}$ to be the missing data. Denote $\underset{\sim}{y} = (y_1, \cdots, y_n)'$, then one of the biggest difference between EM algorithm and other approximation

algorithms is that EM algorithm is based on the log likelihood of the complete data $\underset{\sim}{w}' = (\underset{\sim}{y}', \underset{\sim}{\beta}')$. The complete data log likelihood can be written as

$$\log f(\underset{\sim}{y}, \underset{\sim}{\beta}) = \log f(\underset{\sim}{y}|\underset{\sim}{\beta}) + \log f(\underset{\sim}{\beta})$$

$$= \sum_{i=1}^{n} \log f(y_i|\underset{\sim}{\beta}) + \log f(\underset{\sim}{\beta}) \qquad (54)$$

From (54), we see the advantages of assuming the random effect as the missing data: first, the y_i are independent conditional on $\underset{\sim}{\beta}$. Second, $\underset{\sim}{\alpha}$ enters only the first portion of the log likelihood and *D* enters only the portion coming from the random effects through $f(\underset{\sim}{\beta})$. Thus, EM algorithm can be conducted using the following iterative procedure:

i. Choose starting value $\underset{\sim}{\alpha}^{(0)}$, $\mathrm{D}^{(0)}$ and set $h = 0$.

ii. Calculate (with expectations evaluated under current values):

(a) $\underset{\sim}{\alpha}^{(h+1)}$ to maximize $E_{\underset{\sim}{\beta}|\underset{\sim}{Y}}[\log f_{\underset{\sim}{Y}|\underset{\sim}{\beta}}(\underset{\sim}{y}|\underset{\sim}{\beta}, \underset{\sim}{\alpha})]$

(b) $\mathrm{D}^{(h+1)}$ to maximize $E_{\underset{\sim}{\beta}|\underset{\sim}{Y}}[\log f_{\underset{\sim}{\beta}}(\underset{\sim}{\beta}|D)]$

(c) Set $h = h + 1$

iii. If convergence, declare the current values to be the MLEs; otherwise return to step ii.

Generally the expectations in step ii(b) can be computed in closed form. When the closed form solutions is not available, one can produce random draws from the

conditional distribution of $\underset{\sim}{\beta}$ given $\underset{\sim}{Y} = \underset{\sim}{y}$, and use these draws to obtain Monte Carlo approximations to the required expectations, which is the so called MCEM algorithm.

There are several ways to generate draws from a conditional distribution, Metroplis-Hastings algorithm has been used for GLMM problems and we refer the details to McCulloch (1994, 1997, 2001). Incorporating the Metroplis step into the EM algorithm gives a MCEM algorithm as follows:

i. Choose starting value $\underset{\sim}{\alpha}^{(0)}$, $\mathrm{D}^{(0)}$ and set $h = 0$.

ii. Generate draws $\beta^{(1)}, \beta^{(2)}, \cdots, \beta^{(H)}$ from the conditional distribution of $\underset{\sim}{\beta}$ given $\underset{\sim}{Y} = \underset{\sim}{y}$ using Metroplis algorithm.

(a) Calculate $\underset{\sim}{\alpha}^{(\mathrm{h}+1)}$ to maximize $\frac{1}{H}\sum_{k=1}^{H} \log f_{\underset{\sim}{Y}|\underset{\sim}{\beta}}(\underset{\sim}{y} \big| \underset{\sim}{\beta}^{(k)}, \underset{\sim}{\alpha})$

(b) $\mathrm{D}^{(\mathrm{h}+1)}$ to maximize $\frac{1}{H}\sum_{k=1}^{H} \log f_{\underset{\sim}{\beta}}(\underset{\sim}{\beta}^{(k)} \big| D)$

(c) Set $h = h + 1$

iii. If convergence, declare the current values to be the MLEs; otherwise return to step ii.

MCEM algorithm has been applied to spatial GLMM (Booth and Hobert 1999, Zhang 2002, Christensen et. al. 2006). However applying MCEM to spatial GLMM together with sequential procedure may be trouble-some in practice since the performance of the algorithm heavily depends on the observed data. Constructing an algorithm with robust mixing and convergence characteristics for sequential procedure will be a very desirable future work.

5.3. *Extension of Sequential GLMMs to Group Sequential Contexts*

In chapter 2, we extend single-individual SPRT to GLMM in which the observations are non-identically and non-independently distributed. Group sequential methods have gained in popularity in particular among clinical trials studies recently due to implementation advantages. Interesting further works would be a study of group sequential methods for GLMM and a comparison with single-individual sequential methods for GLMM.

BIBLIOGRAPHY

Baksh, F.M., Todd, S., Whitehead, J. and Lucini, M.M., 2005. Design Consideration in the Sequential Analysis of Matched Case-control Data. *Statistics in Medicine* 24, 853-867.

Barchia, I. M., G. A. Herron, and Gilmour, A. R. 2003. Use of a Generalized Linear Mixed Model to Reduce Excessive Heterogeneity in Petroleum Spray Oil Bioassay Data. *Journal of Economic Entomology* 96(3): 983-989.

Bartlett, M. S. 1946. The large sample theory of sequential tests. *Proc. Cambridge Philos. Soc.* 42: 239 – 244.

Bennett, K. E. et. al. 2008. Blood-Feeding Behavior of Vesicular Stomatitis Virus Infected Culicoides Sonorensis (Diptera: Ceratopogonidae). *Journal of Medical Entomology* 45: 921-926.

Bianchi E. J. J. A., P. W. Goedhart and J. M. Baveco. 2008. Enhanced Pest Control in Cabbage Crops Near Forest in The Netherlands. *Landsc. Ecol.* 23: 595:602.

Binns, M. R., J. P. Nyrop, and Van Der Werf, W. 2000. *Sampling and Monitoring in Crop Protection*. CABI Publishing.

Booth, J. G. and Hobert, J. P. 1999. Maximizing Generalized Linear Mixed Model Likelihoods with an Automated Monte Carlo EM Algorithm. *Journal of the Royal Statistical Society* 671: 265:285.

Breslow, N. E., and Clayton, D.G. 1993. Approximate Inference in Generalized Linear Mixed Models. *Journal of American Statistics Association* 88: 9-25.

Candy S. G. 2000. The Application of Generalized Linear Mixed Models to Multi-level Sampling for Insect Population Monitoring. *Environ. Ecol. Stat.* 7: 217-238.

Christensen, O. F., Roberts, G. O. and Sköld M., R. 2006. Robust Markov Chain Monte Carlo Methods for Spatial Generalized Linear Mixed Models. *Journal of Computational and Graphical Statistics* 15: 1-17.

Coad, S.D. and Ivanova, A., 2005. The Use of the Triangular Test with Response-adaptive Treatment Allocation. *Statistics in Medicine* 24, 1483-1493.

Elias, S. P. et. al. 2006. Deer Browse Resistant Exotic-Invasive Understory: An Indicator of Elevated Human Risk of Exposure to Ixodes Scapularis (Acari: Ixodidae) in Southern Coastal Maine Woodlands. *Journal of Medical Entomology* 43: 1142-1152.

Elston, D. A., Moss, R., Boulinier, T., Arrowsmith C., and Lambin, X. 2001. Analysis of Aggregation, A Worked Example: numbers of Ticks on Red Grouse chicks. *Parasitology* 122: 563:569.

Fowler, G. W. and Lynch, A. M. 1987. Sampling Plans in Insect Pest Management Based on Wald's Sequential Probability Ratio Test. *Environmental Entomology* 16: 345-354

Ghosh, B. K., 1970. *Sequential Tests of Statistical Hypotheses*. Addison-Wesley Publishing Company, Inc.

Givens, G. H. and Hoeting J. A., 2005. *Computational Statistics*. A John Wiley & Sons, Inc.

Govindarajulu, Zakkula, 2004. *Sequential Statistics*. World Scientific Publishing Co. Pte. Ltd.

Gotway, C. A., and Stroup, W. W. 1997. A Generalized Linear Model Approach to Spatial Data Analysis and Prediction. *Journal of Agricultural, Biological and Environmental Statistics* 2(2) : 157-178.

Gozé, E., S. Nibouche, and Deguine, J. P. 2003. Spatial and Probability Distribution of Helicoverpa Armigera (Hubner) (Lepidoptera : Noctuidae) in Cotton: Systematic Sampling, Exact Confidence Intervals and Sequential Test. *Environmental Entomology* 32(5): 1203-1210.

Ifoulis, A. A. and Savopoulou-Soultani, M. 2006. Use of Geostatistical Analysis to Characterize the Spatial Distribution of Lobesia botrana (Lepidoptera: Tortricidae) larvae in Northern Greece. *Environmental Entomology* 35(2): 497-506.

Lee, JW and DeMets, DL., 1991. Sequential Comparison of Changes with Repeated Measurements Data. *Journal of the American Statistical Association* 86, 757-762.

McCulloch, C. E. 1994. Maximum Likelihood Estimation of Variance Components for Binary Data. *Journal of the American Statistical Association* 89: 330-335.

McCulloch, C. E. 1997. Maximum Likelihood Algorithms for Generalized Linear Mixed Models. *Journal of the American Statistical Association* 92: 162-170.

McCulloch, C. E. and Searle, S. R. 2001. *Generalized, Linear and Mixed Models.* Wiley, New York

Mulekar, M. S., Young, L. J. and Young, J. H. 1993. Testing Insect Population Density Relative to Critical Densities with 2-SPRT. *Environmental Entomology* 22: 346-351.

Paulsen, K.B., Bremmelgaard, A., Sorensen, A.I., Raahave, D., Petersen, J.V., 1994. Estimated Costs of Postoperative Wound Infections: A Case-Control Study of Marginal Hospital and Social Security Costs. *Epidemiology and Infection* 113, 283-295.

Paterson, S. and Lello, J. 2003. Mixed Models: Getting the Best Use of Parasitological Data. *Parasitology* 19: 370-375.

Ramírez-Dávila, J. F. and Porcayo-Camargo, E. 2008. Spatial Distribution of the Nymphs of Jacobiasca lybica (Hemiptera: Cicadellidae) in a Vineyard in Andalucia, Spain. *Revista Colombiana De Entomologia* 34(2): 169-175.

SAS Institute Inc. 2008. *SAS/STAT® 9.2 User's Guide: The GLIMMIX Procedure (Book Excerpt)*. Cary, NC: SAS Institute Inc.

Schabenberger, O. and Gotway, C. A. 2005. *Statistical Methods for Spatial Data Analysis.* Chapman & Hall/CRC Press.

Scharfstein, D.O., Tsiatis, A.A., and Robins, J.M., 1997. Semiparametric Efficiency and Its Implication on the Design and Analysis of Group-Sequential Studies. *Journal of the American Statistical Association* 92, 1342-1350.

Schotzko, D. J. and Okeeffe, L. E. 1989. Geostatistical Description of the Spatial-Distribution of Lygus Hesperus (Heteroptera, Miridae) in lentils. *Journal of Economic Entomology* 82(5): 1277-1288.

Shah, P., D. R. Jeske, and Luck, R. 2009. Sequential Hypothesis Testing Techniques for Pest Count Models with Nuisance Parameters. *Journal of Economic Entomology* 102: 1970-1976.

Spiessens, B., Lesaffre, E., Verbeke, G., Kim, K. and DeMets, D.L., 2000. An Overview of Group Sequential Methods in Longitudinal Clinical Trials. *Statistical Methods in Medical Research* 9, 497-515.

Spiessens, B., Lesaffre, E., Verbeke, G. and Kim, K., 2002. Group Sequential Methods for an Ordinal Logistic Random-Effects Model Under Misspecification. *Biometrics* 58, 569-575.

Spiessens, B., Lesaffre, E. and Verbeke, G., 2003. A Comparison of Group Sequential Methods for Binary Longitudinal Data. *Statistics in Medicine* 22, 501-515.

Takakura, K. 2009. Reconsiderations on Evaluating Methodology of Repellent Effects: Validation of Indices and Statistical Analyses. *Journal of Economic Entomology* 102: 1977-1984.

UC Integrated Pest Management. 1992. UC IPM Pest Management Guidelines: Grape, UC ANR Publication 3343, Davis, CA.

Wald, A., 1947. *Sequential Analysis*. Dover Publications.

Wald, A. and Wolfowitz, J., 1948. Optimum Character of the Sequential Probability Ratio Test. *Ann. Math. Statist* 19 326-339.

Whitehead, J., 1997. *The Design and Analysis of Sequential Clinical Trials*. 2nd edition. John Wiley.

Williams, L., Schotzko, D. J. and Mccaffrey, J. P. 1992. Geostatistical Description of the Spatial Distribution of Limonius californicus (Coleoptera, Elateridae) Wireworms in the Northwestern United States, with Comments on Sampling. *Environmental Entomology* 21(5): 983-995.

Wolfinger, R. and O'Connell, M. 1993. Generalized Linear Mixed Models: A Pseudo-likelihood Approach. *Journal of Statistical Computation and Simulation* 48: 233-243.

Wu, MC and Lan, KK., 1992. Sequential Monitoring for Comparison of Changes in a Response Variable in Clinical Studies. *Biometrics* 48, 765-779.

Young, L. J. and Young, J. H. 1998. *Statistical Ecology*. Kluwer Academic Publishers.

Zhang, H. 2002. On Estimation and Prediction for Spatial Generalized Linear Mixed Models. *Biometrics* 58: 129-136.