Power Analysis of Finite Mixtures of Poisson Distributions

A Dissertation Presented

by

Christine Marie Brady

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

May 2008

Stony Brook University

The Graduate School

Christine Marie Brady

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Nancy R. Mendell – Dissertation Advisor
Professor
Applied Mathematics and Statistics Department

Hongshik Ahn – Chairperson of Defense
Professor
Applied Mathematics and Statistics Department

Wei Zhu
Professor
Applied Mathematics and Statistics Department

Yungtai Lo
Associate Professor
Department of Community and Preventive Medicine
Mount Sinai School of Medicine

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Power Analysis of Finite Mixtures of Poisson Distributions

by

Christine Marie Brady

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2008

One of the most common types of studies that occur is the comparison of responses between a control group and treatment group. Typically, it is assumed that each group is homogeneous. However, when dealing with count data the phenomenon of overdispersion often occurs. This phenomenon may be due to heterogeneity that exists within the group. In such cases, a mixture of distributions is often used to account for such heterogeneity.

We developed a likelihood ratio test for comparing two groups assuming a two-component Poisson mixture exists within each group. We conducted a power study for the family of alternatives with a one parameter difference from the null hypothesis. For each model considered, we compared the power for the Likelihood Ratio Test to the Welch-Satterthwaite t-test, Wilcoxon Test, and Adjusted Wilcoxon Test.

The power study was done using a user-friendly software that we developed which simulated our data. The software obtains the maximum likelihood estimates of the parameters under the null hypothesis, such that the control and treatment groups each follow a two-component Poisson mixture with equal mixing proportions and component means. As well, it can compute the MLE's for the two groups differing in either mixing proportions or exactly one component mean. In addition to simulating data, our program has the capability to input actual data and run similar studies.

We compared the power of the Likelihood Ratio Test using the asymptotic 95^{th} percentile critical value of the chi-squared distribution with one degree of freedom and the 95^{th} percentile asymptotic critical value of the standard normal distribution for the other three tests for sample sizes of 100 and 250 per group. As well, we investigated the empirical null distribution for the LRT. We conducted a similar power study for sample sizes of 100 per group using the 95^{th} percentile empirical value. Generally speaking, the LRT was found to be significantly more powerful than the other tests considered.

We applied our testing procedure for comparing two groups of two-component Poisson mixtures for two sets of count data that were provided. One data set that we studied consisted of the number of fibromas which existed on patients suffering from the disease tuberous sclerosis. The other data set that we applied our procedure to consisted of the number of deviant verbalizations from a study on schizophrenia.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost I would like to thank my doctoral advisor, Dr. Nancy R. Mendell, for all of her guidance and support throughout the years. She has always had faith and confidence in me. Without her, I would not have been able to accomplish this. She made my time at Stony Brook University an enjoyable and unforgettable experience.

I would like to thank my dissertation committee members: Dr. Hongshik Ahn, Dr. Wei Zhu and Dr. Yungtai Lo for contributing their time and for all of their helpful suggestions.

One person in particular that I would like to thank is Greg Davidson. He believed in me even when I did not believe in myself. Without his help this dissertation would not have been possible.

Lastly, I would like to thank my family for their unconditional support. Their confidence, faith, and constant encouragement have made me the person I am today. Without them, I could have never done this.

Chapter 1

Introduction and Literature Review

1.1 Introduction

When conducting a study, one common assumption is that the population under consideration is homogeneous. Under this assumption, one may try and use a single density function to model the population. However, sometimes using a single probability distribution function may in fact lead to incorrect results. This may be due to the fact that the population is heterogeneous. One common attempt to overcome this problem is to use a mixture of distributions, such as Poisson distributions, to reflect such heterogeneity. By using a finite mixture to model heterogeneity amongst a population, it is possible to increase the power and precision of a test.

In general, a finite mixture model has the form:

$$f\left(Y;\underline{X}\right)=\sum_{j=1}^{k}\pi_j\cdot g_j\left(Y;\underline{\theta}_j\right) \tag{1.1}$$

where $g_j\left(Y;\underline{\theta}_j\right)$ are called the component densities, the π_j for $j=1,\ldots,k$ are called the mixing proportions, such that $\sum_{j=1}^{k}\pi_j=1$.

Finite mixture models have been used in a variety of fields such as biology, medicine, and finance to model heterogeneity.   Some common examples of heterogeneity that may exist within a population may be due to gender, age, genetics, or presence of a disease.

The following literature review concentrates on the history of mixtures, specifically past studies involving mixtures of Poisson distributions and recent advancements in regression analysis incorporating mixtures. A review of estimation techniques for the parameters of a mixture and hypothesis tests incorporating mixtures. As well, a comparison of the tests considered in our power study is given.

1.2 Literature Review

1.2.1 General Background on Mixture Models

The study of mixtures models is over a century old. One of the first major contributions to the study of mixtures models was conducted by Karl Pearson in 1894. He fitted a mixture of two univariate normal distributions with unequal component means and variances. A majority of the study of mixture models has dealt with normal distributions. McLachlan and Basford (1988) concentrated on modeling mixtures of normal distributions for a variety of fields. An extensive review on mixture models are included in the books written by Everitt and Hand (1981), Titterington et al. (1985) and McLachlan and Peel (2000). Mixtures of discrete distributions, such as the Poisson and binomial distribution, have been studied by various researchers such as Blischke (1965) and Schilling (1947).

When dealing with count data, the Poisson distribution is a natural choice of a probability distribution. However, one common problem that is exhibited when using a single Poisson distribution to model such data is overdispersion. Overdispersion is a violation of the mean-variance relationship for the Poisson distribution, where the variance is greater than the mean. Occasionally, underdispersion occurs where the sample variance is less than the mean (Faddy, 1994). A finite mixture of Poisson distributions is commonly used to model the heterogeneity that may be the reason for the apparent overdispersion.

1.2.2 Zero Inflated Poisson (ZIP) and Poisson Mixture Regression Models

Poisson Regression Models have widely been used to model count data (McCullagh and Nelder, 1989). However for overdispersed count data, the use of finite mixture Poisson regression models has been used to model heterogeneity (Wang et al 1996). They stated that if the dispersion is ignored, a single Poisson regression model may lead to seriously biased estimates of the parameters which would lead to incorrect inferences (Wang et al., 1996).

Applications of mixture Poisson Regression models have been done by Wang et al (1996), who applied a mixture Poisson regression model to analyze daily epileptic seizure frequency. Xiang et al. (2005) developed influence diagnostics for a two-component Poisson mixture regression model. They claimed to be able to identify a cluster of observations that may be causing overdispersion in the model. Therefore, using such a mixture would result in valid conclusions about the population. Xiang et al. (2005) applied their method to two count data sets exhibiting overdispersion arising from public health.

One of the more recent interests of investigation of Poisson mixture models has been the Zero-Inflated Poisson (ZIP). This can be viewed as a special case of the two-component mixture model, where the first component mean is zero. A mixture such as this occurs when dealing with count data that exhibits an excess amount of zero counts. Meng (1997) reviewed the history of analyzing count data with excess zeros. Cohen (1963) and Johnson and Kotz (1969) studied ZIP models without covariates. Lambert (1992) introduced a ZIP regression model, which deals with zero inflated data with covariates. Zero inflated Poisson models have as well been considered by Heilbron (1989), Gupta et al. (1996), Bohning (1998), Bohning et al. (1999), Li et al. (1999), Mullahy (1997), Fong and Yip (1993), Johnson et al. (1992), Campbell et al. (1991), and Xie et al. (2001).

1.2.3 Methods to Estimate Parameters in Mixture

In order to estimate the parameters for a mixture, it must first be verified that the mixture is identifiable. Identifiable means that it must be uniquely characterized so that two distinct sets of parameters cannot yield the same mixed distribution. Teicher (1960) proved that the class of finite mixtures of Poisson distributions is identifiable.

There are various methods that exist to estimate the parameters in a mixture model. A few of the main techniques used have been method of moments, minimum distance, and maximum likelihood estimation. The use of high speed computers in recent years has concentrated research on the method of maximum likelihood. Titterington (1996) commented that estimation has been a main topic of interest since there are no explicit formulas for the maximum likelihood estimates.

One of the oldest estimation techniques is the method of moments. Pearson (1894) estimated the parameters of the two-component normal mixture using this method. Rider (1961; 1962) used moment estimation for mixtures of Poisson distributions. Cohen (1965) estimated a two-component Poisson mixture using the first two sample moments and a third equation based on the frequency in the zero cell. He also estimated a two-component Poisson mixture missing zero-cell frequencies using factorial moments. John (1970) derived the moment estimators and asymptotic distributions for 2-component Poisson mixtures. Moment estimates of binomial and Poisson distributions have been studied also by Pearson (1915) , Muench (1936, 1938), Arley and Buch (1950), Rider (1962), Blischke (1962, 1964) and Cohen (1963).

The use of minimum-distance method has also been applied to mixtures by several researchers. Titterington et al. (1985) as well as McLachlan and Peel (2000) have reviewed several minimum distance methods for mixture models. One distance that has received much attention has been the Hellinger distance. Woodward et al. (1995) and Cutler and Cordero-Brana (1996) applied the Hellinger distance to two-component normal mixtures. Simpson (1987) applied this method to count data and showed it worked well with data that contained outliers. Lindsay (1994) compared the MHD method to ML method. It was shown that typically the maximum likelihood method worked better for well-specified models compared to the MHD.

Karlis and Xekalaki (1998) compared the Hellinger distance method to maximum likelihood method specifically for Poisson mixtures. They found that if the model was well specified and sample size was large that the two methods were comparable. However given outliers in a data set, the MHD method resulted in better estimates compared to the ML method, similar to the results of Simpson (1987).

The most common technique of fitting mixture models is the maximum likelihood method. Rao (1948) is attributed to being one of the first researchers to use this method. He estimated the parameters of a two-component mixture of univariate normal distributions with equal variances by applying Fisher's scoring method. Various researchers such as Baker (1940), Mendenhall and Hader (1958), Day (1969) and Wolfe (1965, 1967, 1970) considered this method as well. However, due to computational reasons this method was not very feasible in the past. As computers became more advance, MLE method was considered for a variety of mixture distributions.

Karlis and Xekalaki (2003a) compared the method of moments to the maximum likelihood method for finite mixtures of Poisson distributions. It was found that the ML method is favorable compared to the method of moments method for finite Poisson mixtures in terms of both small sample size and asymptotic efficiency. The method of moment estimates are not always attainable and if so, the parameter estimates have a higher variance compared to the MLE's. They also proposed a modification of the method of moment technique, called the zero frequency method, where the third moment equation is replaced by the zero frequency equation. They found that the zero frequency method was more efficient for distributions with low means compared to the method of moments. This method worked well when there was an excess count of zeros, which is common in mixtures of Poisson distributions. They felt it was a possible alternative method to the maximum likelihood method for small values of means. As well, Tan and Chang (1972) showed that the maximum likelihood estimation method was superior to that of method of moments for normal mixtures.

The crucial paper that stimulated interest in modeling heterogeneous data by maximum likelihood was written by Dempster, Laird, and Rubin (1977) regarding the EM algorithm, since mixtures can be considered an incomplete data problem. However, the idea of the EM algorithm was thought of far before 1977. Newcomb (1886) suggested using an iterative approach to compute the MLE of the common mean of a mixture in known proportions of a finite number of univariate normal populations with common variances. And, Hasselbad (1969) used an iterative method to find the maximum likelihood estimates for finite Poisson mixtures. Redner and Walker (1984) provided a history of the various methods, concentrating on the maximum likelihood estimation method along with the EM algorithm.

There are various concerns when applying the EM Algorithm. One of the concerns with the EM algorithm is the choice of starting values. A review of the literature on choice of starting values is given in Chapter 2. Another main criticism of the EM algorithm is the slow convergence rate. Hasselbad (1969) fit a mixture of two Poisson's using the EM algorithm with the initial values for the mixing proportion and component means set to the moment estimates. The EM algorithm took over 1000 iterations to converge. However, the maximum likelihood Poisson mixture fit the data very well. As well, Everitt and Hand (1981) reported on the convergence rate for a sample of 200 observations drawn from a four-component Poisson Mixture. For one set of starting values, the EM algorithm took 192 iterations. However for another set of initial values, the EM algorithm took 365 iterations to converge.

There have been many attempts made at speeding up the convergence rate of the EM algorithm. The Incremental EM (IEM), Sparse EM (SPEM) and Lazy EM are some examples of modifications of the EM algorithm to speed up of the convergence rate (McLachlan and Peel, 2000). Karlis and Xekalaki (1999b) showed that for the one-parameter exponential family one of the estimating equations for the MLE is the first moment equation. Using this result, it was suggested that the speed of convergence for the EM algorithm can be improved due to a reduction in computational time. The result was applied to two-component mixtures of normal distributions and Poisson distributions and a significant improvement in convergence rates was found.

1.2.4 Hypothesis Tests Involving Mixture Models

A majority of the research has been on tests for homogeneity, a single density function, versus heterogeneity, such as a two-component mixture model for a single sample. The homogeneity case is a special version of the two-component mixture where the mixing proportion equals 0 or 1 or the means of the two components are equal. Titterington et al. (1985) considered the mixture of two Poisson's for a data set consisting of the number of death notices for women aged 80 years and older from the Times newspaper for each day from 1910 to 1912. The fit of a single Poisson distribution resulted is a very poor fit whereas a mixture modeled the data much better.

One major area of study has been on the choice of the number of component density functions in the mixture model. McLachlan and Peel (2000) devoted an entire chapter on different methods to assess the number of components in a mixture. One such method of testing the number of components in a mixture model is using a Likelihood Ratio Test, where m components versus m+1 components are tested. However, the use of the LRT has problems since the form of the null distribution is unknown. It has been shown that the regularity conditions for the LRT do not hold true for mixture models (Self and Liang, 1987 McLachlan and Peel 2000).

Wolfe (1971) conducted one of the first simulation studies on such a Likelihood Ratio test for a mixture of normal distributions. Simar (1976), Symons et al. (1983), Bohning et al, (1994), Leroux and Puterman (1992) have considered this problem for mixtures of Poisson distributions. A majority of the literature has concentrated on testing a homogeneous Poisson distribution versus a two-component Poisson mixture, which is a special case of this type of LRT. Chen and Chen (2001) studied the asymptotic behavior for the LRT for testing homogeneity versus a two-component mixture for normal, binomial and Poisson distributions.

Karlis and Xekalaki (1999a) derived a procedure that determines the optimal number of components in a mixture of Poisson distributions. They formulated a sequential method that adopts the LRT while using the bootstrapping approach for constructing the null distribution of the test statistic at each stage. The method not only reveals the number of components in the mixture but as well a goodness of fit test. They showed that when comparing a single Poisson versus a two-component mixture, the power of their test increases as the distance between the component means increase. This result is similar for the case of normal mixtures as shown by Mendell et al. (1991). In general when testing m components versus m+1 components, the power was shown to be low when the component means were close and one component has a small mixing probability. The null distribution seemed to rely heavily on the number of components and sample size. With regards to the asymptotic distribution of the test statistic, it appeared that the chi-squared with one degree of freedom did not hold, which is the standard choice for the LRT under these conditions.

Karlis and Xekalaki (2000) described an alternative test to the LRT, called the Hellinger deviance test (HDT), for comparing a single Poisson versus a two-component mixture. The HDT is based on the Hellinger distance. In their study, they compared the power of HDT to the LRT. They found that the HDT seldom resulted in lower power than the LRT for various two-component alternatives. However, they did show for contaminated data sets that the HDT was more robust compared to the LRT.

1.2.5 Two-Independent Group Comparison

Another area of study that has received attention has been two group comparison involving mixtures.

The general two-group mixture model is defined as:

$$f\left(Y;X\right)=\sum_{j=1}^{k}\pi_{xj}\cdot g_{xj}\left(Y;\underline{\theta}_{xj}\right) \quad \text{for } x=1,2 \tag{1.2}$$

where x denotes the group, $g_{xj}\left(Y;\underline{\theta}_{xj}\right)$ is the density of the jth component in group x, $\underline{\theta}_{xj}$ denotes the parameters for g_{xj} and π_{xj} is the mixing proportion for the jth component in group X.

Good (1979) considered a special case of the two-group mixture model involving normal distributions, where the control group was homogeneous and a two-component mixture existed in the treatment group. The mixture considered in the treatment group consisted of a proportion of the sample having the same distribution as the control group and the remaining proportion having the same distribution but a shifted mean. The hypotheses that Good (1979) considered were:

$$\begin{aligned} H_o &: G_2(X) = G_1(X) \\ H_1 &: G_2(X) = p \cdot G_1(X - \Delta) + (1-p) \cdot G_1(X) \end{aligned} \quad (1.3)$$

where $0 < p \leq 1$, $\Delta \neq 0$, and G_1 and G_2 denote the cdf for the control group and treatment group respectively.

Good (1979) suggested that there would be a reduction in the power of the t-test and that the Wilcoxon Rank Sum test would perform even poorer. The reduction in power was claimed to be due to a decrease in the absolute difference between means and to an increase in the variance in the treatment group. He suggested an alternative randomization test that would take into account such properties due to a mixture:

$$v(\theta) = \theta\left(\frac{1}{n} + \frac{1}{m}\right)^{-1} \left(\bar{X} - \bar{Y}\right)^2 + (1-\theta) S_y^{\,2} \text{ for } 0 \leq \theta \leq 1 \quad (1.4)$$

where $\bar{X}, S_x^{\,2}, n\left(\bar{Y}, S_y^{\,2}, m\right)$ denote the mean, sum of squares of deviations about the mean, and size of the control (treatment) sample.

Good investigated the power of $v(0.67)$, suggesting that it would be sensitive a mean shift and increase in variance in the treatment group. It was found through simulation that the power of their test was comparable to that of the ordinary t-test in detecting a treatment effect in the presence of non-responders.

Boos and Brownie (1986) followed up a study on the randomization test that Good (1979) had suggested. They believed that the Good (1979) test statistic was effective when the mixing proportion in the treatment group mixture was small.

However, they found that when the mixing proportion was at least 0.6, that the Wilcoxon rank sum test, t-test and $v(0.67)$ test all appeared to be as effective as one another. Boos and Brownie (1986) suggested that the use of the Wilcoxon rank sum test or t-test are favorable in such situations since they are much easier to apply and simpler to interpret.

The use of mixtures to model non-response in treatment groups has as well been considered by various other researchers (Salsburg, 1986; Conover and Salsburg, 1988; Razzaghi and Nanthakumar, 1992, 1994; Razzaghi and Kodell 2000). Lo et al (2002) considered a two-sample permutation test incorporating mixtures using the likelihood ratio test statistic. Their study concentrated on comparing controls versus relatives of schizophrenia patients for three different measurements. They tested to see if these two groups of individuals arise from the same distribution versus the alternative that an unknown proportion of the relatives arise from a different distribution. The null hypothesis was that both groups had equal component means. They considered three different alternatives involving the normal and exponential distributions, for which the relatives consisted of a two-component mixture in which one of the means is the same mean as that of the controls. The likelihood ratio test did not follow an asymptotic chi-squared distribution with degrees of freedom equal to the difference in the number of parameters between the two hypotheses. This discrepancy may be due to either small sample size or the fact that under the null hypothesis the mixing proportions are on the boundary of the parameter space and therefore are not identifiable (Hartigan, 1985). An empirical power study was conducted to assess the power of the likelihood ratio test they derived for small and moderate samples. Based on their results, the power of the test depended on sample sizes, the mixing proportion, difference of component means and ratio of component variances.

Duan (2005) derived a generalized likelihood ratio test for comparing two groups under the assumption of a two-component normal mixture within each group. A study was conducted to compare the power of the LRT compared to that of the Welch-Satterthwaite t-test and Wilcoxon Rank Sum test. She found for sample sizes of one hundred per group that the LRT was significantly more powerful than the other tests when differences existed in either the mixing proportions or exactly one component mean. For the other models that they studied, the power of the t-test was comparable to the LRT if not slightly more powerful. Duan et al. (2005) used the generalized likelihood ratio test comparing sex adjusted neurological values of alcohol-dependent individuals to controls assuming that within each group there existed a two-component normal mixture distribution.

Kim (2007) considered the LRT of normal mixtures for testing equality of mixing proportions in two groups under the assumption that the two mixtures are the same but their parameters are unknown. He derived the asymptotic power of this test in terms of a non-central chi-squared distribution. As well, they provided a comparison of power for the LRT of mixtures to the chi-square test of independence.

1.2.6 Review of Statistical Power of Tests Comparing Two Independent Group Means

A natural test to see whether two independent samples have been drawn from different distributions with unequal means is Student's t-test. This test assumes the populations drawn from are normally distributed with a common population variance. An alternative non-parametric test, which relaxes this assumption of normality, is the Wilcoxon rank sum test (Wilcoxon, 1945). As a result, interest has been in comparing the power for these two tests.

For the Poisson distribution for which there exists a mean-variance relationship, we are interested in comparing the power and size of Wilcoxon Rank Sum test to the Welch-Satterthwaite t-test (Welch, 1938; Satterthwaite, 1946), which assumes unequal population variances. The use of a two-sample independent mean test assuming equality of variance versus inequality of variance has been a concern for many. Ogenstad (1998) stated that the assumption of homoscedasticity is typically made for simplicity and mathematical ease. There seems to be a movement of researchers who prefer to use the Welch-Satterthwaite t-test, which assumes inequality of population variances (Moser et al. 1989, Moser and Stevens 1992, Neuhauser 2001).

Under normality but with variance inequality, the actual Type I error rate of the independent sample t-test tends to be (1) near the nominal significance level when the sample sizes are equal or sufficiently large; (2) larger than the nominal level when the smaller sample size is paired with the larger population variance, and (3) smaller than the nominal level when the larger sample size is paired with the larger population variance

(Ramsey, 1980; Scheffe, 1956). Algina et al. (1994) estimated the Type I error rates for three different tests comparing means from two independent samples, including the independent samples t-test and Welch-Satterthwaite approximate t-test. Type I error rates were estimated using skewed distributions (lognormal, exponential, and beta) for various total sample sizes with equal and unequal variances, and equal and unequal sample sizes. Their results indicated that Welch-Satterthwaite t-test tended to control the Type I error rate the same as the independent sample t-test if not better. Moser et al. (1989) showed similar results for Student's t-test based on the additional assumption of normality. Zimmerman and Zumbo (1993) showed that the Welch-Satterthwaite t-test performed on ranked data performed just as well in controlling Type I error as the Wilcoxon Rank Sum test when variance are equal and considerably better when the variances were unequal.

One main concern has been whether to conduct a preliminary variance test to see which t-test should be used. Many researchers feel that performing a statistical test on the basis of the outcome of another test is not a good technique. The acceptance or rejection of a null hypothesis doesn't establish scientific knowledge with the same degree of conviction (Neuhauser 2002). Failure to reject the null hypothesis does not prove it true. Fisher (1935) stated "the null hypothesis is never proved or established, but is possibly disproved in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis." Conclusions based on nonsignificant results are often not valid (Neuhauser 2002) Therefore this two-stage procedure does not seem optimal since they are based on nonsignificant results. Zimmerman (1996, 2004) showed that such preliminary test for homogeneity of variance affects the error rates of the test substantially leading the test to be invalid. The primary problem is that the test statistics distribution of the test of interest is conditional on the outcome of the preliminary test (Wells and Hintze, 2007). Thus, if an error occurs in the preliminary test then the significance level of the second stage is affected.

Moser et al (1989) and Moser & Stevens (1992) suggest when the variance ratio is unknown, which is most common, that Welch-Satterthwaite t-test should be used. This test provides reasonable type I error rates and powers (Neuhauser, 2002). Moser et al, (1989) showed the main advantage of the Welch-Satterthwaite t-test over Students t-test is that it has acceptable type I error rates when the variances differ. As well, the power of the Welch-Satterthwaite t-test is similar to Student's t-test even if population variances are equal (Moser et al. 1989, Moser and Stevens 1992; Coombs et al. 1996).

There has been a long standing debate regarding the differences in power between non-parametric and parametric tests. It has been shown through simulation and theory that the Wilcoxon Rank Sum test and Students t-test have similar, basically equivalent, power when sampled from normal distributions (Hodges and Lehmann, 1956; Lehmann 1975). Lehmann (1975) stated that the t-test is more powerful compared to the Wilcoxon Rank Sum test when the dealing with normal distributions, however the efficiency loss of the Wilcoxon Rank Sum test is only about 5 percent. However, much of the claim of superiority of Student's t-test was based on normal theory.

A simple way to compare such tests is to use the Pitman efficiency (Pitman, 1948) also referred to as the asymptotic relative efficiency (A.R.E), which is used to compare large sample power for two tests. Hodges and Lehmann (1956) showed that the Pitman efficiency of the Wilcoxon Rank Sum test compared to the t-test will always be greater than 0.864 for any distribution with a finite variance. When normality holds true, the Wilcoxon Rank Sum test has a Pitman efficiency of $\pi / 3 = 0.955$ compared to Students t-test (Lehmann, 1975). However, one criticism made by Bradley (1968) is that A.R.E's apply to very large sample sizes. Therefore, the asymptotic result may be misleading for more realistic situations where one would be dealing with smaller sample sizes.

Since Student's t-test and Wilcoxon Rank Sum test differ based on the assumption of normality it seems obvious that a comparison of the two tests should be done for non-normal distributions. Boneau (1962) conducted a simulation study that compared these two tests when sampling from normal, rectangular and exponential distributions. They found that the t-test was only slightly more powerful than its non-parametric counterpart. And, he implied that Hodges and Lehmann (1956) asymptotic results are not valid for small sample sizes. Toothaker (1972) compared these two tests for the normal, uniform and skewed distributions for small sample sizes (at most five). They found a minor difference in the power of the two tests, which are similar to the results found by Boneau (1962).

Neave and Granger (1968) considered the power difference for a non-normal population formed by the super position of two normal distributions. They found that the Wilcoxon statistic was significantly more powerful compared to the t-test. Blair et al. (1980) compared the relative power for these two tests while sampling from an exponential distribution. They found that the Wilcoxon Rank Sum test had power advantages over the t-test. As well, they commented results found by Boneau (1962) and Toothaker (1972) may be questionable since they used such small sample sizes. Blair et al. (1980) suggested that using more common moderate sample sizes would lead to different results.

Lehmann (1975) compared these two tests for four non-normal distributions (logistic, double exponential, rectangular, and exponential). The values of the Pitman efficiencies that were found are shown in Table 1.1

Table 1.1 - Pitman efficiency of Wilcoxon Rank Sum test to Student's t-test

Distribution	Logistic	Double Exponential	Rectangular	Exponential
Pitman Efficiency	$\pi^2/9 = 1.097$	1.5	1	3

For distributions close to normal, the power of the tests was approximately equal. And, for distributions whose tails are much heavier than that of the normal distribution, it is possible the Wilcoxon Rank Sum test would be more efficient than the t-test.

This debate is what led Blair et al. (1980) to conduct a study comparing the relative power between the Wilcoxon Rank Sum test and Student's t-test. They considered a variety of non-normal distributions, including mixtures of normal and uniform distributions, as well as an assortment of moderate sample sizes. Blair et al. (1980) claimed that (1) the Wilcoxon Rank Sum test statistic had significantly larger power advantage compared to the Student's t-test statistic, where the term power advantage refers to the quantity obtained when the proportion of hypotheses rejected by the less powerful statistic is subtracted from the proportion of rejections by the more powerful statistic with both proportions being calculated at a particular difference of means, (2) the asymptotic relative efficiencies (Pitman efficiencies) were good indicators of the relative power of the two statistics, (3) there were significant differences in power for small samples versus large samples.

1.3 Outline of Dissertation

As one can see based on the literature review, much attention regarding mixtures has been on homogeneity versus heterogeneity. Studies regarding the two-group comparison test involving mixtures have been focused primarily on the normal distribution. Very few studies have investigated the power of test procedures involving mixtures.

In this thesis, we conduct a power analysis comparing four different tests for the two-group comparison problem assuming heterogeneity within each group. We focus our attention on mixtures of Poisson distributions. We considered three different alternatives differing by one additional parameter from the null hypothesis:

Alternative I – The two groups differ in mixing proportions only.

i.e., $\pi_1 \neq \pi_2$, and $0 < \pi_x < 1$ for $x = 1, 2$, $\lambda_{11} = \lambda_{21}, \lambda_{12} = \lambda_{22}$

Alternative II – The two groups differ by the value of second component means.

i.e., $\pi_{12} = \pi_{22} = \pi$, and $0 < \pi < 1$, $\lambda_{11} = \lambda_{21}, \lambda_{12} \neq \lambda_{22}$

Alternative III – The two groups differ by the value of first component means.

i.e., $\pi_{12} = \pi_{22} = \pi$, and $0 < \pi < 1$, $\lambda_{11} \neq \lambda_{21}, \lambda_{12} = \lambda_{22}$

Chapter 2

Two Sample Likelihood Ratio Test in the Presence of Mixture

2.1 The Likelihood Ratio Test

The general model for the two-group comparison in the presence of a mixture is a special case of the general mixture model where the variable X is binary. For the two-group comparison in the presence of a mixture, the null and alternative models are as follows:

$$H_o : f\left(Y; X = 1\right) = f\left(Y; X = 2\right) = \sum_{j=1}^{k} \pi_j \cdot g\left(Y; \underline{\theta}\right) \tag{2.1}$$

$$H_a : \begin{cases} (control) & f\left(Y; X = 1\right) = \sum_{j=1}^{k} \pi_{1j} \cdot g_{1j}\left(Y; \underline{\theta}_{1j}\right) \\ (treatment) & f\left(Y; X = 2\right) = \sum_{j=1}^{k} \pi_{2j} \cdot g_{2j}\left(Y; \underline{\theta}_{2j}\right) \end{cases}$$

such that $\sum_{j=1}^{k} \pi_{xj} = 1$, $0 < \pi_{xj} < 1$, where k represents the number of components in group X for $x = 1, 2$.

More specifically, for the two-component Poisson mixture, $g_{xj}(Y;\underline{\theta}) = g_{xj}(Y;\lambda_{xj})$ represents the Poisson density function where λ_{xj} represents the mean and variance for the jth component density, where $\mathrm{j}=1,2$ and $x=1,2$.

Therefore, the null and alternative models for the two-group comparison in the presence of a two-component Poisson mixture are:

$$H_o: f(Y;X=1) = f(Y;X=2) = (1-\pi)g(Y;\lambda_1) + \pi g(Y;\lambda_2)$$

$$H_a:\begin{cases} (control) & f(Y;X=1) \\ & =(1-\pi_{12})\,g(Y;\lambda_{11}) + \pi_{12} g(Y;\lambda_{12}) \\ (treatment) & f(Y;X=2) \\ & =(1-\pi_{22})\,g(Y;\lambda_{21}) + \pi_{22} g(Y;\lambda_{22}) \end{cases} \tag{2.2}$$

Based on the above model, the likelihood function for an observation from the control group (X=1) is:

$$f(Y;X=1) = (1-\pi_{12})\,g(Y;\lambda_{11}) + \pi_{12} g(Y;\lambda_{12}) \tag{2.3}$$

and the likelihood function for an observation from the treatment group (X=2) is:

$$f(Y;X=2) = (1-\pi_{22})\,g(Y;\lambda_{21}) + \pi_{22} g(Y;\lambda_{22}) \tag{2.4}$$

It follows that the likelihood function under the null hypothesis is:

$$\begin{aligned} L_o &= \prod_{k\in C} f(y_k;\underline{\theta}_1) \cdot \prod_{k\in T} f(y_k;\underline{\theta}_2) \\ &= \prod_{k\in C \text{ or } T} \left[(1-\pi)g(y_k;\lambda_1) + \pi g(y_k;\lambda_2)\right] \end{aligned} \tag{2.5}$$

and the likelihood function under the alternative hypothesis is:

$$\begin{aligned} L_o &= \prod_{k\in C} f\left(y_k; X=1\right)\cdot \prod_{k\in T} f\left(y_k; X=2\right) \\ &= \prod_{k\in C}\left[(1-\pi_{12})g\left(y_k;\lambda_{11}\right)+\pi_{12}g\left(y_k;\lambda_{12}\right)\right] \\ &\quad\times \prod_{k\in T}\left[(1-\pi_{22})g\left(y_k;\lambda_{21}\right)+\pi_{22}g\left(y_k;\lambda_{22}\right)\right] \end{aligned} \tag{2.6}$$

where C denotes the control group (X=1) and T denotes the treatment group (X=2).

The likelihood ratio test statistic, $G^2 = -2\ln\Lambda$ is computed, where $\Lambda = \dfrac{L_0}{L_a}$ is the ratio of the maximum value of the likelihood function under the null hypothesis and the maximum value of the likelihood function under the alternative model being considered. The likelihood ratio test statistic, G^2, asymptotically follows the chi-squared distribution with degrees of freedom equal to the difference in the number of parameters under the null and alternative hypotheses (Cox and Hinkley, 1974).

2.2 Models for Generalized Two-Group Comparison in the Presence of Mixtures

The models considered for the generalized two-group comparison in the presence of mixtures have been summarized in Table and will be discussed in detail in this section.

In Table 2.1, we denote $\pi_x = \pi_{x2}$ for $x = 1,2$. We refer to a given hypothesized model as H_{ijk} where

$$i=\begin{cases}0 & \text{if } \pi_1=\pi_2\\ 1 & \text{if } \pi_1\neq\pi_2\end{cases},\quad j=\begin{cases}0 & \text{if } \lambda_{11}=\lambda_{21}\\ 1 & \text{if } \lambda_{12}\neq\lambda_{22}\end{cases} \text{ and } k=\begin{cases}0 & \text{if } \lambda_{11}\neq\lambda_{21}\\ 1 & \text{if } \lambda_{12}=\lambda_{22}\end{cases}$$

Table 2.1 - Models for two-group comparison in the presence of two component Poisson mixtures

	$\lambda_{11} = \lambda_{21}$ $\lambda_{21} = \lambda_{22}$	$\lambda_{11} = \lambda_{21}$ $\lambda_{21} \neq \lambda_{22}$	$\lambda_{11} \neq \lambda_{21}$ $\lambda_{12} = \lambda_{22}$
$\pi_1 = \pi_2$	H_{000}	H_{010}	H_{001}
$\pi_1 \neq \pi_2$	H_{100}	n/a	n/a

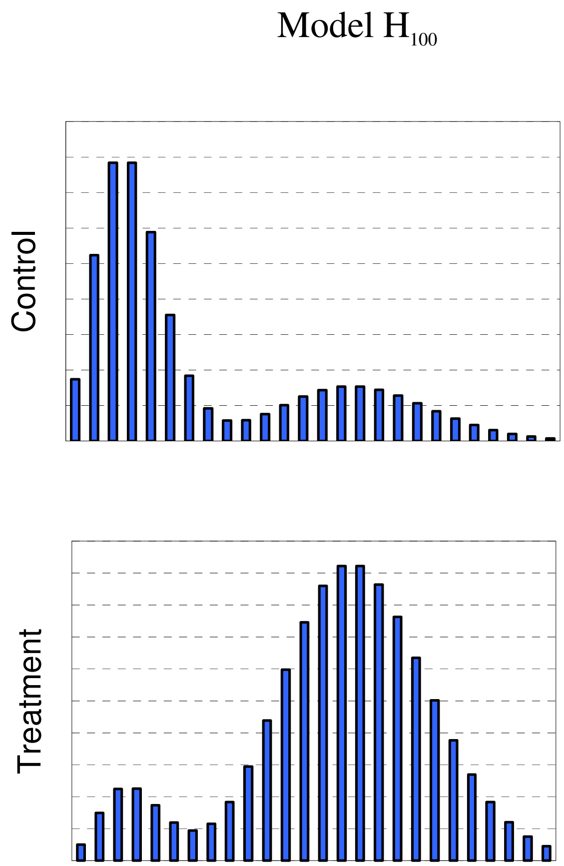
The null hypothesis is denoted by Model H_{000}. The null hypothesis claim is that the mixing proportions and component means for both the control and treatment groups are the same, i.e., $\pi_1 = \pi_2$, $\lambda_{11} = \lambda_{21} = \lambda_{+1}$, $\lambda_{12} = \lambda_{22} = \lambda_{+2}$.

We considered three different alternative hypotheses.

(1) Alternative Model H_{100}:

For this alternative, the claim is that there is a difference in only the mixing proportions for the two groups, i.e., $\pi_1 \neq \pi_2$, $\lambda_{1k} = \lambda_{2k} = \lambda_{+k}$ for $k = 1,2$. By reparameterizing, we have $\alpha = \lambda_{+1}, \alpha + \beta = \lambda_{+2}$ where $\beta > 0$. An example of this alternative model is illustrated in Figure 2.1

Figure 2.1 - Illustration of Alternative Model H_{100} ($\pi_1 \neq \pi_2$)

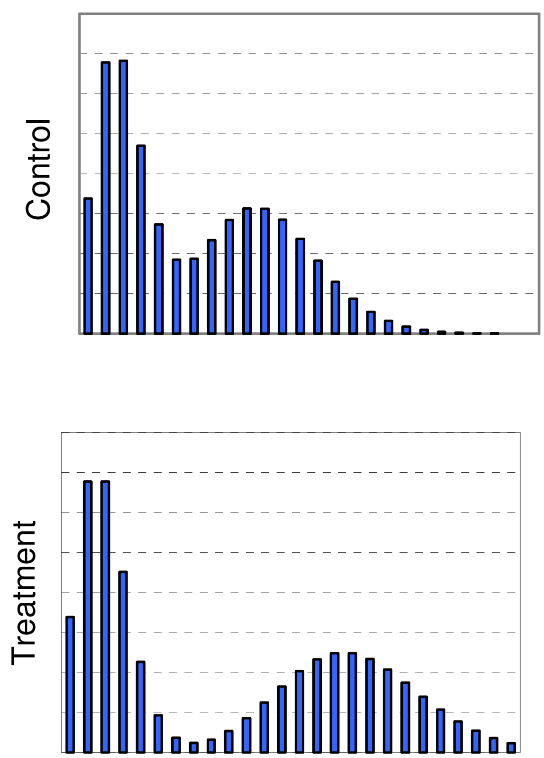


Under this alternative, the only difference between the two groups is their mixing proportions, i.e., $0 < \pi_1 \neq \pi_2 < 1$. Thus, the difference between the number of parameters under the null and alternative hypotheses is one. Therefore, the LRT statistic, $G^2 = -2\ln\Lambda$, asymptotically follows the chi-squared distribution with one degree of freedom.

(2) Alternative Model H_{010}:

For this alternative, the claim is that there is a difference in only the second component means for the two groups, i.e., $\pi_1 = \pi_2$ such that $0 < \pi_i < 1$ for $i = 1, 2$, $\lambda_{11} = \lambda_{21} = \lambda_{+1}$, and $\lambda_{12} \neq \lambda_{22}$. By reparameterizing, we have $\alpha = \lambda_{+1}, \alpha + \beta = \lambda_{12}$, and $\alpha + \beta + \gamma = \lambda_{22}$. An example of this alternative model is illustrated in Figure 2.2 .

Figure 2.2 - Illustration Alternative Model H_{010} ($\lambda_{12} \neq \lambda_{22}$)



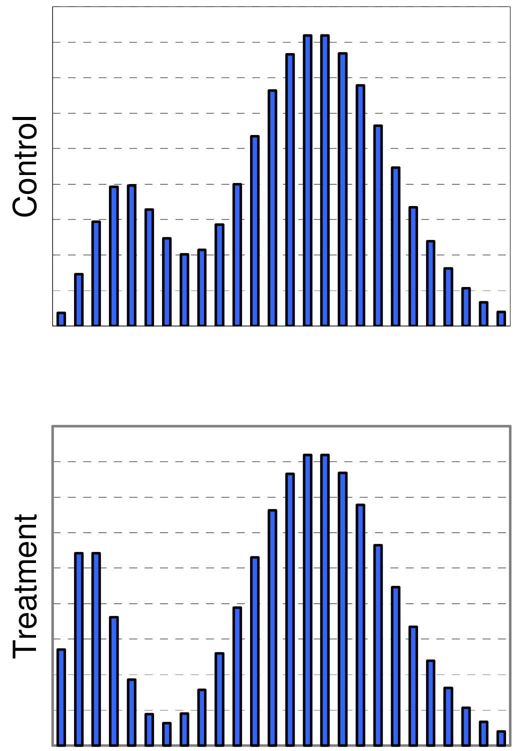
`

Under this alternative, the only difference between the two groups is their second component means, i.e., $\lambda_{12} \neq \lambda_{22}$. Thus the difference between the number of parameters under the null and alternative hypotheses is one. Therefore, the LRT statistics, $G^2 = -2\ln\Lambda$, asymptotically follows the chi-squared distribution with one degree of freedom.

(3) Alternative Model H_{001}:

For this alternative, the claim is that there is a difference in only the first component means for the two groups, i.e., $\pi_1 = \pi_2$ such that $0 < \pi_i < 1$ for $i = 1,2$, $\lambda_{11} \neq \lambda_{21}$, and $\lambda_{12} = \lambda_{22} = \lambda_{+2}$. By reparameterizing, we have $\alpha = \lambda_{12} = \lambda_{22} = \lambda_{+2}$, $\alpha - \beta = \lambda_{11}$, and $\alpha - \beta - \gamma = \lambda_{21}$. An example of this alternative model is illustrated in Figure 2.3 .

Figure 2.3 - Illustration of Alternative Model H_{001} ($\lambda_{11} \neq \lambda_{21}$)



Under this alternative, the only difference between the two groups is their first component means, i.e., $\lambda_{11} \neq \lambda_{21}$. Thus the difference between the number of parameters under the null and alternative hypotheses is one. Therefore, the LRT statistics, $G^2 = -2\ln\Lambda$, asymptotically follows the chi-squared distribution with one degree of freedom.

2.3 Maximum Likelihood Estimation

2.3.1 Expectation Maximization (EM) Algorithm

The EM algorithm (Dempster et al., 1977) is an iterative approach to compute maximum likelihood estimates when the observations can be viewed as incomplete data. The EM algorithm is widely used with mixture models. Typically, the component density function associated with an observation is unknown causing the data to be incomplete.

The application of the EM algorithm to mixture distributions is done in the following manner. Suppose that y consisting of n observations, $y = (y_1, y_2, \ldots, y_n)$, is observed from a k- component Poisson mixture. Thus, y can be viewed as the incomplete data since the components for each observation is unknown. The complete data can be viewed as $x_1 = (y_1, z_1)$, $x_2 = (y_2, z_2)$, …, $x_n = (y_n, z_n)$ where $z_j = (z_{j1}, z_{j2}, ..., z_{jk})$ for j=1,…,n is an unobserved indicator vector whose components are all zero except for the one equal to unity representing the component density function that observation j belongs to (Dempster et al., 1977). The estimation of the parameters can not be done using the complete data likelihood function. A numerical determination of the maximum likelihood estimates is found by maximizing the expectation of the incomplete data log likelihood function.

The incomplete likelihood function for $y=\left(y_1, y_2, \ldots, y_n\right)$ is

$$G\left(Y \mid \underline{\theta}\right)=\prod_{j=1}^{n} g\left(y_j ; \underline{\theta}\right)=\prod_{j=1}^{n} \sum_{i=1}^{k} \pi_i f_i\left(y_j ; \lambda_i\right) \tag{2.7}$$

and the complete likelihood function for $x=\left(x_1, \ldots, x_n\right)$ is

$$F\left(X \mid \underline{\theta}\right)=\prod_{j=1}^{n} \prod_{i=1}^{k} \pi_i^{z_{ji}} f_i\left(y_j ; \lambda_i\right)^{z_{ji}} \tag{2.8}$$

where $\underline{\theta}=\left(\pi_1, \ldots, \pi_k, \lambda_1, \ldots, \lambda_k\right)$ is a vector of unknown parameters and z_{ji} is a indicator variable for the component density function that the jth observation arises from.

The conditional likelihood function of x given y is given by

$$K\left(X \mid Y, \underline{\theta}\right)=\frac{F\left(X \mid \underline{\theta}\right)}{G\left(Y \mid \underline{\theta}\right)}=\prod_{j=1}^{n} \frac{\prod_{i=1}^{k} \pi_i^{z_{ji}} f_i\left(y_j ; \lambda_i\right)^{z_{ji}}}{\sum_{i=1}^{k} \pi_i f_i\left(y_j ; \lambda_i\right)} \tag{2.9}$$

Therefore, the incomplete log likelihood function is

$$\log G\left(Y \mid \underline{\theta}\right)=\log F\left(X \mid \underline{\theta}\right)-\log K\left(X \mid Y, \underline{\theta}\right) \tag{2.10}$$

The conditional expectation of the incomplete data log likelihood function, $\log G\left(Y \mid \underline{\theta}\right)$, is

$$L\left(\underline{\theta}, \underline{\theta}^t\right)=Q\left(\underline{\theta}, \underline{\theta}^t\right)-H\left(\underline{\theta}, \underline{\theta}^t\right) \tag{2.11}$$

where $Q\left(\underline{\theta}, \underline{\theta}^t\right)=E\left[L_{com}\left(Z, Y, \underline{\theta}\right) \mid Y, \underline{\theta}^t\right]$ is the conditional expectation of the complete data log likelihood function is and $H\left(\underline{\theta}, \underline{\theta}^t\right)=E\left[\log K\left(X \mid Y, \underline{\theta}\right) \mid Y, \underline{\theta}^t\right]$.

By applying Jensen's inequality to the above convex function, $H\left(\underline{\theta},\underline{\theta}^{t}\right)$, it can be shown that $H\left(\underline{\theta},\underline{\theta}^{t+1}\right) \leq H\left(\underline{\theta},\underline{\theta}^{t}\right)$ where equality holds true if $K\left(X \mid Y,\underline{\theta}^{t+1}\right) = K\left(X \mid Y,\underline{\theta}^{t}\right)$. Therefore, it follows that in order to maximize the expectation of the incomplete data log likelihood function, $L\left(\underline{\theta},\underline{\theta}^{t}\right)$, one must maximize $Q\left(\underline{\theta},\underline{\theta}^{t}\right)$.

The EM algorithm is an iterative method that involves two steps: the expectation (E) step and the maximization (M) step. In order to start the procedure, one must set initial values for the parameters being estimated. In the E-Step, one computes

$$E\left[z_{ji} \mid Y,\underline{\theta}^{t}\right] = P\left(z_{ji} = 1 \mid Y,\underline{\theta}^{t}\right) = \frac{\pi_i^t f_i\left(y_j;\underline{\theta}_i^t\right)}{\sum_{i=1}^{k} \pi_i^t f_i\left(y_j;\underline{\theta}_i^t\right)}$$ and calculates $L\left(\underline{\theta},\underline{\theta}^{t}\right)$. In the M-Step, one maximizes $Q\left(\underline{\theta},\underline{\theta}^{t}\right)$ with respect to $\underline{\theta}$ in order to find $\underline{\theta}^{t+1}$. This procedure continues until the stopping criteria is meet. The stopping criteria that I choose to use is based on the relative change of the log likelihood function and parameter values in consecutive iterations, $\left|L^{t+1} - L^{t}\right| + \left|\underline{\theta}^{t+1} - \underline{\theta}^{t}\right| < .0001$

2.3.2 MLE based on EM Algorithm

For the two-component Poisson mixture groups, the estimation of λ_{ij} for $i = 1,2$ and $j = 1,2$ can be viewed as an incomplete data problem since the component densities which the individual observations are drawn from is not known (Dempster et al,. 1977). Suppose the observations $y_{ik,}$ $i = 1,2$ and $k = 1,2,\ldots,n_i\,(n_1 + n_2 = N)$ are independent and identically distributed. Then the incomplete data, $y_{ik,}$ can be thought of as complete by considering $(y_{ik,} z_{ik})$, where $y_{ik,}$ is the observed measurement and $z_{ik} = 1$ if $y_{ik,}$ belongs to the second component density in group i.

Let $x_i = 0$ for the control group and $x_i = 1$ for the treatment group. The complete data and observed data likelihood functions for the cases we considered are given, respectively, as follows:

1. Model H_{100} : comparison of two two-component Poisson mixture groups in the presence of differences in only the mixing proportions.

i.e., $\pi_1 \neq \pi_2$ and $0 < \pi_i < 1$, $\lambda_{1j} = \lambda_{2j}$ for $i, j = 1,2$, or equivalently $\pi_1 \neq \pi_2$ and $0 < \pi_i < 1$, $\beta > 0$

The observed (incomplete) data likelihood function is written as

$$
\begin{aligned}
Likelihood(Y,\underline{\theta}) &= \prod_{k\in control} f_1\left(y_k;x_1,\underline{\lambda}_1\right)\cdot \prod_{k\in treatment} f_2\left(y_k;x_2,\underline{\lambda}_2\right) \\
&= \prod_{k\in control}\left[\left(1-\pi_1\right)\cdot f\left(y_k;x_1,\alpha\right)+\pi_1\cdot f\left(y_k;x_1,\alpha+\beta\right)\right] \\
&\times \prod_{k\in treatment}\left[\left(1-\pi_2\right)\cdot f\left(y_k;x_2,\alpha\right)+\pi_2\cdot f\left(y_k;x_2,\alpha+\beta\right)\right]
\end{aligned}
\tag{2.12}
$$

The complete data likelihood function is

$$
\begin{aligned}
&Likelihood\left(Y,Z,\underline{\theta}\right) \\
&=\prod_{i=1}^{2}\prod_{k=1}^{n_i}\left[f\left(y_{ik};x_i,\alpha\right)^{1-z_{ik}}\ f\left(y_{ik};x_i,\alpha+\beta\right)^{z_{ik}}\cdot\pi_i^{z_{ik}}\cdot\left(1-\pi_i\right)^{1-z_{ik}}\right]
\end{aligned}
\tag{2.13}
$$

Therefore, the complete data log likelihood function is

$$
\begin{aligned}
&L_{com}\left(Y,Z,\underline{\theta}\right) \\
&=\log\left[\prod_{i=1}^{2}\prod_{k=1}^{n_i}\left(g\left(y_{ik};x_i,\alpha\right)^{1-z_{ik}}\ g\left(y_{ik};x_i,\alpha+\beta\right)^{z_{ik}}\ \pi_i^{z_{ik}}\left(1-\pi_i\right)^{1-z_{ik}}\right)\right] \\
&=\log\left[\prod_{i=1}^{2}\prod_{k=1}^{n_i}\left(\frac{\alpha^{y_{ik}}e^{-\alpha}}{y_{ik}!}\right)^{1-z_{ik}}\left(\frac{\left(\alpha+\beta\right)^{y_{ik}}\ e^{-(\alpha+\beta)}}{y_{ik}!}\right)^{z_{ik}}\pi_i^{z_{ik}}\left(1-\pi_i\right)^{1-z_{ik}}\right] \\
&=-N\alpha+\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[\begin{array}{l}-\log\left(y_{ik}!\right)+y_{ik}\log\left(\alpha\right)-z_{ik}y_{ik}\log\left(\dfrac{\alpha}{\alpha+\beta}\right)-z_{ik}\beta \\ +z_{ik}\log\left(\pi_i\right)+\left(1-z_{ik}\right)\log\left(1-\pi_i\right)\end{array}\right]
\end{aligned}
\tag{2.14}
$$

The steps of the EM algorithm iteratively update the maximum likelihood estimates

$$\underline{\theta}^{t+1} = \left(\pi_i^{t+1}, \alpha^{t+1}, \beta^{t+1}\right) \tag{2.15}$$

using the current estimates of the parameters, $\underline{\theta}^{t} = \left(\pi_i^{t}, \alpha^{t}, \beta^{t}\right)$ by choosing the values of $\underline{\theta}$ that maximize $Q\left(\underline{\theta}, \underline{\theta}^{t}\right) = E\left[L_{com}\left(Z, Y, \underline{\theta}\right) \mid Y, \underline{\theta}^{t}\right]$. During the E-step, z_{ik} is replaced by the conditional expectation

$$E\left(z_{ik} \mid y_{ik}, \underline{\theta}\right) = \frac{\pi_i g\left(y_{ik}; \alpha+\beta\right)}{\pi_i g\left(y_{ik}; \alpha+\beta\right) + \left(1-\pi_i\right) g\left(y_{ik}; \alpha\right)} \tag{2.16}$$

During the M-step, the derivatives of $Q\left(\underline{\theta}, \underline{\theta}^{t}\right)$ with respect to π_i, α and β are set equal to zero to obtain the new estimates, $\underline{\theta}^{t+1}$, of the parameters.

Let

$$\frac{\partial Q\left(\underline{\theta}, \underline{\theta}^{t}\right)}{\partial \pi_i} = 0. \text{ i.e.,}$$

$$\sum_{k=1}^{n_i}\left[z_{ik}{}^{t} \cdot \frac{1}{\pi_i} + \left(1 - z_{ik}{}^{t}\right) \cdot \frac{-1}{1-\pi_i} \right] = 0$$

$$\Rightarrow \pi_i = \frac{\sum_{k=1}^{n_i} z_{ik}{}^{t}}{n_i}$$

Therefore,

$$\pi_1{}^{t+1} = \frac{\sum_{k=1}^{n_1} z_{1k}{}^{t}}{n_1} \quad \text{and} \quad \pi_2{}^{t+1} = \frac{\sum_{k=1}^{n_2} z_{2k}{}^{t}}{n_2} \tag{2.17}$$

Let

$$\frac{\partial Q\left(\underline{\theta},\underline{\theta}^{t}\right)}{\partial \alpha}=0. \text{ i.e.,}$$

$$\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[-1+y_{ik}\left(\frac{1}{\alpha}\right)-z_{ik}{}^{t}y_{ik}\left(\frac{1}{\alpha}\right)+z_{ik}{}^{t}y_{ik}\left(\frac{1}{\alpha+\beta}\right)\right]=0 \tag{2.18}$$

Let

$$\frac{\partial Q\left(\underline{\theta},\underline{\theta}^{t}\right)}{\partial \beta}=0. \text{ i.e.,}$$

$$\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[z_{ik}{}^{t}y_{ik}\left(\frac{1}{\alpha+\beta}\right)-z_{ik}{}^{t}\right]=0 \tag{2.19}$$

By using equations 2.18 and 2.19, we obtain:

$$\alpha=\frac{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[y_{ik}-y_{ik}z_{ik}{}^{t}\right]}{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[1-z_{ik}{}^{t}\right]} \tag{2.20}$$

and

$$\beta=\frac{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[y_{ik}z_{ik}{}^{t}\right]}{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[z_{ik}{}^{t}\right]}-\frac{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[y_{ik}-y_{ik}z_{ik}{}^{t}\right]}{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[1-z_{ik}{}^{t}\right]} \tag{2.21}$$

2. Model H_{010}: comparison of two two-component Poisson mixture groups in the presence of a difference only in the second component means.

i.e., $\pi_1 = \pi_2 = \pi$ and $0<\pi_i < 1$, $\lambda_{11} = \lambda_{21}$ and $\lambda_{12} \neq \lambda_{22}$ for $i = 1, 2$.
Therefore, $\beta > 0$ and $\gamma > 0$.

The observed data likelihood function is:

$$\begin{aligned} Likelihood\left(Y, \underline{\theta}\right) &= \prod_{k \in Control} f_1\left(y_k; x_1, \underline{\lambda}_1\right) \prod_{k \in Treatment} f_2\left(y_k; x_2, \underline{\lambda}_2\right) \\ &= \prod_{k \in Control} \left[(1-\pi)\, g\left(y_k; x_1, \alpha\right) + \pi g\left(y_k; x_1, \alpha+\beta\right)\right] \\ &\times \prod_{k \in Treatment} \left[(1-\pi)\, g\left(y_k; x_2, \alpha\right) + \pi g\left(y_k; x_2, \alpha+\beta+\gamma\right)\right] \end{aligned} \tag{2.22}$$

The complete data likelihood function for this case is the following:

$$Likelihood\left(Y, Z, \underline{\theta}\right) =$$
$$\prod_{i=1}^{2} \prod_{k=1}^{n_i} \left[g\left(y_{ik}; x_i, \alpha\right)^{1-z_{ik}} \; g\left(y_{ik}; x_i, \alpha+\beta+\gamma x_i\right)^{z_{ik}} \; \pi^{z_{ik}} \left(1-\pi\right)^{1-z_{ik}} \right] \tag{2.23}$$

Thus, it follows that the complete log likelihood function is the following:

$$L_{com}\left(Y,Z,\underline{\theta}\right)$$
$$=\log\left[\prod_{i=1}^{2}\prod_{k=1}^{n_i}\left[g\left(y_{ik};x_i,\alpha\right)^{1-z_{ik}}\ g\left(y_{ik};x_i,\alpha+\beta+\gamma x_i\right)^{z_{ik}}\ \pi^{z_{ik}}\left(1-\pi\right)^{1-z_{ik}}\right]\right] \tag{2.24}$$

$$=\log\left[\begin{array}{l}\prod_{i=1}^{2}\prod_{k=1}^{n_i}\left(\frac{\alpha^{y_{ik}}e^{-\alpha}}{y_{ik}\,!}\right)^{1-z_{ik}}\left(\frac{\left(\alpha+\beta+\gamma x_i\right)^{y_{ik}}\ e^{-\left(\alpha+\beta+\gamma x_i\right)}}{y_{ik}\,!}\right)^{z_{ik}}\\ \qquad\cdot\pi^{z_{ik}}\left(1-\pi\right)^{1-z_{ik}}\end{array}\right]$$
$$=-N\alpha+\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[\begin{array}{l}-\log\left(y_{ik}\,!\right)+\left(1-z_{ik}\right)y_{ik}\log\left(\alpha\right)+z_{ik}\left(y_{ik}\log\left(\alpha+\beta+\gamma x_i\right)-\beta-\gamma x_i\right)\\ +z_{ik}\log\left(\pi\right)+\left(1-z_{ik}\right)\log\left(1-\pi\right)\end{array}\right]$$

The steps of the EM algorithm iteratively update the maximum likelihood estimates

$$\underline{\theta}^{t+1}=\left(\pi^{t+1},\alpha^{t+1},\beta^{t+1},\gamma^{t+1}\right) \tag{2.25}$$

using the current estimates of the parameters, $\underline{\theta}^t=\left(\pi^t,\alpha^t,\beta^t,\gamma^t\right)$ by choosing the values of $\underline{\theta}$ that maximize $Q\left(\underline{\theta},\underline{\theta}^t\right)=E\left[L_{com}\left(Z,Y,\underline{\theta}\right)\mid Y,\underline{\theta}^t\right]$. During the E-step, z_{ik} is replaced by the conditional expectation

$$E\left[z_{ik}\mid y_{ik},\underline{\theta}\right]=\frac{\pi g\left(y_{ik};\alpha+\beta+\gamma x_i\right)}{\pi g\left(y_{ik};\alpha+\beta+\gamma x_i\right)+\left(1-\pi\right)g\left(y_{ik};\alpha\right)} \tag{2.26}$$

During the M-step, the derivatives of $Q\left(\underline{\theta},\underline{\theta}^t\right)$ with respect to π_i,α and β are set equal to zero to obtain the new estimates, $\underline{\theta}^{t+1}$, of the parameters.

Let

$$\frac{\partial Q\left(\underline{\theta},\underline{\theta}^{t}\right)}{\partial \pi}=0. \text{ i.e.,}$$

$$\sum_{\mathrm{i}=1}^{2}\sum_{k=1}^{n_i}\left[z_{ik}{}^{t}\cdot\frac{1}{\pi}+\left(1-z_{ik}{}^{t}\right)\cdot\frac{-1}{1-\pi}\right]=0 \tag{2.27}$$

$$\Rightarrow \pi=\frac{\sum_{\mathrm{i}=1}^{2}\sum_{k=1}^{n_i}\left[z_{ik}{}^{t}\right]}{N}$$

Let

$$\frac{\partial Q\left(\underline{\theta},\underline{\theta}^{t}\right)}{\partial \alpha}=0. \text{ i.e.,}$$

$$-N+\sum_{\mathrm{i}=1}^{2}\sum_{k=1}^{n_i}\left[\left(1-z_{ik}{}^{t}\right)y_{ik}\frac{1}{\alpha}+z_{ik}{}^{t}y_{ik}\frac{1}{\alpha+\beta+\gamma x_i}\right]=0 \tag{2.28}$$

Let

$$\frac{\partial Q\left(\underline{\theta},\underline{\theta}^{t}\right)}{\partial \beta}=0. \text{ i.e.,}$$

$$\sum_{\mathrm{i}=1}^{2}\sum_{k=1}^{n_i}\left[z_{ik}{}^{t}y_{ik}\frac{1}{\alpha+\beta+\gamma x_i}-z_{ik}{}^{t}\right]=0 \tag{2.29}$$

Let

$$\frac{\partial Q\left(\underline{\theta},\underline{\theta}^{t}\right)}{\partial \gamma}=0. \text{ i.e.,}$$

$$\sum_{\mathrm{i}=1}^{2}\sum_{k=1}^{n_i}\left[z_{ik}{}^{t}y_{ik}\frac{x_i}{\alpha+\beta+\gamma x_i}-x_i z_{ik}{}^{t}\right]=0 \tag{2.30}$$

By setting equations (2.28) and (2.29) equal, we can solve for α and obtain the following:

$$\alpha = \frac{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[y_{ik}\left(1 - z_{ik}{}^{t}\right)\right]}{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left(1 - z_{ik}{}^{t}\right)} \tag{2.31}$$

By setting equations (2.29) and (2.30) equal, we can solve for β and obtain the following:

$$\beta = \frac{\sum_{k=1}^{n_1} z_{1k}{}^{t} y_{1k}}{\sum_{k=1}^{n_1} z_{1k}{}^{t}} - \frac{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[y_{ik}\left(1 - z_{ik}{}^{t}\right)\right]}{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left(1 - z_{ik}{}^{t}\right)} \tag{2.32}$$

By using equations (2.30), (2.31) and (2.32), we can solve for γ and obtain the following:

$$\gamma = \frac{\sum_{k=1}^{n_2} z_{2k}{}^{t} y_{2k}}{\sum_{k=1}^{n_2} z_{2k}{}^{t}} - \frac{\sum_{k=1}^{n_1} z_{1k}{}^{t} y_{1k}}{\sum_{k=1}^{n_1} z_{1k}{}^{t}} \tag{2.33}$$

3. Model H_{001}, comparison of two two-component Poisson mixture groups in the presence of a difference only in the first component means.

i.e., $\pi_1 = \pi_2 = \pi$ and $0<\pi_i < 1$, $\lambda_{11} \neq \lambda_{21}$ and $\lambda_{12} = \lambda_{22}$ for $i = 1, 2$.
Therefore, $\beta > 0$ and $\gamma > 0$.

The observed data likelihood function is:

$$\begin{aligned} Likelihood\left(Y,\underline{\theta}\right) &= \prod_{k\in Control} f_1\left(y_k;x_1,\underline{\lambda}_1\right) \prod_{k\in Treatment} f_2\left(y_k;x_2,\underline{\lambda}_2\right) \\ &= \prod_{k\in Control} \left[\left(1-\pi\right) g\left(y_k;x_1,\alpha-\beta\right)+\pi g\left(y_k;x_1,\alpha\right)\right] \\ &\times \prod_{k\in Treatment} \left[\left(1-\pi\right) g\left(y_k;x_2,\alpha-\beta\gamma\right)+\pi g\left(y_k;x_2,\alpha\right)\right] \end{aligned} \tag{2.34}$$

The complete data likelihood function for this case is the following:

$$\begin{aligned} &Likelihood\left(Y,Z,\underline{\theta}\right)= \\ &\prod_{i=1}^{2}\prod_{k=1}^{n_i}\left[g\left(y_{ik};x_i,\alpha-\beta-\gamma x_i\right)^{1-z_{ik}}\ g\left(y_{ik};x_i,\alpha\right)^{z_{ik}}\ \pi^{z_{ik}}\left(1-\pi\right)^{1-z_{ik}}\right] \end{aligned} \tag{2.35}$$

Thus, it follows that the complete log likelihood function is the following:

$$\begin{aligned} &L_{com}\left(Y,Z,\underline{\theta}\right) \\ &=\log\left[\prod_{i=1}^{2}\prod_{k=1}^{n_i}\left[g\left(y_{ik};x_i,\alpha-\beta-\gamma x_i\right)^{1-z_{ik}}\ g\left(y_{ik};x_i,\alpha\right)^{z_{ik}}\ \pi^{z_{ik}}\left(1-\pi\right)^{1-z_{ik}}\right]\right] \\ &=\log\left[\begin{array}{l}\prod_{i=1}^{2}\prod_{k=1}^{n_i}\left(\frac{\left(\alpha-\beta-\gamma x_i\right)^{y_{ik}}\ e^{-\left(\alpha+\beta+\gamma x_i\right)}}{y_{ik}!}\right)^{1-z_{ik}}\left(\frac{\alpha^{y_{ik}}e^{-\alpha}}{y_{ik}!}\right)^{z_{ik}} \\ \qquad\cdot\pi^{z_{ik}}\left(1-\pi\right)^{1-z_{ik}}\end{array}\right] \\ &=-N\alpha+\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[\begin{array}{l}-\log\left(y_{ik}!\right)+\left(1-z_{ik}\right)\left[y_{ik}\log\left(\alpha-\beta-\gamma x_i\right)+\beta+\gamma x_i\right]+z_{ik}y_{ik}\log\left(\alpha\right) \\ +z_{ik}\log\left(\pi\right)+\left(1-z_{ik}\right)\log\left(1-\pi\right)\end{array}\right] \end{aligned} \tag{2.36}$$

The steps of the EM algorithm iteratively update the maximum likelihood estimates

$$\underline{\theta}^{t+1} = \left(\pi^{t+1}, \alpha^{t+1}, \beta^{t+1}, \gamma^{t+1}\right) \tag{2.37}$$

using the current estimates of the parameters, $\underline{\theta}^{t} = \left(\pi^{t}, \alpha^{t}, \beta^{t}, \gamma^{t}\right)$ by choosing the values of $\underline{\theta}$ that maximize $Q\left(\underline{\theta}, \underline{\theta}^{t}\right) = E\left[L_{com}\left(Z, Y, \underline{\theta}\right) \mid Y, \underline{\theta}^{t}\right]$. During the E-step, z_{ik} is replaced by the conditional expectation

$$E\left[z_{ik} \mid y_{ik}, \underline{\theta}\right] = \frac{\pi g\left(y_{ik}; \alpha\right)}{\pi g\left(y_{ik}; \alpha\right) + (1-\pi)\, g\left(y_{ik}; \alpha - \beta - \gamma x_{i}\right)} \tag{2.38}$$

During the M-step, the derivatives of $Q\left(\underline{\theta}, \underline{\theta}^{t}\right)$ with respect to π_{i}, α and β are set equal to zero to obtain the new estimates, $\underline{\theta}^{t+1}$, of the parameters.

Let

$$\frac{\partial Q\left(\underline{\theta}, \underline{\theta}^{t}\right)}{\partial \pi} = 0. \text{ i.e.,}$$

$$\sum_{\mathrm{i}=1}^{2} \sum_{k=1}^{n_{i}} \left[z_{ik}{}^{t} \cdot \frac{1}{\pi} + \left(1 - z_{ik}{}^{t}\right) \cdot \frac{-1}{1-\pi} \right] = 0$$

$$\Rightarrow \pi = \frac{\sum_{\mathrm{i}=1}^{2} \sum_{k=1}^{n_{i}} \left[z_{ik}{}^{t} \right]}{N} \tag{2.39}$$

Let

$$\frac{\partial Q\left(\underline{\theta},\underline{\theta}^{t}\right)}{\partial\alpha}=0. \text{ i.e.,}$$

$$-N+\sum_{\text{i}=1}^{2}\sum_{k=1}^{n_i}\left[\left(1-z_{ik}{}^{t}\right)y_{ik}\,\frac{1}{\alpha-\beta-\gamma x_i}+z_{ik}{}^{t}\,y_{ik}\,\frac{1}{\alpha}\right]=0 \tag{2.40}$$

Let

$$\frac{\partial Q\left(\underline{\theta},\underline{\theta}^{t}\right)}{\partial\beta}=0. \text{ i.e.,}$$

$$\sum_{\text{i}=1}^{2}\sum_{k=1}^{n_i}\left[\left(1-z_{ik}{}^{t}\right)y_{ik}\,\frac{-1}{\alpha-\beta-\gamma x_i}+\left(1-z_{ik}{}^{t}\right)\right]=0 \tag{2.41}$$

Let

$$\frac{\partial Q\left(\underline{\theta},\underline{\theta}^{t}\right)}{\partial\gamma}=0. \text{ i.e.,}$$

$$\sum_{\text{i}=1}^{2}\sum_{k=1}^{n_i}\left[\left(1-z_{ik}{}^{t}\right)y_{ik}\,\frac{-x_i}{\alpha-\beta-\gamma x_i}+x_i\left(1-z_{ik}{}^{t}\right)\right]=0 \tag{2.42}$$

By setting equations (2.41) and (2.42) equal, we can solve for $\alpha-\beta$ and obtain the following:

$$\left(\alpha-\beta\right)=\frac{\sum_{k=1}^{n_1}\left[y_{ik}\left(1-z_{ik}{}^{t}\right)\right]}{\sum_{k=1}^{n_1}\left(1-z_{ik}{}^{t}\right)} \tag{2.43}$$

By plugging equations (2.43) into (2.42), we can solve for γ and obtain the following:

$$\gamma = \frac{\sum_{k=1}^{n_2}\left(1-z_{2k}{}^{t}\right) y_{2k}\left(-1\right)}{\sum_{k=1}^{n_2}\left(1-z_{2k}{}^{t}\right)} - \frac{\sum_{k=1}^{n_1}\left(1-z_{1k}{}^{t}\right) y_{1k}}{\sum_{k=1}^{n_1}\left(1-z_{1k}{}^{t}\right)} \tag{2.44}$$

By using equations (2.40) and (2.42) and (2.43), we can solve for α and get the following:

$$\alpha = \frac{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[z_{ik}{}^{t} y_{ik}\right]}{\sum_{i=1}^{2}\sum_{k=1}^{n_i} z_{ik}{}^{t}} \tag{2.45}$$

By using equations (2.43) and (2.45), we can solve for β and obtain the following:

$$\beta = \frac{\sum_{i=1}^{2}\sum_{k=1}^{n_i}\left[y_{ik} z_{ik}{}^{t}\right]}{\sum_{i=1}^{2}\sum_{k=1}^{n_i} z_{ik}{}^{t}} - \frac{\sum_{k=1}^{n_1}\left(1-z_{1k}{}^{t}\right) y_{1k}}{\sum_{k=1}^{n_1}\left(1-z_{1k}{}^{t}\right)} \tag{2.46}$$

2.3.3 Selection of Starting Values for the EM Algorithm and Calculation of the Likelihood

One of the major considerations that have to be made while using the EM algorithm is the choice of initial values for the unknown parameters. A set of good initial values is important because they can affect the convergence rate and the algorithms ability to locate the global maximum. There are various methods that have been used to choose appropriate initial values.

Laird (1978) suggested a grid search, Leroux (1992) and Woodward et al (1984) suggested clustering techniques, and McLachlan (1988) proposed the use of principal component analysis. Other suggestions have been to use estimates obtained by other estimation techniques, such as the method of moments. Hasselbad (1969) fit a mixture of two Poisson’s using the EM algorithm with the initial values for the mixing proportion and component means set to the moment estimates.

Karlis and Xekalaki (2003b) compared various existing methods for choosing initial values for the EM algorithm as well as modifications of existing techniques through a simulation study for finite normal and Poisson mixtures. They noted that Bohning et al. (1999) suggested the use of well-separated values, which resulted in an increase in the convergence rate of the algorithm. However, Karlis and Xekalaki (2003b) found in their simulation study that for the finite Poisson mixtures, where the variance depends on the mean, finding well separated initial values is not very simplistic.

Many suggestions involve various ways of partitioning the data sets to obtain estimates. Atwood et al. (1992) suggested five different mixing proportion values based on different partitions of the data into groups. Whereas, Bohning (1999) suggested partitioning the data by maximizing the within sum of squares criterion.

McLachlan and Basford (1988, Section 3.2) noted that finding an accurate estimate for the mixing proportion is very important. Fowlkes (1979) recommended using the point of inflection in quantile-quantile (Q-Q) plots to estimate the mixing proportion. When dealing with a two-component normal mixture, Thode et al. (1988) suggested that only an initial estimate for the mixing proportion was needed, since the rest of the parameters could be estimated by partitioning the data set by this value . Thode et al. (1988) used the Engelman-Hartigan (1969) test procedure for clusters to split the data set into two groups, which maximizes the ratio of the between-sample variation to the within-sample variation. They as well suggested using ¼, 3/4, (n-1)/n and 1/n as initial estimates for the mixing proportion. Karlis and Xekalaki (2003b) found that Thode et al. (1988) method was a good contender for three-component normal mixtures and two-component Poisson mixtures compared to the other methods used in their simulation study, since for distributions with high overdispersion it requires fewer iterations. Overall, the consensus is to use various initial values to ensure that the global maximum is obtained while using the EM algorithm.

For each model that we considered, we used 101 different sets of initial values for the parameters. A description of the algorithm used to calculate the MLE's for Model H_{010} is described below:

In Model H_{010}, the null hypothesis is that the both groups being compared are drawn from the same two-component Poisson mixture distribution. The alternative is that the groups are drawn from two-component Poisson mixtures with the same mixing proportion and first component means but unequal second component means. The model is described as:

$$H_o : f(Y;X=1) = f(Y;X=2) = (1-\pi)g(Y;\lambda_1) + \pi g(Y;\lambda_2)$$

$$H_a : \begin{cases} (control) & f(Y;X=1) = (1-\pi)g(Y;\lambda_{+1}) + \pi g(Y;\lambda_{12}) \\ (treatment) & f(Y;X=2) = (1-\pi)g(Y;\lambda_{+1}) + \pi g(Y;\lambda_{22}) \end{cases} \qquad (2.47)$$

Let $y_{x1}, y_{x2}, ..., y_{xn_x}$ be a random sample of size n_x, for x=1,2 drawn from the alternative hypotheses.

For the null hypothesized model, the algorithm used to find the MLE's for the parameters $\underline{\theta} = (\pi, \lambda_{+1}, \lambda_{+2})$ is as follows:

Step 1: Use Thode et al. (1988) method to find an optimal starting value for the mixing proportion.

The following is a description of the Engelman-Hartigan test procedure for clusters (1969):

Given a set of data consisting of n observations, $x_1,....,x_n$, which are in ascending order. Such a set of data can be partitioned into two samples $x_{11}, x_{12},..., x_{1n_1}$ and $x_{21}, x_{22},..., x_{2n_2}$, which maximize the ratio of the between-sample variation to the within-sample variation. There are $n-1$ partitions that must be considered:

Partition	Sample 1	Sample 2
$i=1$	$\{x_1\}$	$\{x_2,...,x_n\}$
$i=2$	$\{x_1,x_2\}$	$\{x_3,...,x_n\}$
$\dots$	$\dots$	$\dots$
$i=(n-1)$	$\{x_1,...,x_{n-1}\}$	$\{x_n\}$

For $i=1,...,(n-1)$, the following must be computed

$$C_i = \frac{B}{W} = \frac{n_1 \cdot n_2 \cdot (\bar{x}_1 - \bar{x}_2)^2}{\left[(n_1-1)s_1^2 + (n_2-1)s_2^2\right](n_1+n_2)} \tag{2.48}$$

where $n_i, \bar{x}_i$ and s_i^2 are the sample size, sample mean and sample variance of the i^{th} sample respectively.

We are interested in finding the value of i for which the maximum value of C_i occurs. Using this value of i, we obtain the following estimates:

$$\hat{\pi} = \frac{i}{n},\ \hat{\lambda}_1 = \bar{x}_1,\ \text{and}\ \hat{\lambda}_2 = \bar{x}_2 \tag{2.49}$$

This procedure is done for both the control sample and treatment sample. Using the values above, we obtain the following initial values for the parameters to be:

$$\hat{\pi}^{(0)} = \frac{\hat{\pi}_C + \hat{\pi}_T}{2},\ \hat{\lambda}_{+1}^{(0)} = \frac{\hat{\lambda}_{1C} + \hat{\lambda}_{1T}}{2},\ \text{and}\ \hat{\lambda}_{+2}^{(0)} = \frac{\hat{\lambda}_{2C} + \hat{\lambda}_{2T}}{2} \tag{2.50}$$

The EM algorithm is run using the program we developed and the log likelihood is calculated.

Step 2: Generate 100 uniform (0, 1) random numbers as initial estimates for the mixing proportion, $\hat{\pi}^{(0)}$. Given a value $\hat{\pi}^{(0)}$, both samples are separated into two parts to obtain estimates of the component means in the mixture.

Let $y_{x1}, y_{x2}, \ldots, y_{xn_x}$ be a random sample of size n_x, for x=1,2. Define n_{xj} for $j = 1, 2$ as

$$n_{12} = \left[n_1 \cdot \hat{\pi}^{(0)} \right], n_{11} = n_1 - n_{12} \qquad (2.51)$$

$$n_{22} = \left[n_2 \cdot \hat{\pi}^{(0)} \right], n_{21} = n_2 - n_{22} \qquad (2.52)$$

where $[\bullet]$ takes the integer part of the number in the bracket (Duan).

The initial estimates for the component means (variances) are given by:

$$\hat{\lambda}_{+1}^{(0)} = \frac{\dfrac{\sum_{i=1}^{n_{11}} y_{1(i)}}{n_{11}} + \dfrac{\sum_{i=1}^{n_{21}} y_{2(i)}}{n_{21}}}{2} \qquad (2.53)$$

$$\hat{\lambda}_{+2}^{(0)} = \frac{\dfrac{\sum_{i=n_{11}+1}^{n_1} y_{1(i)}}{n_{12}} + \dfrac{\sum_{i=n_{21}+1}^{n_2} y_{2(i)}}{n_{22}}}{2} \qquad (2.54)$$

where $y_{x(i)}$ for $i = 1, 2, \ldots, n_x$ and x=1,2 are the order statistics of the random samples. Using these initial values of the parameters, the EM algorithm is applied and the log likelihood is computed.

Step 3: The maximum likelihood of Step 1 and Step 2 is chosen to be the global maximum likelihood; hence the corresponding parameter values are the MLE's.

For the alternative hypothesis, the same steps are followed except we obtain the following initial values of the parameters $\underline{\theta} = (\pi, \lambda_{+1}, \lambda_{12}, \lambda_{22})$:

Step 1: The initial values for the parameters are:

$$\hat{\pi}^{(0)} = \frac{\hat{\pi}_C + \hat{\pi}_T}{2},\ \hat{\lambda}_{+1}{}^{(0)} = \frac{\hat{\lambda}_{1C} + \hat{\lambda}_{1T}}{2},\ \hat{\lambda}_{12}{}^{(0)} = \hat{\lambda}_{2C},\ \text{and}\ \hat{\lambda}_{22}{}^{(0)} = \hat{\lambda}_{2T} \tag{2.55}$$

Step 2: The starting values for the parameters based on the initial estimates of the mixing proportion, $\hat{\pi}^{(0)}$, are:

$$\hat{\lambda}_{+1}{}^{(0)} = \frac{\dfrac{\sum_{i=1}^{n_{11}} y_{1(i)}}{n_{11}} + \dfrac{\sum_{i=1}^{n_{21}} y_{2(i)}}{n_{21}}}{2} \tag{2.56}$$

$$\hat{\lambda}_{12}{}^{(0)} = \frac{\sum_{i=n_{11}+1}^{n_1} y_{1(i)}}{n_{12}} \tag{2.57}$$

and

$$\hat{\lambda}_{22}{}^{(0)} = \frac{\sum_{i=n_{21}+1}^{n_2} y_{2(i)}}{n_{22}} \tag{2.58}$$

Using the log likelihood function values obtained under the null and alternative hypotheses, the LRT statistic is calculated. The power of the Likelihood Ratio Test is measured using the asymptotic chi-squared with one-degree of freedom critical value, 3.84. We then investigate the null distribution of $G^2 = -2\ln\Lambda$ and the power of the LRT is measured using the empirical 95th percentile of G^2.

Chapter 3

Statistical Tests considered in Power Analysis of Two-Component Poisson Mixtures

For each simulated data set, the Likelihood Ratio Test (LRT) statistic, the Welch-Satterthwaite t-test, Wilcoxon Rank Sum test statistic and Adjusted Wilcoxon Rank Sum test statistic were calculated. The LRT statistic is calculated by using the maximum likelihood estimates based on the EM algorithm described earlier in Chapter 2.

A description of the various tests we considered is given in this chapter.

3.1 Welch-Satterthwaite t Statistic

The Welch-Satterthwaite t-test (Welch, 1938 and Satterthwaite 1946) is a test which compares means for independent samples drawn from populations with unequal variances. The Welch-Satterthwaite test assumes the random samples are drawn from two normally distributed populations.

The Welch-Satterthwaite t-test statistic is calculated as follows:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \tag{3.1}$$

where n_i denotes the sample size in group i for $i = 1,2$. $\bar{Y}_i$ denotes the average of the observations in group i for $i = 1,2$; and s_i^2 denotes the sample variance in group i for $i = 1,2$, i.e.,

$$s_i^2 = \frac{\sum_{i=1}^{n_i} \left(Y_{ij} - \bar{Y}_i\right)^2}{n_i - 1} \tag{3.2}$$

The degrees of freedom, $\hat{v}$, are given by :

$$\hat{v} = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2 - 1}} \tag{3.3}$$

where $i = 1,2$, and $\min(n_1 - 1, n_2 - 1) \le \hat{v} \le (n_1 + n_2 - 2)$. The degrees of freedom are estimated from the data and are not a function of the sample sizes alone. Often, they are fractional and rounded down to the nearest integer.

3.2 Wilcoxon Rank Sum Statistic

The Wilcoxon Rank Sum test (Wilcoxon, 1945) is a non-parametric alternative to the Welch-Satterthwaite t-test. The Wilcoxon Rank Sum test compares two independent samples, $y_{ij}, j = 1,2,\ldots,n_i,$ and $i = 1,2$ to see whether they are drawn from identical distributions. The test is based on the observed values combined into a single ordered sample in ascending order for which the means of the ranks for each sample are compared. If the two distributions were to be identical, we would expect to find the same average of ranks for each group.

The procedure for conducting a Wilcoxon Rank Sum test is as follows:

1. Rank the combined sample ($N = n_1 + n_2$) of observations in ascending order while keeping track of which sample each observation was drawn from. Assign ranks to each observation without regard to which sample each value came from. A rank of 1 is assigned to the smallest observation, 2 to the second smallest, and so on. If multiple observations are identical (tied), a rank equivalent to the average of their ranks if no ties occurred is assigned.

2. Sum the ranks for each group separately denoting them by R_1 and R_2. Therefore, R_i is the sum of the ranks for $y_{ij}, j = 1,2,\ldots,n_i,$ and $i = 1,2$. Since the ranks range over the integers 1,2,…, N, we have

$$R_1 + R_2 = 1 + 2 + \cdots + N = \frac{N(N+1)}{2}$$

3. Reject H_o if R_1 is larger or smaller than expected under the null hypothesis.

For large n_1 and n_2, the null distribution of R_1 can be well approximated by a normal distribution with a mean and variance given by

$$E(R_1) = \frac{n_1(N+1)}{2} \text{ and } Var(R_1) = \frac{n_1 n_2 (N+1)}{12}$$

It follows a large sample z-test can be based on the following test statistic

$$Z = \frac{R_1 - \left(\frac{n_1(N+1)}{2}\right)}{\sqrt{\frac{n_1 n_2(N+1)}{12}}} \tag{3.4}$$

For a two-sided hypothesis test, the p-value is given by $2 * P(Z \geq |z|)$

3.3. Adjusted Wilcoxon Rank Sum Statistic

One issue that arises when dealing with count data is the duplication of values. The distribution of R_1 is not solely based on each sample size but as well on the number of observations tied at each value. When ties occur frequently and sample sizes are large, the distribution of R_1 can be approximated by the normal distribution with the same mean as described above. However, the variance of R_1 is decreased by the amount

$\frac{n_1 n_2 \sum_{i=1}^{e} \left(d_i^3 - d_i\right)}{12N(N-1)}$ where e denotes the number of distinct values observed and d_j represents the number of ties for value i, $i = 1,...,e$. One other condition in order to use the normal approximation is that $\max_{i=1,...,e}\left(\frac{d_i}{N}\right)$ is bounded away from 1 as $N \to \infty$ (Lehmann, 1975)

Therefore, the Adjusted Wilcoxon Rank Sum Test statistic is:

$$Z = \frac{R_1 - \left(\frac{n_1(N+1)}{2}\right)}{\sqrt{\frac{n_1 n_2 (N+1)}{12} - \frac{n_1 n_2 \sum_{i=1}^{e}\left(d_i^3 - d_i\right)}{12N(N-1)}}} \tag{3.5}$$

3.4 Approximate Power Calculation for the Two-Independent-Sample t-Test

In order to verify that our simulation results were reasonable, we calculated the approximate power of the t-test in order to compare it to the observed t-test power. In order to do so, we made the assumption that each group followed a single normal distribution with unequal variances and means. As a result, the power is approximate because we are actually comparing two-component mixtures of Poisson's.

In our simulations, the sample sizes we considered were large ($n_1 = n_2 = n > 50$). Therefore the critical value for the test statistic, $T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$, with degrees of freedom greater than 50 is approximately 2. As well, since we simulated large sample sizes $(n = 100 \text{ or n=250})$, the t-test statistic can be approximated by the standard normal distribution.

As a result, we used the following test statistic to approximate the power of the unequal variance t-test,

$$Z^* = \frac{\mu_2 - \mu_1}{\sqrt{\frac{Var(Y_1)}{n_1} + \frac{Var(Y_2)}{n_2}}} - Z_{\alpha/2} \tag{3.6}$$

where μ_i and $Var(Y_i)$ represents the mean and variance for group i where i =1,2. The mixture of distributions is taken into account by the formulas for λ_i and $Var(Y_i)$ for i =1,2 which are given in Table 3.1

Table 3.1 - Mean differences between groups in the presence of mixtures

H_{000}	$\mu_2 - \mu_1$	0
	$\mathrm{Var}(\mathrm{Y}_1)$	$[\alpha + \pi\beta] + \pi(1-\pi)[\beta^2]$
	$\mathrm{Var}(\mathrm{Y}_2)$	$[\alpha + \pi\beta] + \pi(1-\pi)[\beta^2]$
H_{100}	$\mu_2 - \mu_1$	$\beta(\pi_2 - \pi_1)$
	$\mathrm{Var}(\mathrm{Y}_1)$	$[\alpha + \beta\pi_1] + \pi_1(1-\pi_1)\beta^2$
	$\mathrm{Var}(\mathrm{Y}_2)$	$[\alpha + \beta\pi_2] + \pi_2(1-\pi_2)\beta^2$
H_{010}	$\mu_2 - \mu_1$	$\pi\gamma$
	$\mathrm{Var}(\mathrm{Y}_1)$	$[\alpha + \pi\beta] + \pi(1-\pi)[\beta^2]$
	$\mathrm{Var}(\mathrm{Y}_2)$	$[\alpha + \pi(\beta+\gamma)] + \pi(1-\pi)(\beta+\gamma)^2$
H_{001}	$\mu_2 - \mu_1$	$-\gamma(1-\pi)$
	$\mathrm{Var}(\mathrm{Y}_1)$	$[(\alpha-\beta) + \pi\beta] + \pi(1-\pi)[\beta^2]$
	$\mathrm{Var}(\mathrm{Y}_2)$	$[(\alpha-\beta-\gamma) + \pi(\beta+\gamma)] + \pi(1-\pi)(\beta+\gamma)^2$

μ_1: Mean of Control group (X=1) $= (1\text{-}\pi_1)\cdot\alpha + \pi_1\cdot(\alpha+\beta)$

μ_2: Mean of Treatment group (X=2) $= (1\text{-}\pi_2)\cdot\alpha + \pi_2\cdot(\alpha+\beta+\gamma)$

$\mathrm{Var}(\mathrm{Y}_1)$: Variance of Control Group (X=1)

$\mathrm{Var}(\mathrm{Y}_2)$: Variance of Treatment Group (X=2)

3.5 Using McNemar's test to compare the power of the LRT to the Adjusted Wilcoxon Rank Sum test and Welch-Satterthwaite t-test.

We are interested in determining whether the Likelihood Ratio Test is more powerful compared to the Adjusted Wilcoxon Rank Sum test and Welch-Satterthwaite t-test. Since all three test statistics was computed for every sample, the power comparison of the LRT to the Adjusted Wilcoxon Rank Sum test, and power comparison of the LRT to the Welch-Satterthwaite t-test., fits a matched pair design. Therefore, we are able to apply McNemar's test to compare the power of the tests. We compared the power of the tests using the 95^{th} and 99.9^{th} percentile values of the chi-squared distribution with one-degree of freedom.

Chapter 4

Empirical Null Distribution for the Likelihood Ratio Test

According to Cox and Hinkley (1974) the likelihood ratio test statistic, G^2, asymptotically follows the chi-squared distribution with degrees of freedom equal to the difference in the number of parameters under the null and alternative hypotheses. Due to the sample sizes that we considered, we wanted to verify that this holds true for the LRT test comparing mixture models. Therefore, through our simulation we found the empirical null distribution for the LRT statistic. We compared the theoretical asymptotic distribution with the empirical null distributions found from simulation. The 95th percentile for the empirical null distribution of the LRT was calculated and corresponding confidence intervals.

4.1 Simulation Results of the Null Distribution for the Likelihood Ratio Test

We considered fifteen different configurations of the parameter values to generate the null distributions. For each configuration, the first component mean was one. For each of the generating null models, one thousand samples were simulated for each of the three alternative models. Table 4.1 shows the generating model parameters used to simulate the data.

Table 4.1 - Null generating model parameter values

β	π
2	0.1 0.3 0.5 0.7 0.9
3	0.1 0.3 0.5 0.7 0.9
4	0.1 0.3 0.5 0.7 0.9

As well, we considered three different sample sizes per group, $n = 100, 500 \text{ and } 1{,}000$ to model both small and large samples. For each configuration the size of the LRT, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test and the Adjusted Wilcoxon Rank Sum Test were estimated through simulation.

4.1.1 Empirical Null Distribution of LRT for Models H_{100}, H_{010}, and H_{001}

For each alternative considered in this dissertation, the difference between the number of parameters in the null distribution versus the alternative distribution is one. Therefore, the critical value for the LRT for each alternative is the theoretical 95th percentile for the chi-square distribution with 1 degree of freedom, 3.84. For each configuration in Table 4.1, we computed the 95% confidence interval for the empirical 95th percentile of the LRT test static value for each alternative model and sample size. The confidence intervals for these 45 models are displayed in Figures 4.1 – 4.3 .

Figure 4.1 - 95% confidence intervals for the empirical 95th percentile of the LRT statistic ($n = 100$)

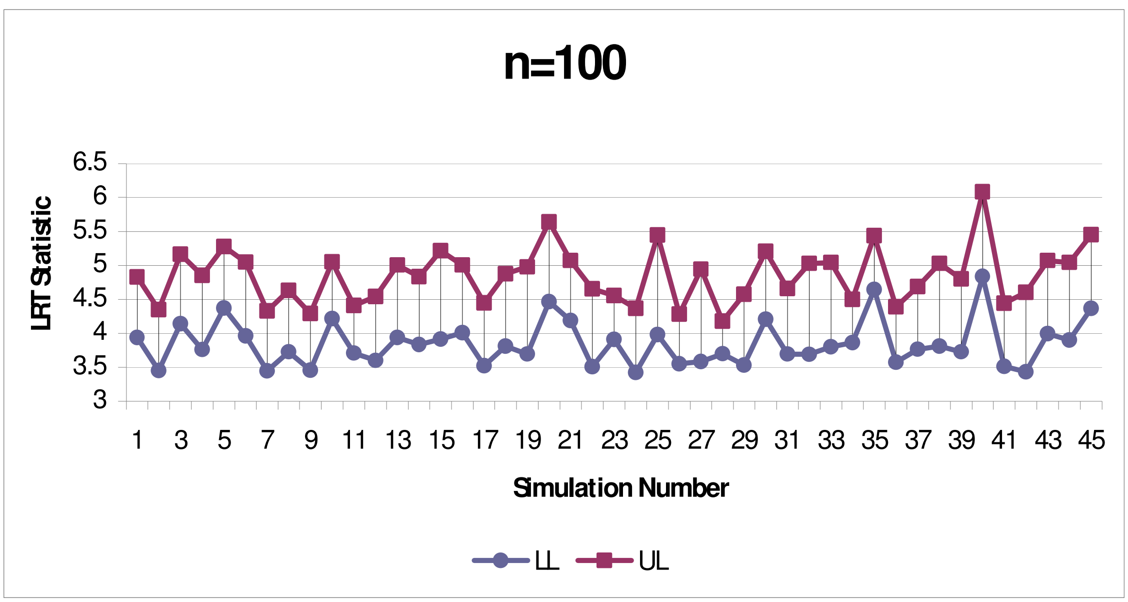


Figure 4.2 - 95% confidence intervals for the empirical 95th percentile of the LRT statistic ($n = 500$)

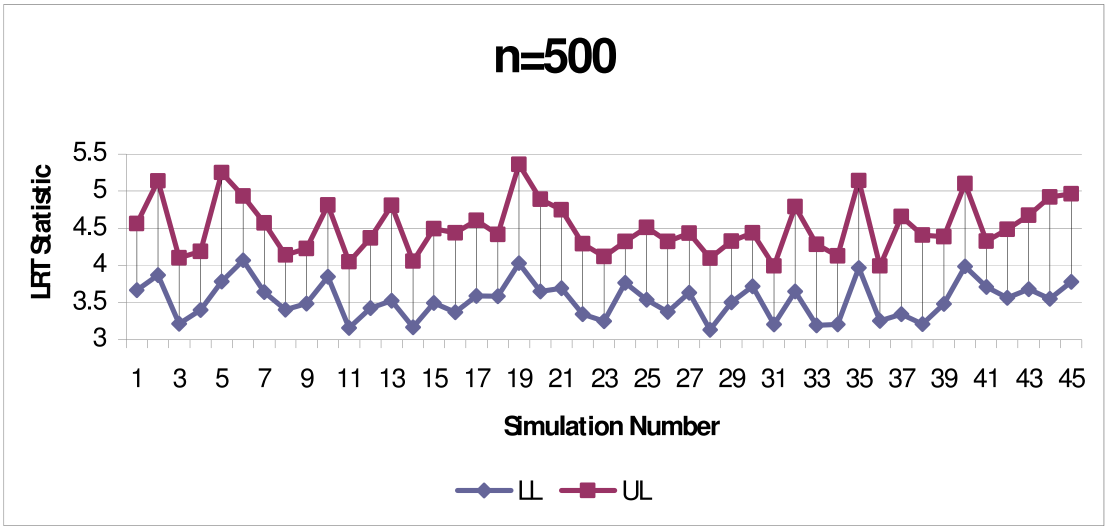
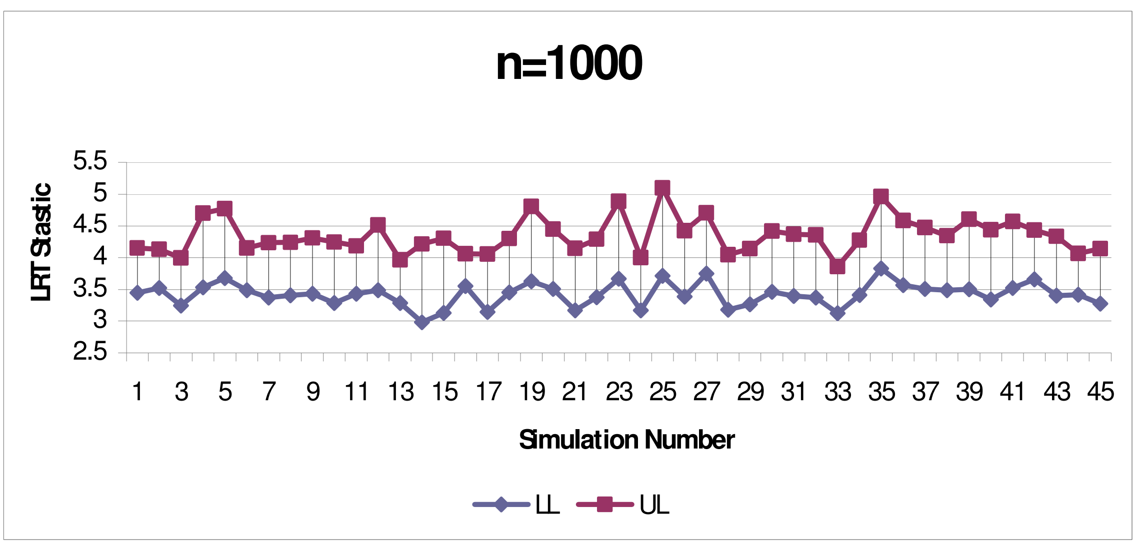


Figure 4.3 - 95% confidence intervals for the empirical 95th percentile of the LRT statistic ($n = 1000$)



As one can see by looking at Figures 4.1- 4.3 , it appears that a majority of the 95% confidence intervals for the 95th percentile overlapped. Therefore, we combined the 45,000 LRT statistics obtained from the simulations to attain a more precise estimate of the empirical 95th percentile of the null distribution. We found the 95th percentile of the null distribution of the LRT statistics for sample size of 100 per group to be 4.26. The 95% confidence interval for the empirical 95th percentile was [4.20, 4.36]. As one can see this value is higher than the asymptotic 95th percentile of the LRT for these models, which is 3.84.

We decided to conduct a similar investigation of the empirical 95th percentile of the null distribution using larger sample sizes per group of $n = 500$ and $n = 1000$. We considered larger sample sizes since the empirical 95th percentile for the null distribution for $n = 100$ per group was significantly higher than the asymptotic 95th percentile, 3.84. We found the 95th percentile of the null distribution of the LRT statistics for sample size of 500 per group to be 3.95. The 95% confidence interval for the empirical 95th percentile was [3.88, 4.03]. We found the 95th percentile of the null distribution of the LRT statistics for sample size of 1000 per group to be 3.82. The 95% confidence interval for the empirical 95th percentile was [3.76, 3.89].

Table 4.2 show the size of the LRT under each generating null model for sample size of one hundred per group ($n = 100$) for each of the three alternative models. For each configuration we ran 1000 replications. The Type I error was estimated using the asymptotic chi-squared distribution with one degree of freedom 95th percentile value, 3.84, for the LRT and the asymptotic 95th percentile standard normal value, 1.96, for the Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank Sum Test. Based on Table 4.2, the size of the LRT seemed to be slightly inflated and the size of the Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank Sum test seemed to be close to the desired value of 0.05.

Table 4.2 – Size of the LRT, Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum test under the generating null hypothesis using asymptotic 95th percentile chi-squared value, 3.84 for the LRT for each alternative model ($n = 100$ per group)

β	π	T-Test	Wilcoxon - Adj	LRT Model H_{100}	LRT Model H_{010}	LRT Model H_{001}
2	0.1	0.047	0.043	0.072	0.074	0.058
	0.3	0.036	0.045	0.052	0.049	0.059
	0.5	0.061	0.060	0.076	0.062	0.062
	0.7	0.048	0.050	0.059	0.060	0.064
	0.9	0.049	0.049	0.084	0.085	0.097
3	0.1	0.048	0.052	0.069	0.084	0.054
	0.3	0.046	0.049	0.053	0.051	0.061
	0.5	0.061	0.059	0.054	0.065	0.062
	0.7	0.044	0.038	0.048	0.049	0.061
	0.9	0.052	0.049	0.081	0.068	0.097
4	0.1	0.036	0.052	0.058	0.048	0.052
	0.3	0.050	0.050	0.058	0.057	0.053
	0.5	0.061	0.066	0.067	0.054	0.069
	0.7	0.059	0.055	0.063	0.055	0.067
	0.9	0.054	0.050	0.066	0.075	0.084

The margin of error of ± 0.03 for each configuration

Table 4.3 show the size of the LRT under each generating null model for sample size of one hundred per group ($n = 100$) for each of the three alternative models. For each configuration, we ran 1000 replicates. The Type I error was estimated using the empirical 95th percentile of the null distribution of 4.26 for the LRT and the asymptotic 95th percentile standard normal value, 1.96, for the other tests considered.

Table 4.3 – Size of the LRT, Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum test under the generating null hypothesis using the empirical 95th percentile of the null distribution of 4.26 for the LRT for each alternative model ($n = 100$ per group)

β	π	T-Test	Wilcoxon Adjusted	LRT Model H_{100}	LRT Model H_{010}	LRT Model H_{001}
2	0.1	0.053	0.044	0.062	0.062	0.049
	0.3	0.053	0.053	0.047	0.052	0.048
	0.5	0.051	0.048	0.051	0.039	0.050
	0.7	0.053	0.050	0.054	0.052	0.055
	0.9	0.047	0.056	0.059	0.054	0.068
3	0.1	0.046	0.040	0.056	0.054	0.033
	0.3	0.043	0.052	0.037	0.047	0.052
	0.5	0.062	0.050	0.043	0.051	0.049
	0.7	0.050	0.046	0.039	0.043	0.052
	0.9	0.055	0.053	0.059	0.063	0.074
4	0.1	0.046	0.054	0.046	0.049	0.043
	0.3	0.058	0.062	0.049	0.051	0.046
	0.5	0.052	0.057	0.045	0.044	0.049
	0.7	0.045	0.043	0.032	0.044	0.051
	0.9	0.058	0.061	0.057	0.059	0.078

The margin of error of ± 0.03 for each configuration.

Figures 4.4- 4.6 illustrate the size of the LRT for the 15 models considered using the asymptotic critical value in Table 4.3 compared to the size of the LRT using the empirical critical value in Table 4.4. As one can see, the size of the LRT using the empirical 95th percentile value seem closer to the desired value of 0.05 compared to when using the asymptotic critical value.

Figure 4.4 – Comparison of Size of LRT for Model H_{100} using asymptotic and empirical critical values

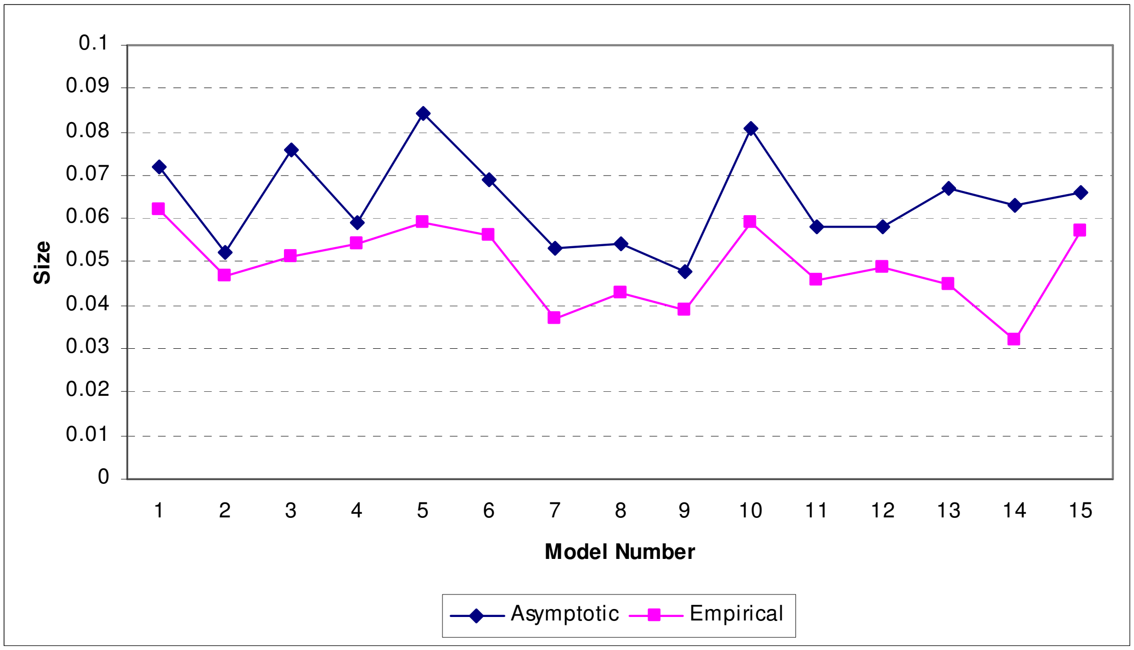


Figure 4.5 – Comparison of Size of LRT for Model H_{010} using asymptotic and empirical critical values

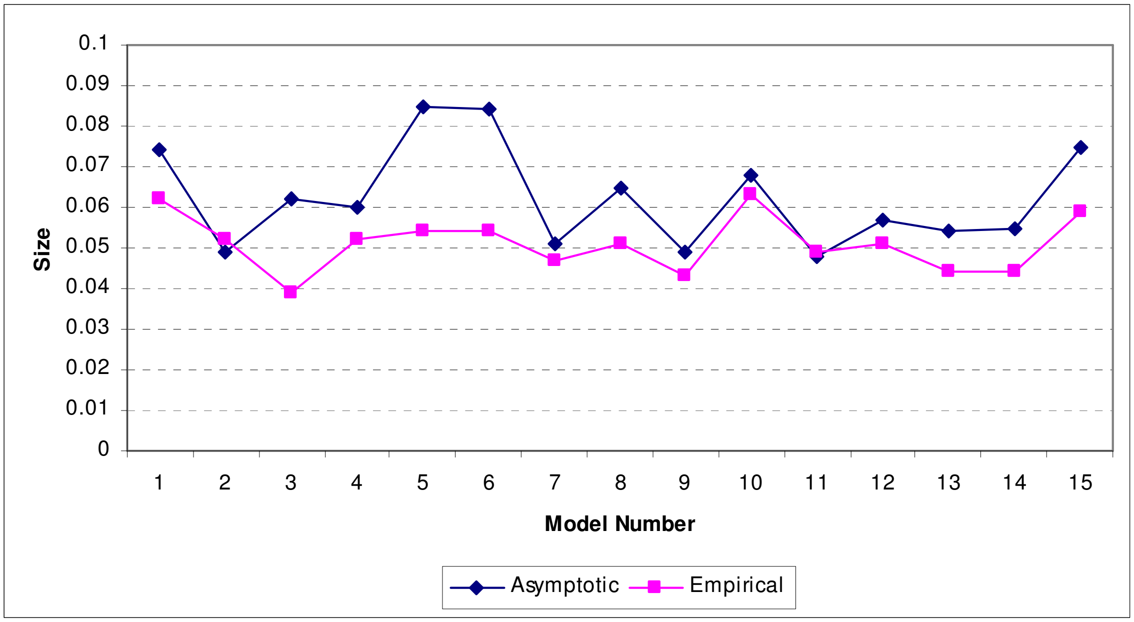
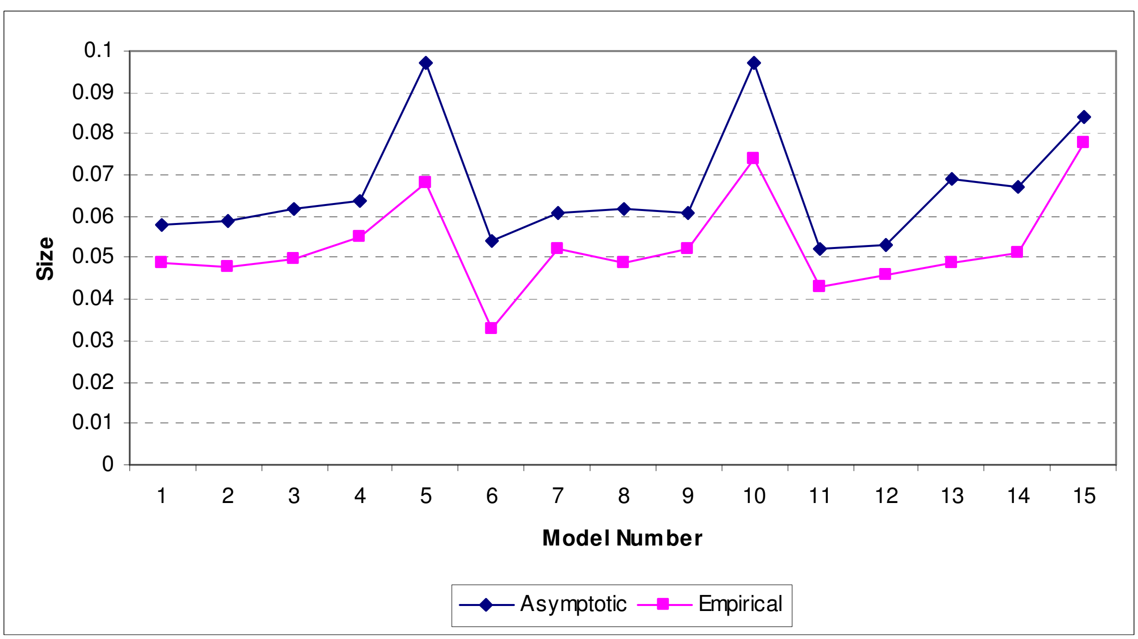


Figure 4.6 – Comparison of Size of LRT for Model H_{001} using asymptotic and empirical critical values



Chapter 5

Power Study of the Likelihood Ratio Test based on the Empirical Null Distribution

5.1 Data Simulation

For our power study, we considered 156 configurations of the parameters. For each configuration, we simulated one thousand samples of size one hundred ($n_1 = n_2 = 100$) using the program that we developed. For each sample the Welch-Satterthwaite t-test statistic, Wilcoxon Rank Sum test statistic, Adjusted Wilcoxon Rank Sum test Statistic, and Likelihood Ratio Test (LRT) statistic were calculated. For each configuration, we then found the power for each of the above statistics through a simulation study using the program that we developed. The power for the Likelihood Ratio Test was based on the 95^{th} percentile of the empirical null distribution, 4.26, which was found in Chapter 4. The power for the three other tests was based on the 95^{th} percentile asymptotic standard normal distribution value, 1.96. As well, we calculated the approximate t-test power as given in equation 3.6 for each configuration.

5.2 Power Study Based on Empirical Null Distribution for the LRT Statistic

5.2.1 Power Study for Model H_{100} using the Empirical Null Distribution for the LRT Statistic

Table 5.1 shows the various configurations of parameters settings used to generate the data. As well, it contains the power of the Welch-Satterthwaite t-test, Wilcoxon Rank Sum Test, Adjusted Wilcoxon Rank Sum Test and Likelihood Ratio Test based on the simulated data for each configuration. For each configuration, we ran 1000 simulations with sample sizes per group of 100. The power of the LRT is calculated based on the empirical 95th percentile critical value, 4.26. The power of the Wilcoxon Rank Sum tests and Welch-Satterthwaite t-test were based on the 95th percentile asymptotic standard normal distribution value, 1.96.

Based on the empirical critical value, it appears that the power of LRT is similar to the other tests considered. In several cases, the power of the Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank sum test were more powerful compared to the LRT. Figure 5.1 illustrates the power results found in Table 5.1. In Figure 5.1, the models are in ascending order based on the power of the LRT.

Table 5.1 - Power of the LRT using the empirical 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{100}; $\alpha = 1$ and n=100 per group.

β	π_1	π_2	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
2	0.1	0.2	0.182	0.194**	0.139	0.157**	0.157
		0.3	0.516	0.541**	0.383	0.413**	0.559
		0.4	0.823	0.851	0.733	0.749**	0.829
	0.2	0.3	0.155	0.142*	0.100	0.121**	0.139
		0.4	0.447	0.459	0.367	0.387**	0.441
		0.5	0.762	0.757	0.715	0.738*	0.728
	0.3	0.4	0.139	0.141*	0.125	0.135*	0.133
		0.5	0.403	0.429*	0.398	0.418**	0.421
		0.6	0.718	0.740*	0.729	0.746*	0.730
	0.4	0.5	0.130	0.134*	0.126	0.131**	0.134
		0.6	0.376	0.375*	0.364	0.384*	0.359
		0.7	0.688	0.669**	0.701	0.729	0.681
3	0.1	0.2	0.262	0.272**	0.168	0.183**	0.288
		0.3	0.695	0.694**	0.541	0.577**	0.693
		0.4	0.943	0.941*	0.867	0.881**	0.948
	0.2	0.3	0.208	0.221*	0.176	0.194**	0.214
		0.4	0.597	0.583**	0.526	0.540**	0.587
		0.5	0.897	0.894**	0.874	0.883**	0.910
	0.3	0.4	0.182	0.169**	0.161	0.175*	0.166
		0.5	0.537	0.556*	0.538	0.550**	0.542
		0.6	0.860	0.855**	0.865	0.870*	0.865
	0.4	0.5	0.167	0.172	0.166	0.172	0.157
		0.6	0.502	0.477**	0.502	0.511*	0.495
		0.7	0.837	0.842**	0.877	0.880*	0.871

$$H_{100} : f\left(x \mid Group\ i\right) = \left(1-\pi_i\right) P\left(\alpha\right) + \pi_i P\left(\alpha+\beta\right) \quad \text{for } x = 1,2$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemars Test
* .05 **.001

Table 5.1 (continued) - Power of the LRT using the empirical 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{100}; $\alpha = 1$ and n=100 per group.

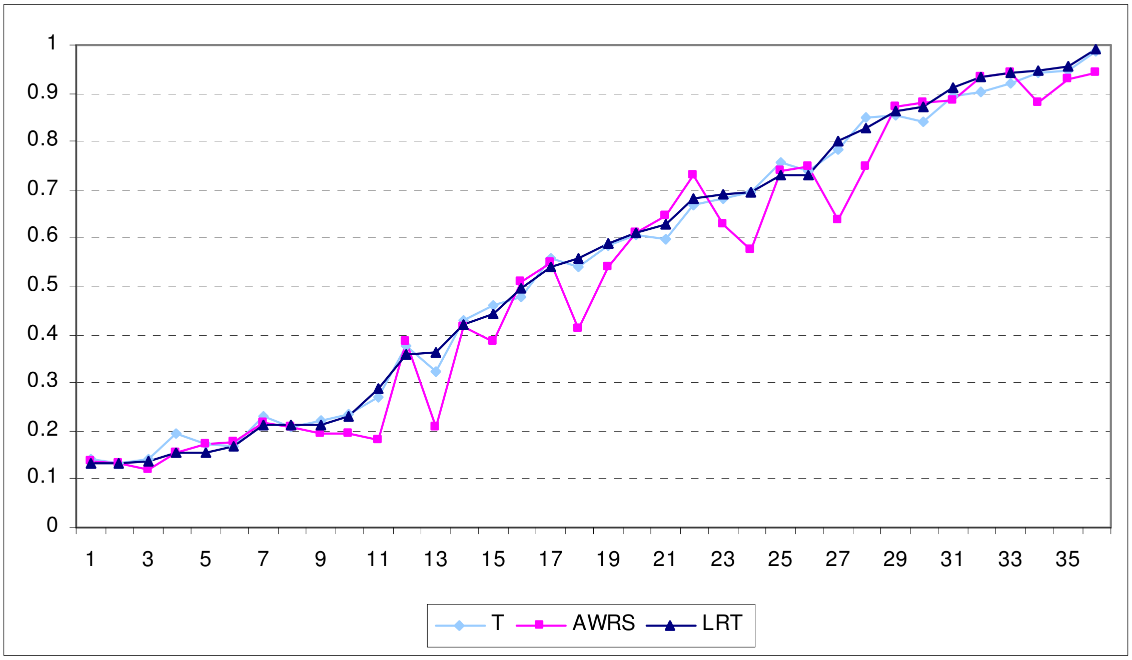
β	π_1	π_2	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
4	0.1	0.2	0.319	0.322**	0.192	0.207**	0.365
		0.3	0.788	0.782**	0.606	0.635**	0.799
		0.4	0.976	0.986	0.937	0.943**	0.991
	0.2	0.3	0.245	0.234*	0.178	0.194**	0.230
		0.4	0.682	0.680**	0.605	0.627**	0.689
		0.5	0.945	0.945*	0.925	0.928**	0.954
	0.3	0.4	0.210	0.230*	0.205	0.219*	0.213
		0.5	0.617	0.605**	0.598	0.612**	0.612
		0.6	0.918	0.922**	0.939	0.941	0.941
	0.4	0.5	0.192	0.209**	0.201	0.208**	0.213
		0.6	0.580	0.596**	0.635	0.644	0.630
		0.7	0.902	0.904**	0.932	0.935*	0.935

$$H_{100} : f\left(x \mid Group\ i\right) = \left(1-\pi_i\right)P\left(\alpha\right) + \pi_i P\left(\alpha+\beta\right) \quad \text{for } x = 1,2$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemars Test
* .05 **.001

Figure 5.1 – Summary of power results for the LRT using the asymptotic chi-squared 95th percentile, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{100}; $\alpha = 1$ and n=100 per group.



5.2.2 Power Study for Model H_{010} using the Empirical Null Distribution for the LRT Statistic

Table 5.2(a) shows the various configurations of parameters settings used to generate the data. As well, it contains the power of the Welch-Satterthwaite t-test, Wilcoxon Rank Sum Test, Adjusted Wilcoxon Rank Sum Test and Likelihood Ratio Test based on the simulated data for each configuration. The power of the LRT is calculated based on the empirical 95th percentile critical value, 4.26. The power of the Wilcoxon Rank Sum tests and Welch-Satterthwaite t-test were based on the 95th percentile asymptotic standard normal distribution value, 1.96.

Based on the simulation, the LRT remains to be significantly more powerful than the Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank Sum Test. Figure 5.2 (a) illustrates the power results from Table 5.2 (a). Given each beta, the models in the figure are in ascending order based on the power of the LRT.

Table 5.2 - Power of the LRT using the empirical 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
2	0.1	0.5	0.046	0.054**	0.034	0.045**	0.088
		1.0	0.075	0.076**	0.032	0.042**	0.151
		1.5	0.113	0.099**	0.044	0.059**	0.255
		2.0	0.157	0.139**	0.050	0.060**	0.343
	0.3	0.5	0.093	0.085**	0.053	0.061**	0.120
		1.0	0.224	0.215**	0.099	0.111**	0.316
		1.5	0.395	0.416**	0.151	0.159**	0.608
		2.0	0.564	0.557**	0.227	0.240**	0.840
	0.5	0.5	0.158	0.153**	0.095	0.100**	0.190
		1.0	0.435	0.446**	0.274	0.284**	0.545
		1.5	0.714	0.731**	0.476	0.491**	0.868
		2.0	0.885	0.892**	0.597	0.609**	0.981
	0.7	0.5	0.253	0.235**	0.181	0.189**	0.258
		1.0	0.684	0.684**	0.541	0.555**	0.751
		1.5	0.932	0.945**	0.867	0.872**	0.969
		2.0	0.991	0.993*	0.972	0.972**	0.999
	0.9	0.5	0.399	0.410**	0.370	0.380**	0.406
		1.0	0.901	0.900*	0.869	0.875**	0.901
		1.5	0.996	0.999	0.998	0.998	0.999
		2.0	1.000	1.000	1.000	1.000	1.000

$$H_{010}: \begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta) \\ f(x \mid X=2) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta+\gamma) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemars Test
* .05 **.001

Table 5.2 (continued) - Power of the LRT using the empirical 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
3	0.1	0.5	0.042	0.057**	0.035	0.045**	0.075
		1.0	0.065	0.065**	0.035	0.041**	0.138
		1.5	0.095	0.100**	0.048	0.055**	0.209
		2.0	0.128	0.132**	0.037	0.049**	0.306
	0.3	0.5	0.075	0.075**	0.045	0.052**	0.112
		1.0	0.164	0.152**	0.060	0.069**	0.289
		1.5	0.286	0.287**	0.096	0.106**	0.534
		2.0	0.423	0.419**	0.132	0.140**	0.752
	0.5	0.5	0.117	0.110**	0.071	0.073**	0.138
		1.0	0.308	0.286**	0.175	0.179**	0.459
		1.5	0.546	0.565**	0.321	0.324**	0.791
		2.0	0.747	0.756**	0.462	0.471**	0.948
	0.7	0.5	0.183	0.181**	0.141	0.149**	0.216
		1.0	0.519	0.504**	0.414	0.420**	0.641
		1.5	0.818	0.836**	0.739	0.743**	0.923
		2.0	0.954	0.944**	0.879	0.881**	0.991
	0.9	0.5	0.302	0.308**	0.286	0.293**	0.323
		1.0	0.791	0.805**	0.774	0.782**	0.815
		1.5	0.979	0.983*	0.981	0.981*	0.990
		2.0	0.999	1.000	0.999	0.999	1.000

$$H_{010}:\begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta) \\ f(x \mid X=2) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta+\gamma) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemars Test
* .05 **.001

Table 5.2 (continued) - Power of the LRT using the empirical 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
4	0.1	0.5	0.039	0.042**	0.028	0.035**	0.068
		1.0	0.058	0.064**	0.043	0.051**	0.094
		1.5	0.081	0.069**	0.050	0.061**	0.163
		2.0	0.108	0.087**	0.036	0.049**	0.243
	0.3	0.5	0.063	0.083	0.052	0.054**	0.093
		1.0	0.128	0.133**	0.061	0.067**	0.260
		1.5	0.217	0.206**	0.074	0.079**	0.491
		2.0	0.324	0.306**	0.073	0.078**	0.713
	0.5	0.5	0.093	0.108**	0.076	0.081**	0.143
		1.0	0.230	0.214**	0.119	0.122**	0.399
		1.5	0.418	0.404**	0.216	0.222**	0.725
		2.0	0.609	0.625**	0.367	0.377**	0.919
	0.7	0.5	0.142	0.160**	0.132	0.134**	0.178
		1.0	0.398	0.401**	0.337	0.340**	0.554
		1.5	0.688	0.699**	0.599	0.603**	0.880
		2.0	0.880	0.876**	0.786	0.790**	0.989
	0.9	0.5	0.240	0.254*	0.251	0.257*	0.260
		1.0	0.679	0.686**	0.671	0.680**	0.733
		1.5	0.941	0.932**	0.921	0.922**	0.964
		2.0	0.995	0.991*	0.993	0.993*	0.997

$$H_{010}:\begin{cases} f\left(x \mid X=1\right)=\left(1-\pi\right)P\left(\alpha\right)+\pi P\left(\alpha+\beta\right) \\ f\left(x \mid X=2\right)=\left(1-\pi\right)P\left(\alpha\right)+\pi P\left(\alpha+\beta+\gamma\right) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemars Test
* .05 **.001

Figure 5.2 – Summary of power results for the LRT using empirical 95th percentile, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=100 per group.

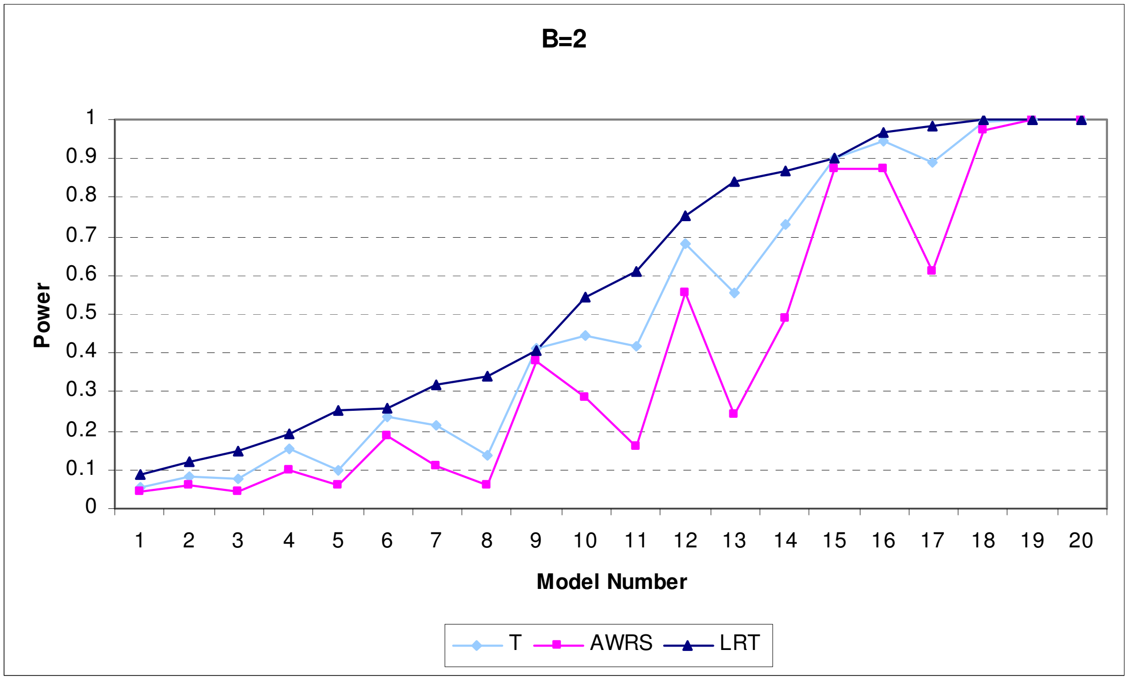


Figure 5.2 (continued) – Summary of power results for the LRT using empirical 95th percentile, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=100 per group.

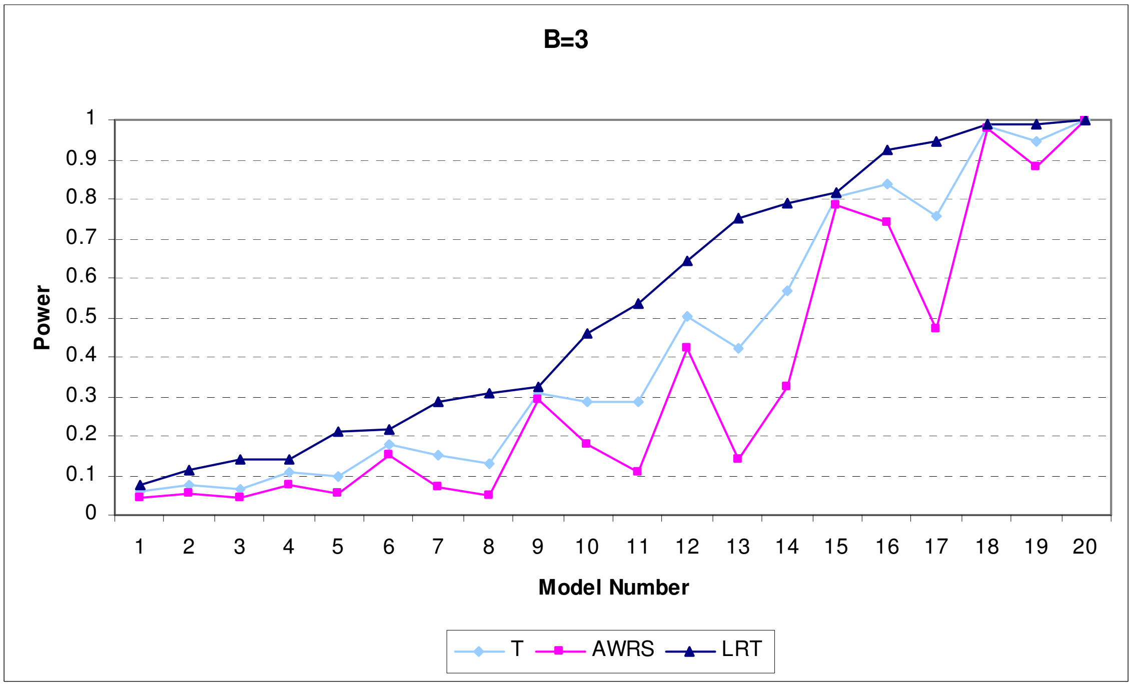
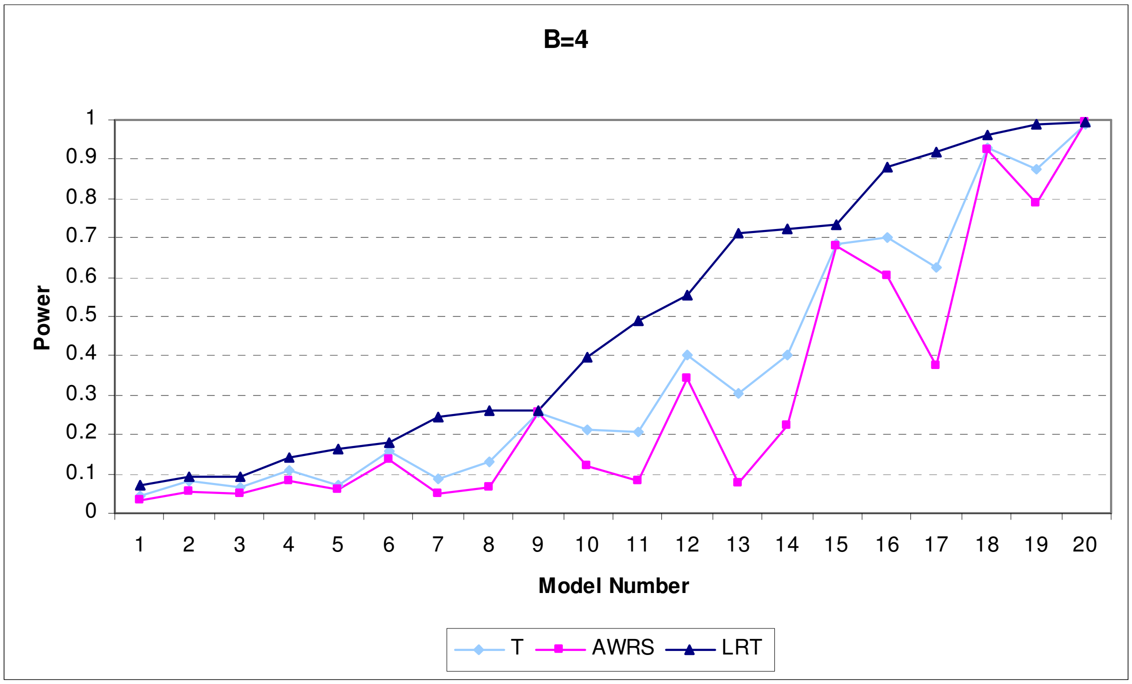


Figure 5.2 (continued) – Summary of power results for the LRT using empirical 95th percentile, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=100 per group.



5.2.3 Power Study for Model H_{001} using the Empirical Null Distribution for the LRT Statistic

Table 5.3 shows the various configurations of parameters settings used to generate the data. As well, it contains the power of the Welch-Satterthwaite t-test, Wilcoxon Rank Sum Test, Adjusted Wilcoxon Rank Sum Test and Likelihood Ratio Test based on the simulated data for each configuration. The power of the LRT is calculated based on the empirical 95th percentile critical value, 4.26. The power of the Wilcoxon Rank Sum tests and Welch-Satterthwaite t-test were based on the 95th percentile asymptotic standard normal distribution value, 1.96.

Based on the empirical critical value, the LRT appeared to be more powerful compared to the Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank Sum test for a majority of the configurations. Figure 5.3 illustrates the power results from Table 5.3. Given each beta, the models in Figure 5.3 are in ascending order based on the power of the LRT.

Table 5.3 - Power of the LRT using the empirical 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon – Adj	LRT
2	0.1	0.5	0.276	0.272*	0.269	0.277*	0.258
		1.0	0.787	0.769**	0.779	0.783**	0.796
		1.5	0.986	0.978**	0.992	0.993*	0.986
		2.0	1.000	1.000	1.000	1.000	1.000
	0.3	0.5	0.161	0.164**	0.174	0.179**	0.200
		1.0	0.486	0.505**	0.549	0.553**	0.589
		1.5	0.814	0.822**	0.883	0.883**	0.914
		2.0	0.963	0.938**	0.976	0.976**	0.993
	0.5	0.5	0.097	0.106**	0.106	0.107**	0.124
		1.0	0.255	0.233**	0.245	0.251**	0.345
		1.5	0.483	0.492**	0.552	0.555**	0.696
		2.0	0.705	0.715**	0.795	0.798**	0.947
	0.7	0.5	0.058	0.072**	0.065	0.068**	0.097
		1.0	0.117	0.123**	0.128	0.130**	0.192
		1.5	0.204	0.238**	0.245	0.248**	0.411
		2.0	0.316	0.315**	0.328	0.329**	0.681
	0.9	0.5	0.034	0.063**	0.059	0.062**	0.073
		1.0	0.045	0.062**	0.064	0.067**	0.083
		1.5	0.058	0.069**	0.069	0.069**	0.118
		2.0	0.074	0.079**	0.070	0.072**	0.166

$$H_{001}:\begin{cases} f\left(x \mid X=1\right)=\left(1-\pi\right)P\left(\alpha-\beta\right)+\pi P\left(\alpha\right) \\ f\left(x \mid X=2\right)=\left(1-\pi\right)P\left(\alpha-\beta-\gamma\right)+\pi P\left(\alpha\right) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemars Test
* .05 **.001

Table 5.3 (continued) - Power of the LRT using the empirical 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon – Adj	LRT
3	0.1	0.5	0.294	0.286^{**}	0.309	0.315^{**}	0.308
		1.0	0.814	0.797^{**}	0.870	0.873^{**}	0.882
		1.5	0.990	0.987^{*}	0.997	0.997	0.997
		2.0	1.000	1.000	1.000	1.000	1.000
	0.3	0.5	0.153	0.165^{**}	0.192	0.197^{**}	0.209
		1.0	0.455	0.453^{**}	0.589	0.595^{**}	0.677
		1.5	0.777	0.767^{**}	0.911	0.913^{**}	0.969
		2.0	0.944	0.935^{**}	0.993	0.993^{*}	1.000
	0.5	0.5	0.091	0.105^{**}	0.113	0.114^{**}	0.159
		1.0	0.230	0.242^{**}	0.285	0.288^{**}	0.458
		1.5	0.432	0.443^{**}	0.551	0.554^{**}	0.834
		2.0	0.642	0.645^{**}	0.775	0.778^{**}	0.982
	0.7	0.5	0.056	0.056^{**}	0.056	0.059^{**}	0.104
		1.0	0.108	0.111^{**}	0.107	0.109^{**}	0.258
		1.5	0.185	0.188^{**}	0.175	0.183^{**}	0.541
		2.0	0.283	0.285^{**}	0.268	0.273^{**}	0.834
	0.9	0.5	0.033	0.050^{**}	0.048	0.050^{**}	0.073
		1.0	0.044	0.063^{**}	0.062	0.062^{**}	0.115
		1.5	0.057	0.064^{**}	0.064	0.068^{**}	0.188
		2.0	0.072	0.082^{**}	0.068	0.070^{**}	0.297

$$H_{001}:\begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha-\beta)+\pi P(\alpha) \\ f(x \mid X=2) = (1-\pi)P(\alpha-\beta-\gamma)+\pi P(\alpha) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemars Test
* .05 **.001

Table 5.3 (continued) - Power of the LRT using the empirical 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
4	0.1	0.5	0.306	0.318**	0.416	0.422**	0.431
		1.0	0.828	0.831**	0.948	0.952	0.953
		1.5	0.992	0.989**	1.000	1.000	1.000
		2.0	1.000	1.000	1.000	1.000	1.000
	0.3	0.5	0.141	0.132**	0.197	0.202**	0.261
		1.0	0.410	0.402**	0.647	0.651**	0.840
		1.5	0.720	0.723**	0.950	0.951**	0.997
		2.0	0.911	0.902**	0.992	0.992*	1.000
	0.5	0.5	0.083	0.068**	0.091	0.097**	0.202
		1.0	0.202	0.208**	0.268	0.272**	0.615
		1.5	0.376	0.370**	0.552	0.556**	0.957
		2.0	0.568	0.591**	0.770	0.772**	1.000
	0.7	0.5	0.053	0.072**	0.069	0.070**	0.123
		1.0	0.099	0.091**	0.094	0.095**	0.373
		1.5	0.165	0.155**	0.149	0.151**	0.729
		2.0	0.249	0.263**	0.243	0.244**	0.962
	0.9	0.5	0.033	0.060**	0.055	0.057**	0.079
		1.0	0.043	0.058**	0.043	0.045**	0.159
		1.5	0.055	0.066**	0.059	0.060**	0.287
		2.0	0.069	0.071**	0.049	0.050**	0.495

$$H_{001}:\begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha-\beta) + \pi P(\alpha) \\ f(x \mid X=2) = (1-\pi)P(\alpha-\beta-\gamma) + \pi P(\alpha) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemars Test
* .05 **.001

Figure 5.3 – Summary of power results for the LRT using the empirical 95th percentile, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=100 per group.

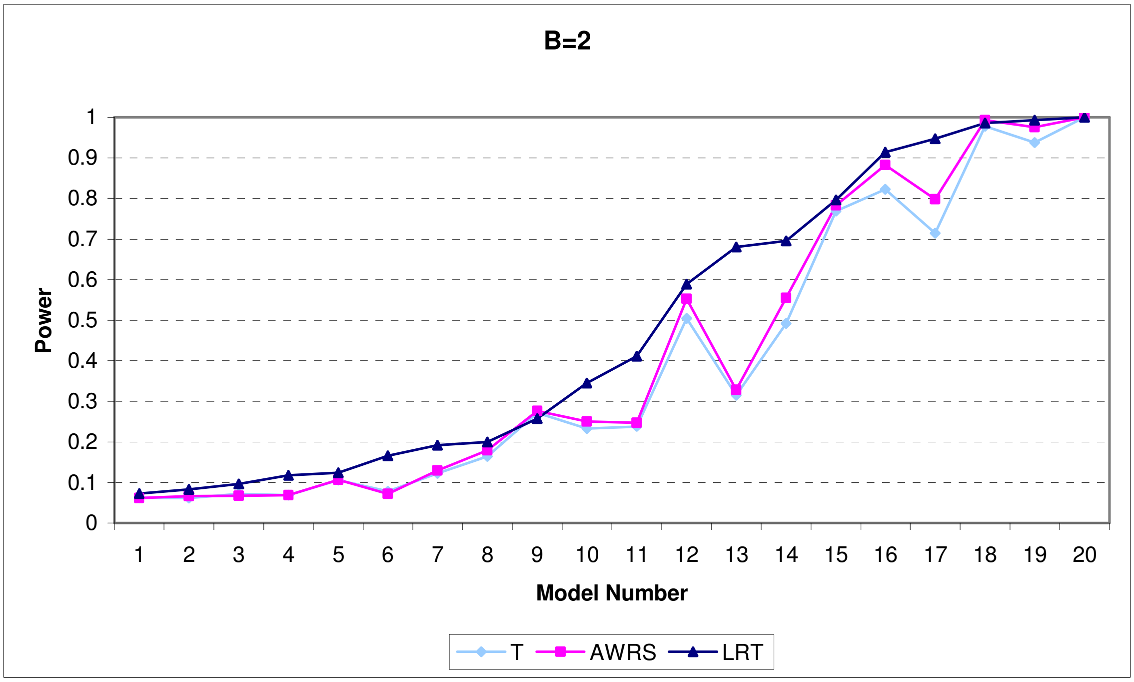


Figure 5.3 (continued) – Summary of power results for the LRT using the empirical 95th percentile, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{001} ; $\alpha = 7$ and n=100 per group.

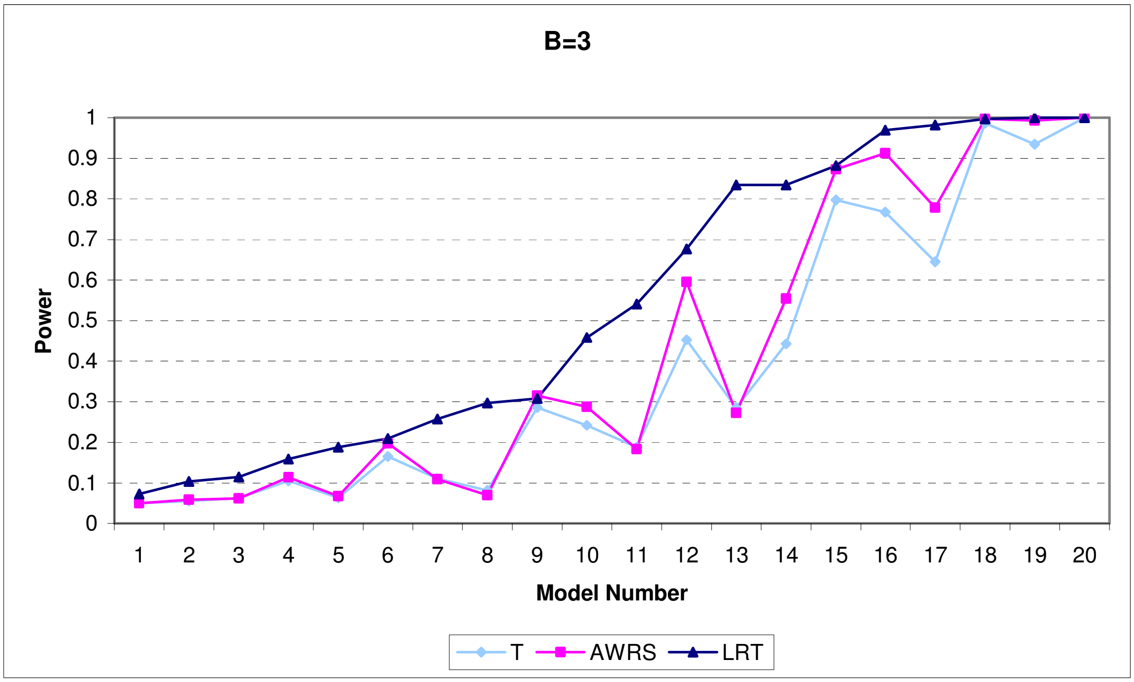
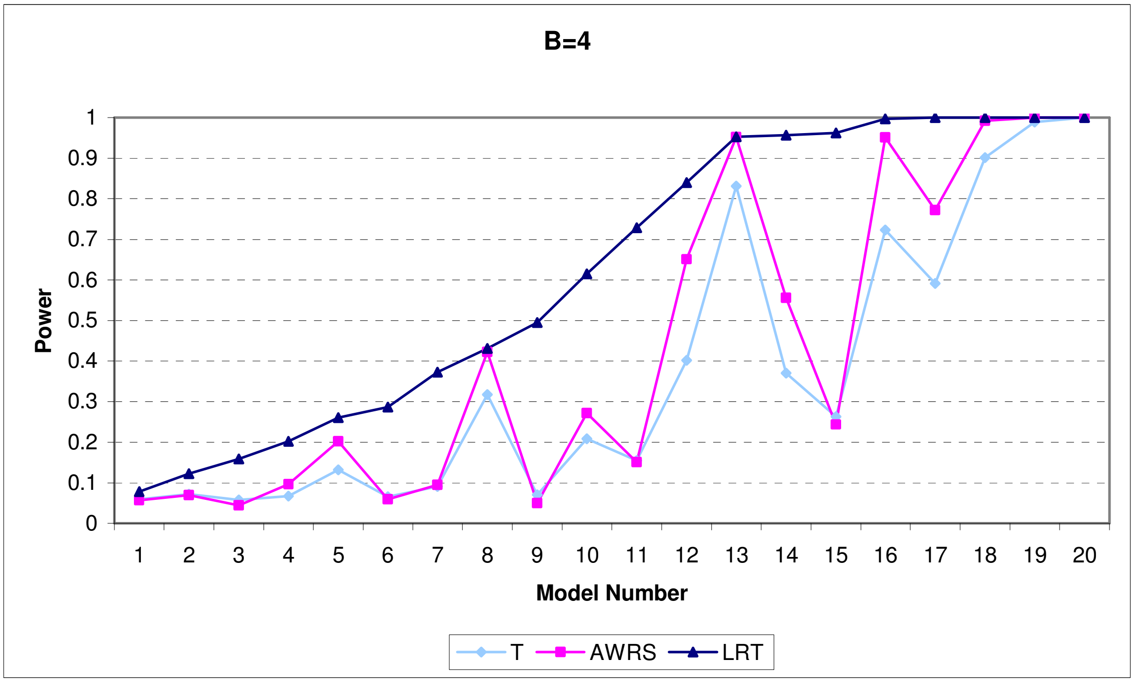


Figure 5.3 (continued) – Summary of power results for the LRT using the empirical 95th percentile, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=100 per group.



Chapter 6

Power Studies of the LRT Based on the Asymptotic Null Distribution

6.1 Data Simulation

For our power study, we considered 156 configurations of the parameters. For each configuration, we simulated one thousand samples of size one hundred ($n_1 = n_2 = 100$) using the program that we developed. For each sample the Welch-Satterthwaite t-test statistic, Wilcoxon Rank Sum test statistic, Adjusted Wilcoxon Rank Sum test Statistic, and Likelihood Ratio Test (LRT) statistic were calculated. For each configuration, we then found the power for each of the above statistics through a simulation study using the program that we developed. As well, we calculated the approximate t-test power as given in equation 3.6 for each configuration. For each alternative model whose approximate t-test power was less than 0.9, we compared the power of the various tests when the sample size per group was increased to 250 ($n_1 = n_2 = 250$) through a similar simulation study.

6.2 Power Comparison Results Based on the Asymptotic Null Distribution

6.2.1 Results of Model H_{100}: Two Component Poisson Mixture with Equal Component Means and Unequal Mixing Proportions

For this model, we considered 36 configurations of the parameter settings. We considered $\pi_1 = 0.1,\ 0.2,\ 0.3, 0.4$ and $\pi_2 - \pi_1 = 0.1,\ 0.2, 0.3$; $\alpha = 1$; $\beta = 2,\ 3,\ 4$ and $n_1 = n_2 = n = 100$. Using the 1000 simulated samples for each configuration of parameter settings, the powers for the various tests considered were estimated.

Table 6.1 (a) shows the various configurations of parameters settings used to generate the data. It contains the power of the Welch-Satterthwaite t-test, Wilcoxon Rank Sum Test, Adjusted Wilcoxon Rank Sum Test and Likelihood Ratio Test based on the simulated data for each configuration. As well, the approximate power of the t-test was calculated using equation (3.6) for each configuration and reported in Table 6.1(a). The power of the LRT is calculated based on the 95th percentile asymptotic critical value of the chi-squared distribution with one degree of freedom, 3.84. The power of the Wilcoxon Rank Sum tests and Welch-Satterthwaite t-test were based on the 95th percentile asymptotic standard normal distribution value, 1.96.

Based on the simulation, we found the power for the LRT to be greater than the Welch-Satterthwaite t-test as well as the Wilcoxon Rank Sum tests for all configurations.

Based on McNemar's test, there was a significant difference in power between the LRT and Welch-Satterthwaite t-test for all configurations where the difference between the mixing proportions was at least 0.2. There was a significant difference between the LRT and Adjusted Wilcoxon Rank Sum test for 28 out of the 36 configurations considered based on McNemar's test. For each beta, the power for the LRT increased as the difference between the mixing proportions increased. And for each configuration of mixing proportions, the power of the LRT increased as the difference between the component means increased.

Figure 6.1 (a) illustrates the power results for the LRT, Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank Sum Test that were reported in Table 6.1 (a) for the various configurations for model H_{100}. The 36 models are ordered by increasing power of the LRT. As one can see by looking at the graph, the power of the LRT is above the power of the other tests considered.

Based on the approximate power calculation of the t-test when the sample size n increased from 100 to 250 per group, there were 17 configurations where the power was less than 0.9. We decided to investigate the power of the tests considered above for these selected models. Table 6.1 (b) shows the generating model parameters used to simulate the data, and the power of the LRT, Welch-Satterthwaite t-test, Adjusted Wilcoxon Rank Sum test, Wilcoxon Rank Sum test as well as the approximate power of the t-test. The results for these selected models are similar to those when the sample size per group was 100. For all 17 settings considered, the power of the LRT is higher than the other tests studied. Based on McNemar's test, there was a significant difference between the LRT and Welch-Satterthwaite t-test for 16 out of the 17 cases and a significant difference between the LRT and Adjusted Wilcoxon Rank Sum test for 13 out of the 17 parameter settings considered. The power results from Table 6.1 (b) are illustrated in Figure 6.1 (b). In Figure 6.1 (b), the models are in ascending order based on the observed LRT power.

Table 6.1 (a) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for alternative H_{100} ; $\alpha = 1$ and n=100 per group.

β	π_1	π_2	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
		0.2	0.182	0.185**	0.124	0.149**	0.234
	0.1	0.3	0.516	0.552**	0.414	0.439**	0.603
		0.4	0.823	0.852*	0.737	0.755**	0.867
		0.3	0.155	0.161*	0.122	0.139**	0.184
	0.2	0.4	0.447	0.473**	0.408	0.431**	0.504
2		0.5	0.762	0.772*	0.734	0.749**	0.790
		0.4	0.139	0.156*	0.147	0.155*	0.175
	0.3	0.5	0.403	0.398**	0.393	0.412*	0.432
		0.6	0.718	0.733*	0.738	0.749	0.758
		0.5	0.130	0.139	0.132	0.142	0.153
	0.4	0.6	0.376	0.421*	0.405	0.416**	0.446
		0.7	0.688	0.695**	0.723	0.737	0.740
		0.2	0.262	0.263*	0.175	0.193**	0.292
	0.1	0.3	0.695	0.742**	0.571	0.592**	0.775
		0.4	0.943	0.946**	0.884	0.892**	0.963
		0.3	0.208	0.196*	0.165	0.179**	0.221
	0.2	0.4	0.597	0.622*	0.558	0.578**	0.649
3		0.5	0.897	0.889**	0.882	0.887**	0.937
		0.4	0.182	0.177	0.174	0.182	0.186
	0.3	0.5	0.537	0.518**	0.524	0.539*	0.566
		0.6	0.860	0.870**	0.892	0.897	0.899
		0.5	0.167	0.170*	0.182	0.186	0.197
	0.4	0.6	0.502	0.522**	0.541	0.561*	0.580
		0.7	0.837	0.847**	0.888	0.891	0.904

$$H_{100} : f\left(x \mid Group\ i\right) = \left(1 - \pi_i\right) P\left(\alpha\right) + \pi_i P\left(\alpha + \beta\right) \quad \text{for } x = 1, 2$$

The margin of error of ± 0.03 for each configuration.

Significantly lower power compared to LRT using McNemar's Test
* .05 **.001

Table 6.1 (a) (continued) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for alternative H_{100} ; $\alpha = 1$ and n=100 per group.

β	π_1	π_2	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
4	0.1	0.2	0.319	0.334**	0.189	0.206**	0.389
		0.3	0.788	0.802**	0.601	0.627**	0.842
		0.4	0.976	0.981*	0.924	0.936**	0.991
	0.2	0.3	0.245	0.242*	0.177	0.192**	0.268
		0.4	0.682	0.688*	0.584	0.607**	0.712
		0.5	0.945	0.952*	0.928	0.933**	0.965
	0.3	0.4	0.210	0.227*	0.214	0.225*	0.246
		0.5	0.617	0.620**	0.626	0.643**	0.706
		0.6	0.918	0.919**	0.932	0.936*	0.951
	0.4	0.5	0.192	0.221	0.207	0.219*	0.237
		0.6	0.580	0.603**	0.620	0.628**	0.665
		0.7	0.902	0.902**	0.936	0.938	0.947

$$H_{100} : f\left(x \mid Group\ i\right) = \left(1-\pi_i\right) P\left(\alpha\right) + \pi_i P\left(\alpha+\beta\right) \quad \text{for } x = 1,2$$

The margin of error of ± 0.03 for each configuration.

Significantly lower power compared to LRT using McNemar's Test
* .05 **.001

Figure 6.1 (a) – Summary of power results for the LRT using the asymptotic chi-squared 95th percentile ordered by ascending power, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{100}; $\alpha = 1$ and n=100 per group.

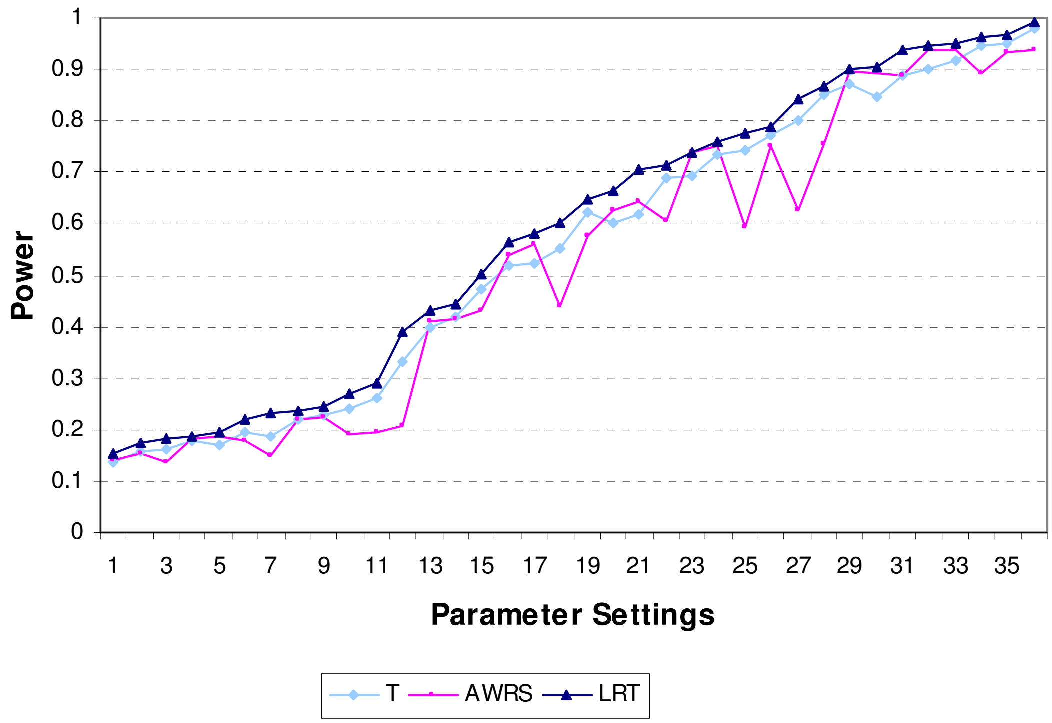


Table 6.1 (b) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{100}; $\alpha = 1$ and n=250 per group.

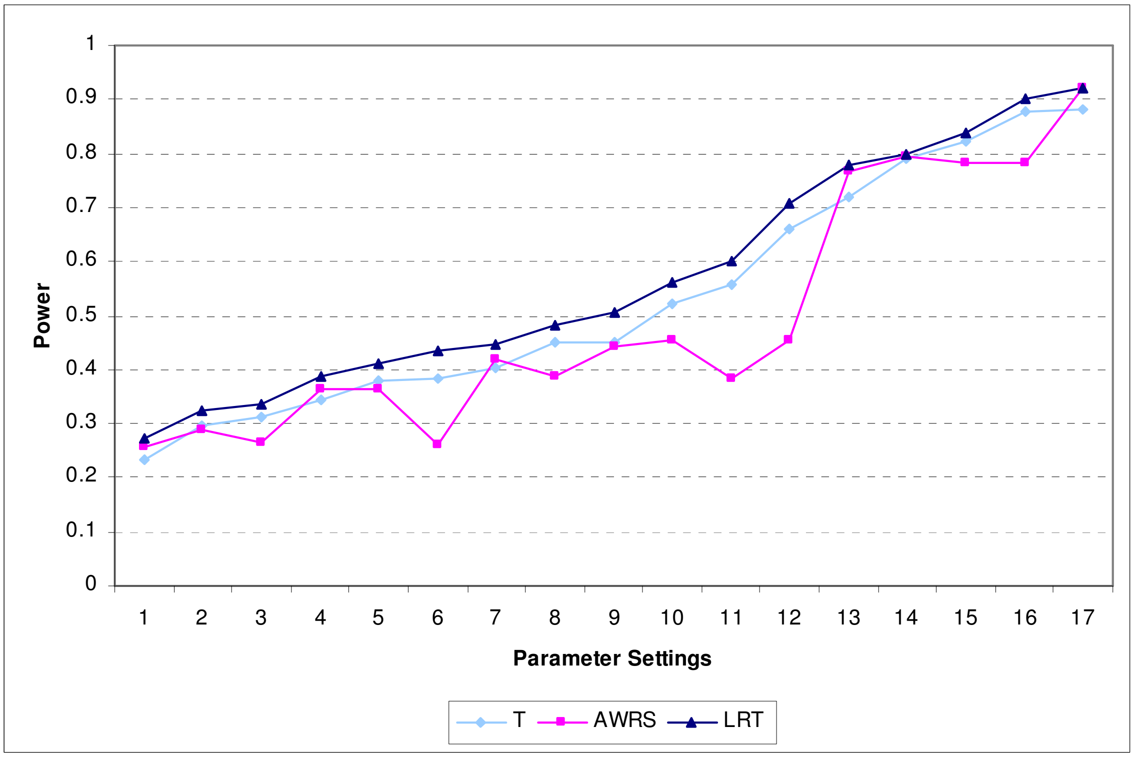
β	π_1	π_2	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
2	0.1	0.2	0.385	0.385**	0.244	0.259**	0.435
		0.3	0.885	0.879*	0.755	0.783**	0.900
	0.2	0.3	0.321	0.314*	0.242	0.263**	0.337
		0.4	0.823	0.824*	0.752	0.781**	0.839
	0.3	0.4	0.283	0.297*	0.278	0.290**	0.323
		0.5	0.774	0.789	0.779	0.794	0.798
	0.4	0.5	0.260	0.235**	0.246	0.258	0.271
		0.6	0.739	0.720**	0.760	0.768	0.779
3	0.1	0.2	0.552	0.558**	0.361	0.385**	0.601
	0.2	0.3	0.442	0.450*	0.363	0.387**	0.481
	0.3	0.4	0.383	0.379*	0.350	0.364**	0.412
	0.4	0.5	0.349	0.345**	0.350	0.363**	0.389
		0.6	0.874	0.883**	0.916	0.919	0.920
4	0.1	0.2	0.654	0.662**	0.421	0.454**	0.707
	0.2	0.3	0.519	0.523**	0.432	0.453**	0.561
	0.3	0.4	0.447	0.451**	0.425	0.442**	0.507
	0.4	0.5	0.407	0.402**	0.406	0.420*	0.448

$$H_{100} : f\left(x \mid Group\ i\right) = \left(1-\pi_i\right) P\left(\alpha\right) + \pi_i P\left(\alpha+\beta\right) \quad \text{for } x = 1,2$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemar's Test
* .05 **.001

Figure 6.1 (b) – Summary of power results for the LRT using the asymptotic chi-squared 95th percentile ordered by ascending power, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{100}; $\alpha = 1$ and n=250 per group.



6.2.2. Results of Model H_{010}: Two Component Poisson Mixture with Equal Mixing Proportions and Unequal Second Component Means

For this model, we considered 60 different configurations of the parameter settings. The configurations that we considered were $\pi_1 = \pi_2 = 0.1,\ 0.3,\ 0.5, 0.7, 0.9$; $\alpha = 1$; $\beta = 2,\ 3, 4$; $\gamma = 0.50,\ 1.00, 1.50$, 2.00 and $n_1 = n_2 = n = 100$. Using the 1000 simulated samples for each configuration of parameter values, the power for the various tests considered were estimated.

Table 6.2 (a) shows the various configurations of parameters settings used to generate the data. It contains the power of the Welch-Satterthwaite t-test, Wilcoxon Rank Sum Test, Adjusted Wilcoxon Rank Sum Test and Likelihood Ratio Test based on the simulated data for each configuration. As well, the approximate power of the t-test was calculated using equation (3.6) for each configuration and reported in Table 6.2 (a). The power of the LRT is calculated based on the 95th percentile asymptotic critical value of the chi-squared distribution with one degree of freedom, 3.84. The power of the Wilcoxon Rank Sum tests and Welch-Satterthwaite t-test were based on the 95th percentile asymptotic standard normal distribution value, 1.96.

The power for the LRT was greater than or equal to the power of the other tests considered for all the configurations. Given a specific beta and mixing proportion, the power of each of the tests increased as the value of gamma increased. As the mixing proportion increased, corresponding to a larger percentage of the groups having different second component means, the power of the tests increased as well. However as beta increased, there was a decrease in the power of all three tests holding the other conditions

constant. The Welch-Satterthwaite t-test test had greater power for each configuration considered compared to the corresponding Adjusted Wilcoxon Rank Sum test. For all differences in component means considered, given the mixing proportion was at most 0.9 there was a significant difference in power between the LRT and Welch-Satterthwaite as well as LRT and Adjusted Wilcoxon Rank Sum test based on McNemar's test.
Figure 6.2 (a) illustrates the power results that were reported in Table 6.2 (a) for the various configurations for model H_{010} .

Based on the approximate power calculation of the t-test when the sample size n increased from 100 to 250 per group, there were 38 configurations where the power was less than 0.9. We decided to investigate the power of the tests considered above for these selected models. Table 6.2 (b) shows the generating model parameters used to simulate the data, and the power of the LRT, Welch-Satterthwaite t-test, Adjusted Wilcoxon Rank Sum test, Wilcoxon Rank Sum test and approximate power of the t-test. The power results from Table 6.2 (b) are illustrated in Figure 6.2 (b). In Figure 6.2 (b), the models are in ascending order based on the observed LRT power. As one can see by looking at the graph, the power of the LRT is more powerful than the Welch-Satterthwaite t-test and the Wilcoxon Rank Sum test when $n = 250$ per group for the selected models we examined.

Table 6.2 (a) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
2	0.1	0.5	0.046	0.065**	0.041	0.048**	0.107
		1.0	0.075	0.077**	0.052	0.061**	0.158
		1.5	0.113	0.117**	0.053	0.067**	0.274
		2.0	0.157	0.149**	0.049	0.060**	0.418
	0.3	0.5	0.093	0.097**	0.046	0.052**	0.135
		1.0	0.224	0.211**	0.091	0.102**	0.366
		1.5	0.395	0.378**	0.154	0.163**	0.642
		2.0	0.564	0.574**	0.197	0.212**	0.853
	0.5	0.5	0.158	0.196*	0.144	0.147**	0.232
		1.0	0.435	0.468**	0.284	0.294**	0.625
		1.5	0.714	0.729**	0.473	0.483**	0.893
		2.0	0.885	0.889**	0.632	0.645**	0.979
	0.7	0.5	0.253	0.232**	0.176	0.184**	0.293
		1.0	0.684	0.676**	0.559	0.568**	0.769
		1.5	0.932	0.938**	0.846	0.851**	0.968
		2.0	0.991	0.989*	0.958	0.962**	0.998
	0.9	0.5	0.399	0.414**	0.368	0.382**	0.454
		1.0	0.901	0.888*	0.860	0.865**	0.905
		1.5	0.996	0.996	0.993	0.993	0.996
		2.0	1.000	1.000	1.000	1.000	1.000

$$H_{010}: \begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta) \\ f(x \mid X=2) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta+\gamma) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly lower in power compared to LRT using McNemar's Test
* .05 **.001

Table 6.2 (a) (continued) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{010};
$\alpha = 1$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
3	0.1	0.5	0.042	0.062*	0.038	0.047**	0.092
		1.0	0.065	0.064**	0.045	0.051**	0.137
		1.5	0.095	0.092**	0.042	0.054**	0.224
		2.0	0.128	0.120**	0.040	0.050**	0.375
	0.3	0.5	0.075	0.075**	0.044	0.046**	0.114
		1.0	0.164	0.158**	0.062	0.071**	0.320
		1.5	0.286	0.277**	0.093	0.099**	0.580
		2.0	0.423	0.416**	0.119	0.133**	0.795
	0.5	0.5	0.117	0.152*	0.100	0.107**	0.179
		1.0	0.308	0.333**	0.185	0.192**	0.515
		1.5	0.546	0.554**	0.317	0.328**	0.822
		2.0	0.747	0.759**	0.437	0.440**	0.957
	0.7	0.5	0.183	0.164**	0.129	0.137**	0.230
		1.0	0.519	0.509**	0.421	0.425**	0.690
		1.5	0.818	0.811**	0.694	0.701**	0.937
		2.0	0.954	0.947**	0.867	0.872**	0.996
	0.9	0.5	0.302	0.304**	0.271	0.279**	0.351
		1.0	0.791	0.774**	0.752	0.754**	0.827
		1.5	0.979	0.973**	0.968	0.969**	0.990
		2.0	0.999	1.000	1.000	1.000	1.000

$$H_{010}: \begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta) \\ f(x \mid X=2) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta+\gamma) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly lower in power compared to LRT using McNemar's Test
* .05 **.001

Table 6.2 (a) (continued) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
4	0.1	0.5	0.039	0.058*	0.039	0.045**	0.084
		1.0	0.058	0.057**	0.039	0.045**	0.133
		1.5	0.081	0.078**	0.037	0.050**	0.212
		2.0	0.108	0.102**	0.039	0.048**	0.329
	0.3	0.5	0.063	0.067**	0.037	0.045**	0.103
		1.0	0.128	0.129**	0.052	0.055**	0.296
		1.5	0.217	0.213**	0.074	0.080**	0.510
		2.0	0.324	0.306**	0.094	0.100**	0.750
	0.5	0.5	0.093	0.117**	0.091	0.094**	0.170
		1.0	0.230	0.241**	0.142	0.147**	0.454
		1.5	0.418	0.433**	0.229	0.235**	0.766
		2.0	0.609	0.601**	0.318	0.328**	0.933
	0.7	0.5	0.142	0.134**	0.110	0.114**	0.207
		1.0	0.398	0.390**	0.339	0.341**	0.618
		1.5	0.688	0.675**	0.587	0.593**	0.899
		2.0	0.880	0.867**	0.766	0.768**	0.991
	0.9	0.5	0.240	0.245**	0.232	0.239**	0.303
		1.0	0.679	0.672**	0.654	0.660**	0.748
		1.5	0.941	0.934**	0.929	0.931**	0.974
		2.0	0.995	0.994	0.992	0.992*	0.999

$$H_{010}:\begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta) \\ f(x \mid X=2) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta+\gamma) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly lower in power compared to LRT using McNemar's Test
* .05 **.001

Figure 6.2 (a) – Summary of power results for the LRT using the chi-squared 95th percentile in ascending order, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=100 per group.

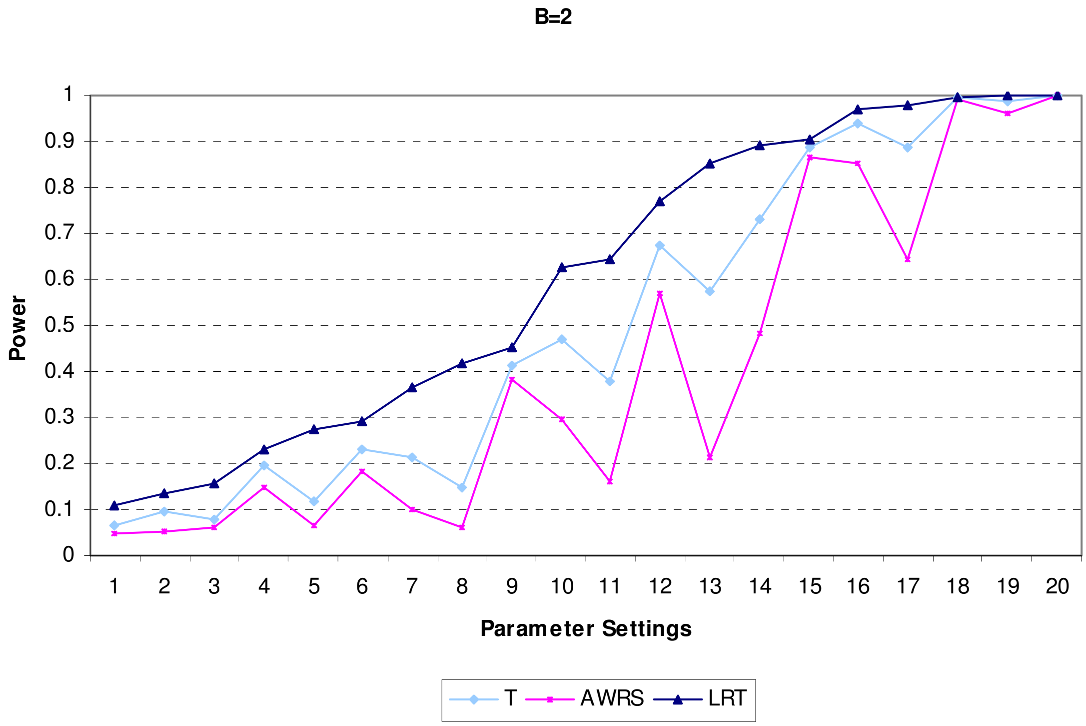


Figure 6.2 (a) (continued) – Summary of power results for the LRT using the chi-squared 95th percentile in ascending order, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=100 per group.

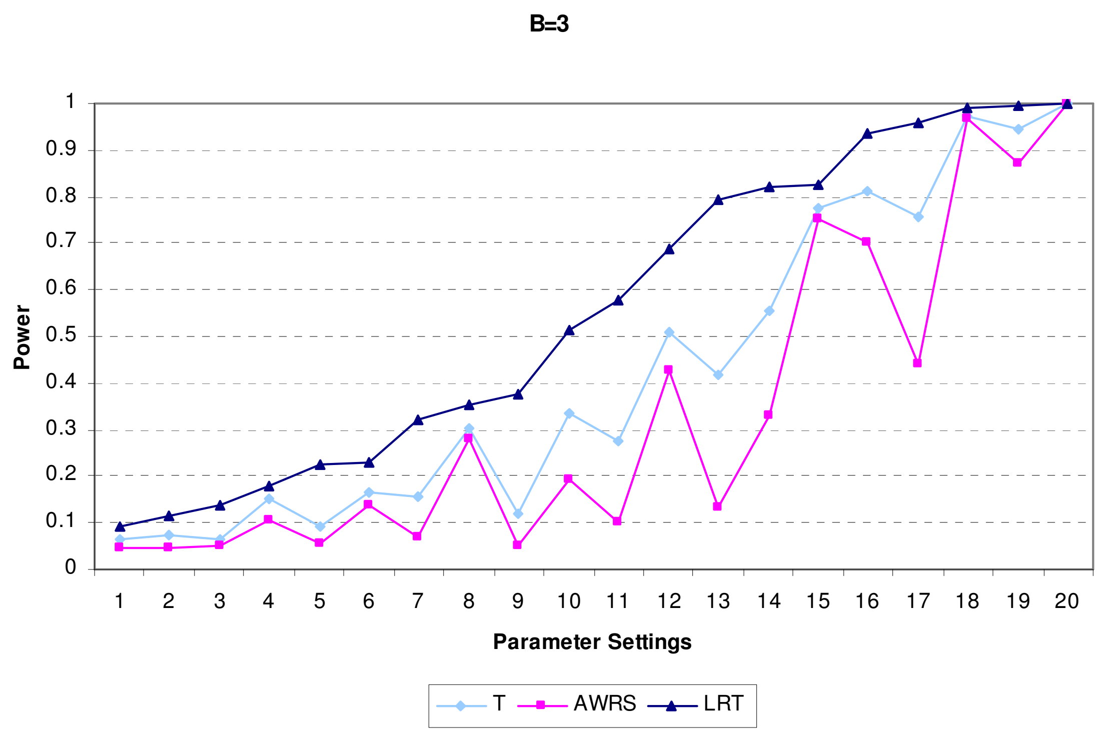


Figure 6.2 (a) (continued) – Summary of power results for the LRT using the chi-squared 95th percentile in ascending order, compared to the power of the Welch-Satterthwaite t-test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=100 per group.

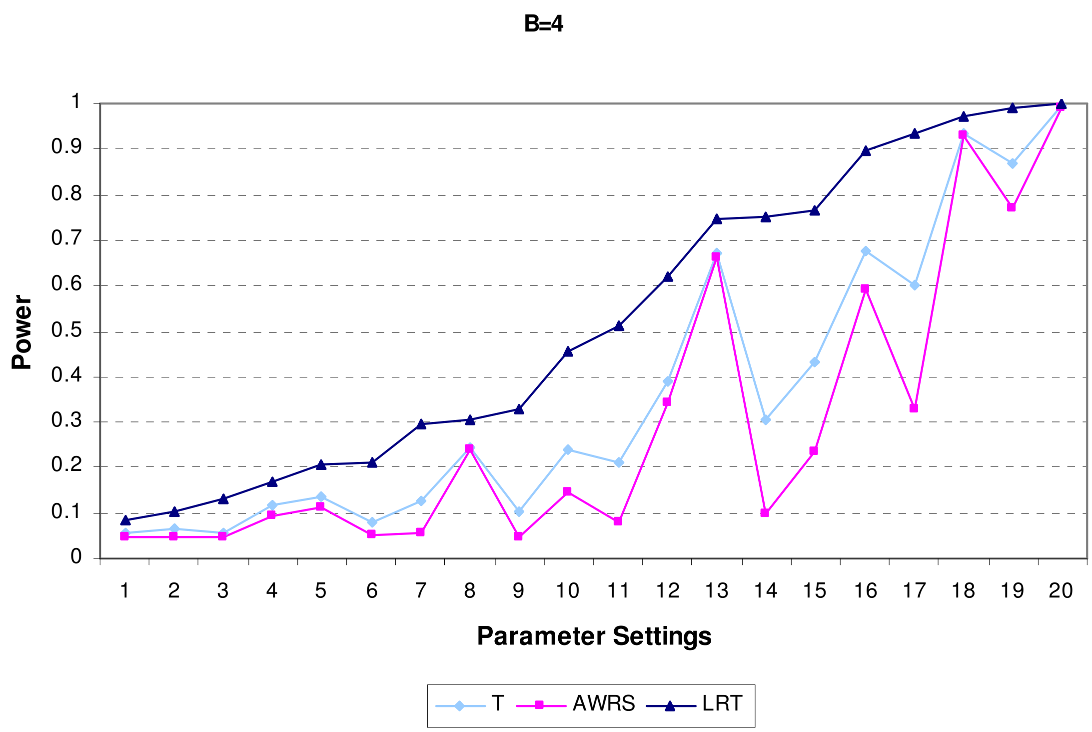


Table 6.2 (b) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=250 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
2	0.1	0.5	0.063	0.067**	0.044	0.056**	0.130
		1.0	0.128	0.133**	0.054	0.067**	0.290
		1.5	0.219	0.241**	0.060	0.067**	0.549
		2.0	0.325	0.324**	0.084	0.099**	0.761
	0.3	0.5	0.171	0.161**	0.096	0.102**	0.267
		1.0	0.476	0.475**	0.193	0.204**	0.718
		1.5	0.763	0.789**	0.326	0.347**	0.964
	0.5	0.5	0.328	0.305**	0.178	0.189**	0.423
		1.0	0.811	0.813**	0.570	0.581**	0.943
	0.7	0.5	0.535	0.542**	0.438	0.455**	0.621
	0.9	0.5	0.769	0.756*	0.719	0.728**	0.778
3	0.1	0.5	0.056	0.071**	0.046	0.054**	0.110
		1.0	0.106	0.107**	0.048	0.059**	0.228
		1.5	0.174	0.182**	0.054	0.066**	0.472
		2.0	0.257	0.252**	0.050	0.062**	0.674
	0.3	0.5	0.127	0.128**	0.074	0.080**	0.208
		1.0	0.341	0.358**	0.132	0.138**	0.690
		1.5	0.598	0.613**	0.201	0.213**	0.930
		2.0	0.797	0.798**	0.270	0.284**	0.995
	0.5	0.5	0.229	0.242**	0.154	0.158**	0.376
		1.0	0.635	0.668**	0.401	0.412**	0.885
	0.7	0.5	0.385	0.410**	0.328	0.329**	0.506
		1.0	0.888	0.895**	0.802	0.805**	0.968
	0.9	0.5	0.624	0.653*	0.611	0.615**	0.684

$$H_{010}:\begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta) \\ f(x \mid X=2) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta+\gamma) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly lower in power compared to LRT using McNemar's Test
* .05 **.001

Table 6.2 (b) (continued) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{010};
$\alpha = 1$ and n=250 per group.

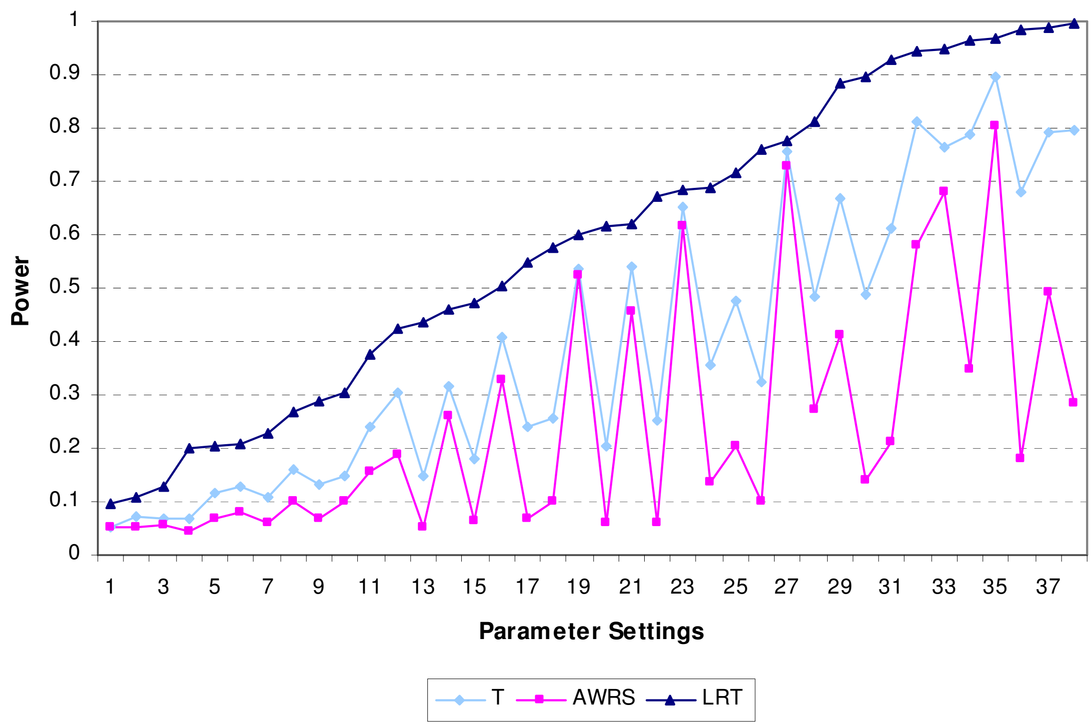
β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
4	0.1	0.5	0.050	0.054**	0.045	0.052**	0.096
		1.0	0.089	0.070**	0.035	0.045**	0.199
		1.5	0.142	0.150**	0.043	0.054**	0.436
		2.0	0.206	0.205**	0.050	0.059**	0.616
	0.3	0.5	0.100	0.117**	0.059	0.070**	0.206
		1.0	0.255	0.256**	0.092	0.102**	0.578
		1.5	0.462	0.488**	0.130	0.139**	0.898
		2.0	0.662	0.681**	0.163	0.180**	0.983
	0.5	0.5	0.171	0.147**	0.098	0.102**	0.305
		1.0	0.489	0.486**	0.269	0.274**	0.811
		1.5	0.791	0.791**	0.481	0.491**	0.987
	0.7	0.5	0.289	0.316**	0.256	0.260**	0.462
		1.0	0.768	0.765**	0.675	0.679**	0.947
	0.9	0.5	0.508	0.535**	0.516	0.526**	0.601

$$H_{010}:\begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta) \\ f(x \mid X=2) = (1-\pi)P(\alpha) + \pi P(\alpha+\beta+\gamma) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly lower in power compared to LRT using McNemar's Test
* .05 **.001

Figure 6.2 (b) – Summary of power results for the LRT using the chi-squared 95th percentile in ascending order, compared to the power of the Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank Sum Test for model H_{010}; $\alpha = 1$ and n=250 per group.



6.2.3. Results of Model H_{001} : Two Component Poisson Mixture with Equal Mixing Proportions and Unequal First Component Means

For this model, we considered 60 different configurations of the parameter settings. The configurations that we considered were $\pi_1 = \pi_2 = 0.1,\ 0.3,\ 0.5, 0.7, 0.9$; $\alpha = 1$; $\beta =\ 2,\ 3,\ 4$; $\gamma = 0.50,\ 1.00, 1.50$, 2.00 and $n_1 = n_2 = n = 100$. Using the 1000 simulated samples for each configuration of parameter values, the power for the various tests considered were estimated.

Table 6.3 (a) shows the various configurations of parameters settings used to generate the data. It contains the power of the Welch-Satterthwaite t-test, Wilcoxon Rank Sum Test, Adjusted Wilcoxon Rank Sum Test and Likelihood Ratio Test based on the simulated data for each configuration. As well, the approximate power of the t-test was calculated using equation (3.6) for each configuration and reported in Table 6.3 (a). The power of the LRT is calculated based on the 95th percentile asymptotic critical value of the chi-squared distribution with one degree of freedom, 3.84. The power of the Wilcoxon Rank Sum tests and Welch-Satterthwaite t-test were based on the 95th percentile asymptotic standard normal distribution value, 1.96.

For a majority of the configurations, the power for the LRT was greater than or equal to the power of the other tests considered. Given a specific beta and mixing proportion, the power of each of the tests increased as the value of gamma increased. As the mixing proportion decreased, corresponding to a larger percentage of the groups having different first component means, the power of the tests increased as well. As well, as beta increased there was an increase in the power of all three tests holding the other conditions constant. For all differences in component means considered, given the mixing proportion was at least 0.3 there was a significant difference in power between the LRT and Welch-Satterthwaite. The same held true for the LRT and Adjusted Wilcoxon Rank Sum test based on McNemar's test, excluding one case when $\beta = 4,\ \pi = 0.3,$ and γ=2.0. Figure 6.3 (a) illustrates the power results that were reported in Table 6.3 (a) for the various configurations for model H_{001} .

Based on the approximate power calculation of the t-test when the sample size n increased from 100 to 250 per group, there were 42 configurations where the power was less than 0.9. We decided to investigate the power of the tests considered above for these selected models. Table 6.3 (b) shows the generating model parameters used to simulate the data, and the power of the LRT, Welch-Satterthwaite t-test, Adjusted Wilcoxon Rank Sum test, Wilcoxon Rank Sum test and approximate power of the t-test. The power results from Table 6.3 (b) are illustrated in Figure 6.3 (b). In Figure 6.3 (b), the models are in ascending order based on the observed LRT power. Similarly, the power of the LRT was higher than all of the other tests considered.

Table 6.3 (a) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
2	0.1	0.5	0.276	0.275**	0.272	0.280**	0.317
		1.0	0.787	0.745**	0.784	0.787	0.793
		1.5	0.986	0.982	0.990	0.990	0.986
		2.0	1.000	1.000	1.000	1.000	1.000
	0.3	0.5	0.161	0.166**	0.174	0.181*	0.211
		1.0	0.486	0.492**	0.548	0.559**	0.632
		1.5	0.814	0.824**	0.889	0.890**	0.921
		2.0	0.963	0.954**	0.982	0.982**	0.994
	0.5	0.5	0.097	0.089**	0.085	0.088**	0.131
		1.0	0.255	0.248**	0.269	0.272**	0.392
		1.5	0.483	0.497**	0.545	0.549**	0.727
		2.0	0.705	0.705**	0.790	0.792**	0.926
	0.7	0.5	0.058	0.078**	0.069	0.072**	0.122
		1.0	0.117	0.126**	0.110	0.113**	0.237
		1.5	0.204	0.213**	0.216	0.218**	0.411
		2.0	0.316	0.354**	0.367	0.370**	0.683
	0.9	0.5	0.034	0.053**	0.049	0.053**	0.087
		1.0	0.045	0.074**	0.062	0.068**	0.111
		1.5	0.058	0.078**	0.076	0.078**	0.145
		2.0	0.074	0.087**	0.078	0.081**	0.200

$$H_{001}:\begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha-\beta)+\pi P(\alpha) \\ f(x \mid X=2) = (1-\pi)P(\alpha-\beta-\gamma)+\pi P(\alpha) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemar's Test
* .05 **.001

Table 6.3 (a) (continued) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon – Adj	LRT
3	0.1	0.5	0.294	0.303**	0.322	0.331*	0.359
		1.0	0.814	0.781**	0.854	0.857*	0.874
		1.5	0.990	0.988*	0.997	0.997	0.997
		2.0	1.000	0.999	1.000	1.000	1.000
	0.3	0.5	0.153	0.138**	0.172	0.175**	0.237
		1.0	0.455	0.468**	0.588	0.590**	0.714
		1.5	0.777	0.769**	0.909	0.912**	0.966
		2.0	0.944	0.937**	0.992	0.993*	1.000
	0.5	0.5	0.091	0.099**	0.114	0.119**	0.159
		1.0	0.230	0.257**	0.314	0.319**	0.538
		1.5	0.432	0.452**	0.538	0.542**	0.848
		2.0	0.642	0.641**	0.751	0.751**	0.986
	0.7	0.5	0.056	0.063**	0.064	0.067**	0.114
		1.0	0.108	0.105**	0.103	0.105**	0.277
		1.5	0.185	0.181**	0.176	0.180**	0.542
		2.0	0.283	0.268**	0.258	0.262**	0.856
	0.9	0.5	0.033	0.055**	0.056	0.057**	0.113
		1.0	0.044	0.057**	0.049	0.049**	0.142
		1.5	0.057	0.072**	0.065	0.067**	0.228
		2.0	0.072	0.079**	0.067	0.070**	0.342

$$H_{001}:\begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha-\beta)+\pi P(\alpha) \\ f(x \mid X=2) = (1-\pi)P(\alpha-\beta-\gamma)+\pi P(\alpha) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemar's Test
* .05 **.001

Table 6.3 (a) (continued) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=100 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
4	0.1	0.5	0.306	0.296**	0.396	0.404**	0.446
		1.0	0.828	0.826**	0.933	0.942**	0.970
		1.5	0.992	0.990*	1.000	1.000	1.000
		2.0	1.000	1.000	1.000	1.000	1.000
	0.3	0.5	0.141	0.143**	0.211	0.215**	0.320
		1.0	0.410	0.412**	0.645	0.652**	0.870
		1.5	0.720	0.722**	0.943	0.944**	0.998
		2.0	0.911	0.902**	0.997	0.997	1.000
	0.5	0.5	0.083	0.089**	0.116	0.118**	0.229
		1.0	0.202	0.174**	0.268	0.273**	0.676
		1.5	0.376	0.371**	0.550	0.555**	0.961
		2.0	0.568	0.591**	0.771	0.771**	0.999
	0.7	0.5	0.053	0.061**	0.059	0.059**	0.142
		1.0	0.099	0.114**	0.106	0.107**	0.420
		1.5	0.165	0.159**	0.150	0.153**	0.764
		2.0	0.249	0.259**	0.232	0.237**	0.965
	0.9	0.5	0.033	0.050**	0.042	0.042**	0.135
		1.0	0.043	0.060**	0.054	0.054**	0.184
		1.5	0.055	0.072**	0.049	0.053**	0.334
		2.0	0.069	0.074**	0.059	0.060**	0.535

$$H_{001}:\begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha-\beta)+\pi P(\alpha) \\ f(x \mid X=2) = (1-\pi)P(\alpha-\beta-\gamma)+\pi P(\alpha) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemar's Test
* .05 **.001

Figure 6.3 (a) – Summary of power results for the LRT using the chi-squared 95th percentile, compared to the power of the Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=100 per group.

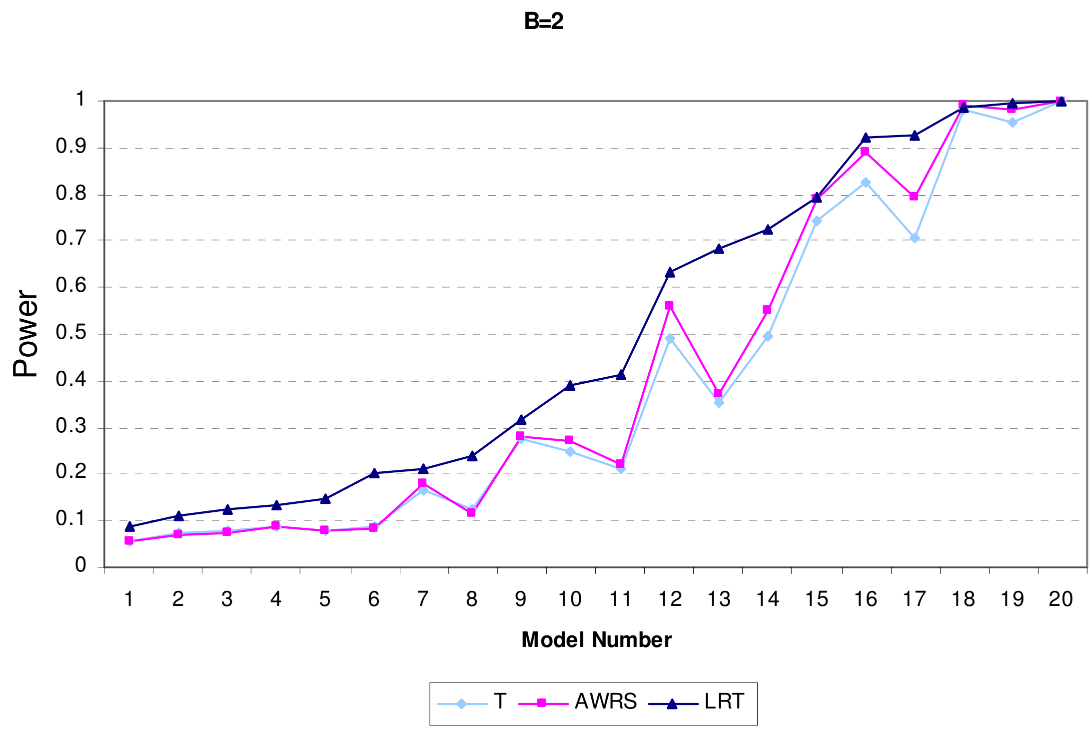


Figure 6.3 (a) (continued)– Summary of power results for the LRT using the chi-squared 95th percentile, compared to the power of the Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=100 per group.

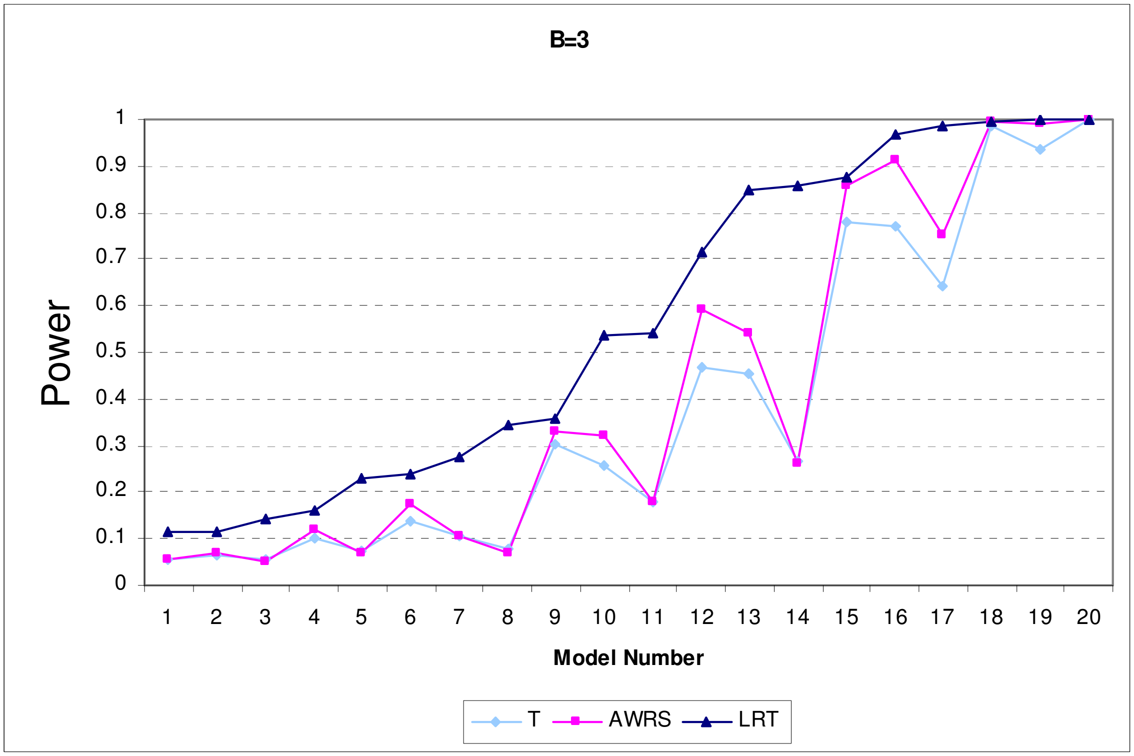


Figure 6.3 (a) (continued)– Summary of power results for the LRT using the chi-squared 95th percentile, compared to the power of the Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=100 per group.

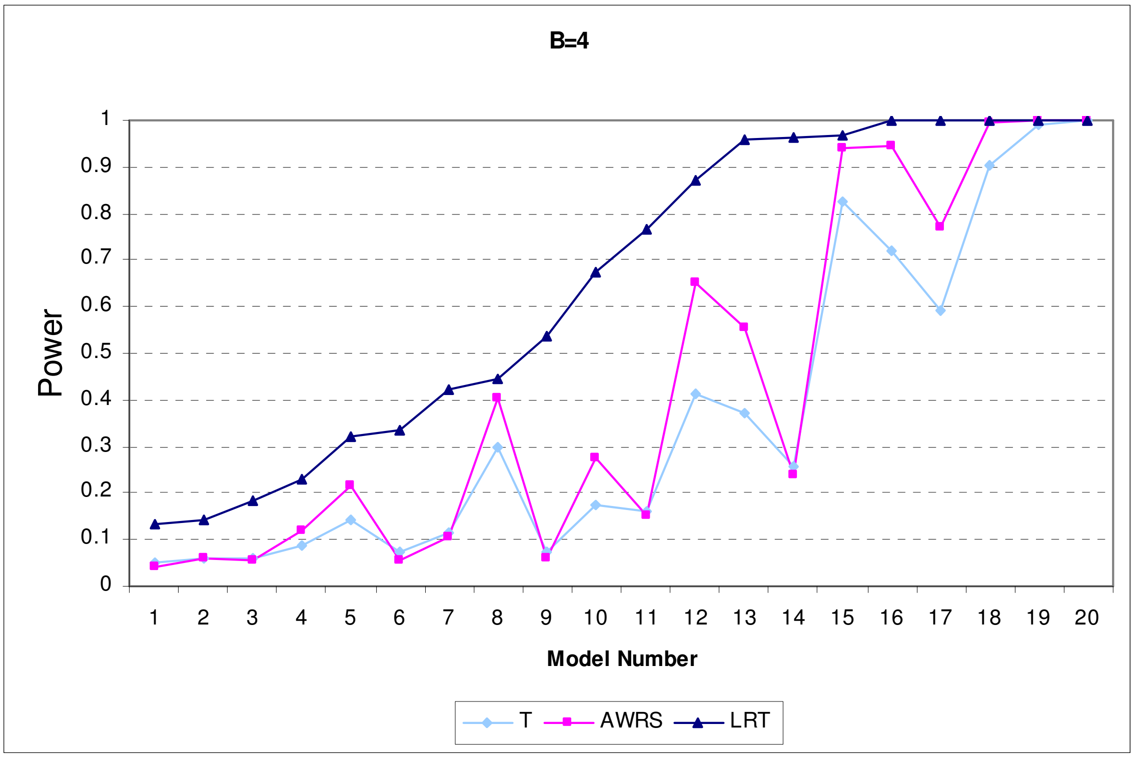


Table 6.3 (b) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=250 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon - Adj	LRT
2	0.1	0.5	0.578	0.575**	0.586	0.596	0.611
	0.3	0.5	0.335	0.342**	0.382	0.390*	0.418
		1.0	0.861	0.842**	0.895	0.895**	0.923
	0.5	0.5	0.180	0.202**	0.209	0.212**	0.276
		1.0	0.539	0.548**	0.611	0.617**	0.756
		1.5	0.858	0.855**	0.917	0.919**	0.985
	0.7	0.5	0.090	0.090**	0.096	0.100**	0.166
		1.0	0.228	0.214**	0.211	0.216**	0.405
		1.5	0.433	0.448**	0.470	0.477**	0.802
		2.0	0.649	0.648**	0.667	0.674**	0.958
	0.9	0.5	0.040	0.057**	0.056	0.057**	0.089
		1.0	0.061	0.079**	0.070	0.070**	0.155
		1.5	0.089	0.097**	0.092	0.093**	0.257
		2.0	0.125	0.146**	0.121	0.125**	0.425

$$H_{001}:\begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha-\beta) + \pi P(\alpha) \\ f(x \mid X=2) = (1-\pi)P(\alpha-\beta-\gamma) + \pi P(\alpha) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemar's Test
* .05 **.001

Table 6.3 (b) (continued) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{001}; $\alpha = 7$ and n=250 per group.

β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon – Adj	LRT
3	0.1	0.5	0.611	0.598**	0.646	0.651*	0.669
	0.3	0.5	0.317	0.315**	0.395	0.406**	0.478
		1.0	0.832	0.814**	0.916	0.916**	0.972
	0.5	0.5	0.165	0.165**	0.190	0.192**	0.323
		1.0	0.489	0.473**	0.574	0.579**	0.862
		1.5	0.808	0.815**	0.918	0.918**	0.999
	0.7	0.5	0.084	0.084**	0.086	0.088**	0.182
		1.0	0.208	0.201**	0.212	0.213**	0.593
		1.5	0.391	0.388**	0.391	0.394**	0.937
		2.0	0.592	0.582**	0.570	0.572**	0.997
	0.9	0.5	0.039	0.066**	0.061	0.062**	0.115
		1.0	0.059	0.070**	0.069	0.070**	0.194
		1.5	0.086	0.102**	0.080	0.083**	0.390
		2.0	0.120	0.118**	0.088	0.088**	0.677

$$H_{001}: \begin{cases} f(x \mid X=1) = (1-\pi)P(\alpha-\beta) + \pi P(\alpha) \\ f(x \mid X=2) = (1-\pi)P(\alpha-\beta-\gamma) + \pi P(\alpha) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemar's Test
* .05 **.001

Table 6.3 (b) (continued) - Power of the LRT using the chi-squared 95th percentile, compared to the power of the Approximate t-test, Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum Test for model H_{001};
$\alpha = 7$ and n=250 per group.

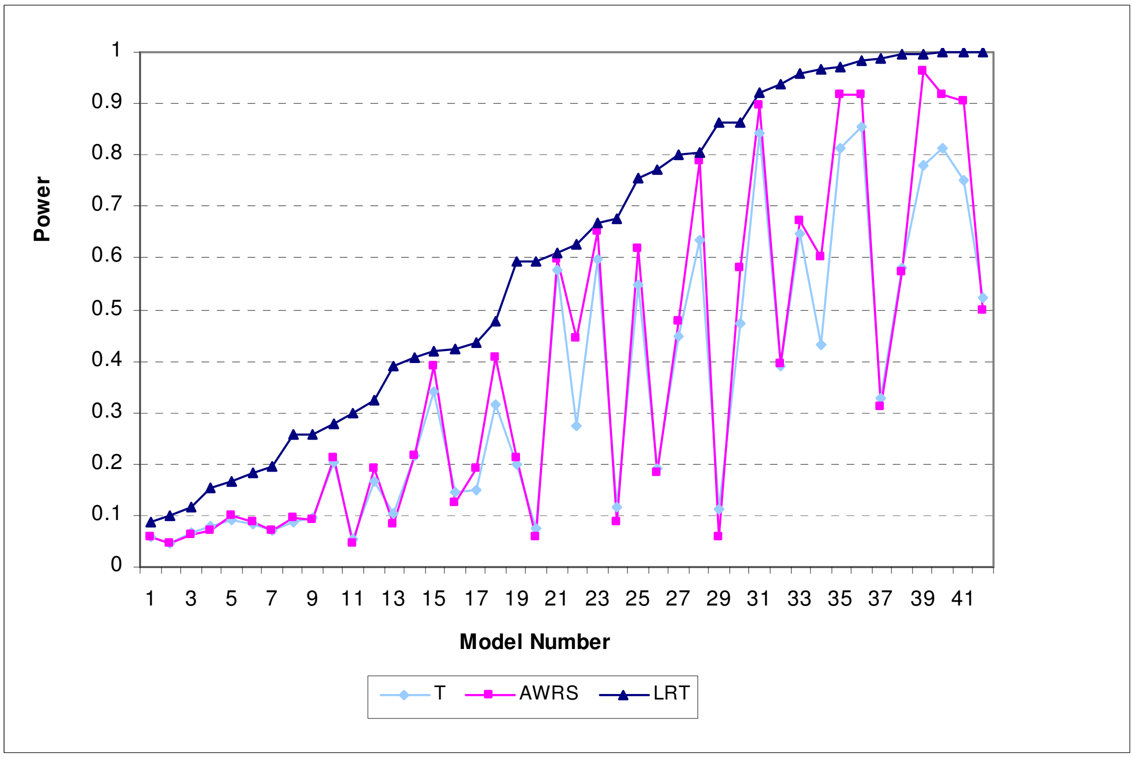
β	π	γ	T Approximate	T Observed	Wilcoxon	Wilcoxon – Adj	LRT
4	0.1	0.5	0.631	0.636**	0.777	0.787*	0.805
	0.3	0.5	0.287	0.275**	0.440	0.446**	0.625
		1.0	0.782	0.782**	0.962	0.963**	0.997
	0.5	0.5	0.147	0.149**	0.192	0.192**	0.434
		1.0	0.428	0.432**	0.600	0.603**	0.966
		1.5	0.739	0.752**	0.901	0.904**	1.000
	0.7	0.5	0.078	0.087**	0.092	0.094**	0.256
		1.0	0.185	0.190**	0.182	0.184**	0.771
		1.5	0.344	0.329**	0.310	0.312**	0.987
		2.0	0.526	0.524**	0.495	0.496**	1.000
	0.9	0.5	0.039	0.045**	0.043	0.044**	0.101
		1.0	0.058	0.055**	0.044	0.046**	0.297
		1.5	0.083	0.076**	0.056	0.059**	0.594
		2.0	0.114	0.111**	0.059	0.060**	0.861

$$H_{001}:\begin{cases} f\left(x \mid X=1\right)=\left(1-\pi\right)P\left(\alpha-\beta\right)+\pi P\left(\alpha\right) \\ f\left(x \mid X=2\right)=\left(1-\pi\right)P\left(\alpha-\beta-\gamma\right)+\pi P\left(\alpha\right) \end{cases}$$

The margin of error of ± 0.03 for each configuration.

Significantly different in power compared to LRT using McNemar's Test
* .05 **.001

Figure 6.3 (b) – Summary of power results for the LRT using the chi-squared 95th percentile, compared to the power of the Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank Sum Test for model H_{001} ; $\alpha = 7$ and n=250 per group.



Chapter 7

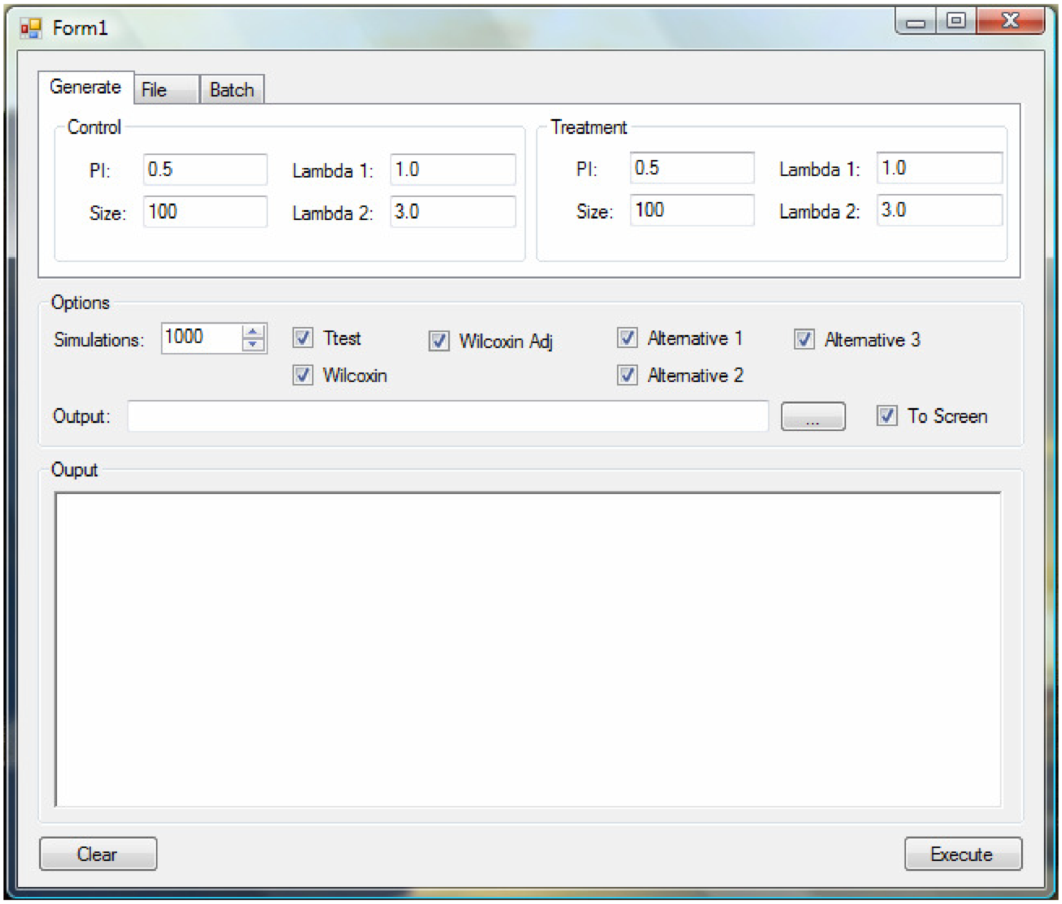
Computer Software

The computer application consists of two assemblies. One assembly is a shared DLL which is written in unmanaged C++. The other assembly is a Microsoft .Net WinForms application that is written in C++\CLI. We choose to use unmanaged C++ for the core functionality of the program because it allowed greater performance and efficiency. However, it comes at the cost of more complicated implementation and memory management concerns.

The initial design consisted of a console, command line driven, interface. This made sense for initial testing purposes but was impractical for production runs. Therefore, to make the software more user-friendly we implemented a WinForm graphical user interface (GUI). The application is designed for the user to interact with the GUI and have it call through to the unmanaged DLL which performs the core calculations.

The application provides the user with the ability to simulate data or input their own data. When simulating data, the user will choose the Generate tab at the top of the GUI. They will then input the mixing proportion, component means, and sample size for each group. The default values for both the control group and treatment group are 0.5 for the mixing proportion, 1.0 for the first component mean, and 3.0 for the second component mean. A picture of the Generate tab is given in Figure 7.1

Figure 7.1 - Picture of Computer Software Graphical User Interface (GUI) Generate Tab

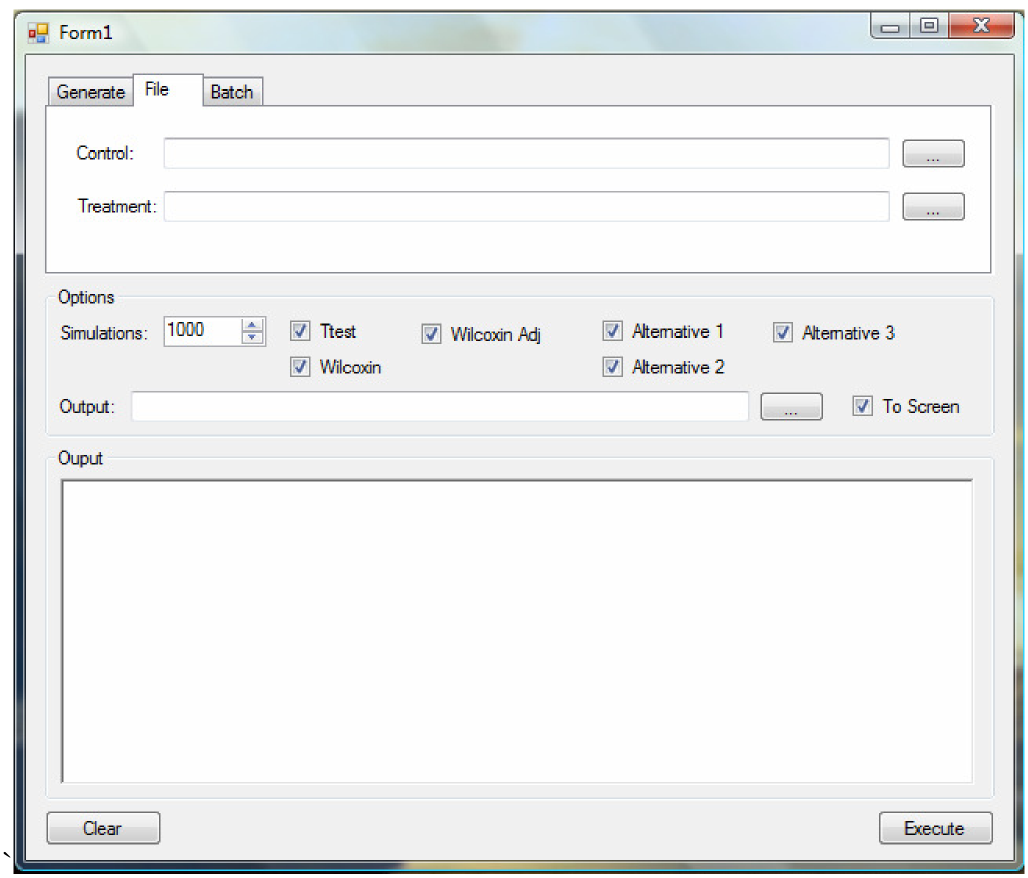


The user has the choice of selecting the number of simulations and which tests they would like to run. The possible test options are the Satterthwaite t-test, Wilcoxon Rank Sum test, Adjusted Wilcoxon Rank Sum test, Alternative 1 (LRT – Model H_{100}), Alternative 2 (LRT – Model H_{010}) and Alternative 3 (LRT – Model H_{001}). The user also has the option of having the simulated data values, tests results, and power results printed to the screen or to an output .xml file which they specify. If the user has a large number of different cases they would like to simulate, they can choose to use the batch option.

They would click on the Batch tab and input an .xml file which includes all the different parameter settings they would like to run along with corresponding output .xml files.

If the user has their own data, which they would like to run through the program, they may they do that as well. To input their own data, they would choose the File tab. They will then input a .txt file containing the count data for the control group and a .txt file containing the count data for the treatment group. They then have the similar options as to when simulating data. They get to choose which tests they want run, the number of simulations, and whether the results are printed to the screen or to an output file. A picture of the File tab is given in Figure 7.2.

Figure 7.2 - Picture of Computer Software Graphical User Interface (GUI) File Tab



Chapter 8

Applications of Finite Mixture of Poisson Distributions

We applied our testing procedure for comparing two groups of two-component Poisson mixtures for two sets of count data that were provided. One data set that we studied consisted of the number of fibromas (benign tumors) which existed on patients suffering from the disease tuberous sclerosis. The other data set that we applied our procedure to consisted of the number of deviant verbalizations from a study on schizophrenia.

8.1 Analysis of Fibroma Data

To begin, we divided the sample into two groups based on the age thirty. We fit single Poisson distributions to each group using the sample mean of each group. We conducted chi-squared goodness of fit tests for each group using the single Poisson distributions which both resulted in p-values less than .0001 indicating extremely poor fits. As well, for each group the variance was much higher than the mean. Thus, it suggested that within each group there existed heterogeneity.

Using our program, we ran each age group versus itself under the null hypothesis to obtain estimates for a two-component Poisson mixture within the group. For the control group (below 30 years old) we obtained 0.22 for the mixing proportion, 0.15 for the first component mean and 3.99 for the second component mean. For the treatment group (at least 30 years old) we obtained 0.37 for the mixing proportion, 1.81 for the first component mean and 10.73 for the second component mean.

Figure 8.1 is a plot of the count of fibromas for the control group and corresponding probability distribution function of the single Poisson distribution and mixture distribution fit to the data. Figure 8.2 is a plot of the count of fibromas for the treatment group and corresponding probability distribution function of the single Poisson distribution and mixture distribution fit to the data. As one can see, it appears as if the two-component mixture for each group fits the data much better. This reiterates that there possibly exists some type of heterogeneity within each group.

Figure 8.1 – Observed Number of Fibromas for Control Group (Below 30 years old) along with corresponding Expected Counts for Single Poisson Distribution and Two-Component Mixture Poisson Distribution.

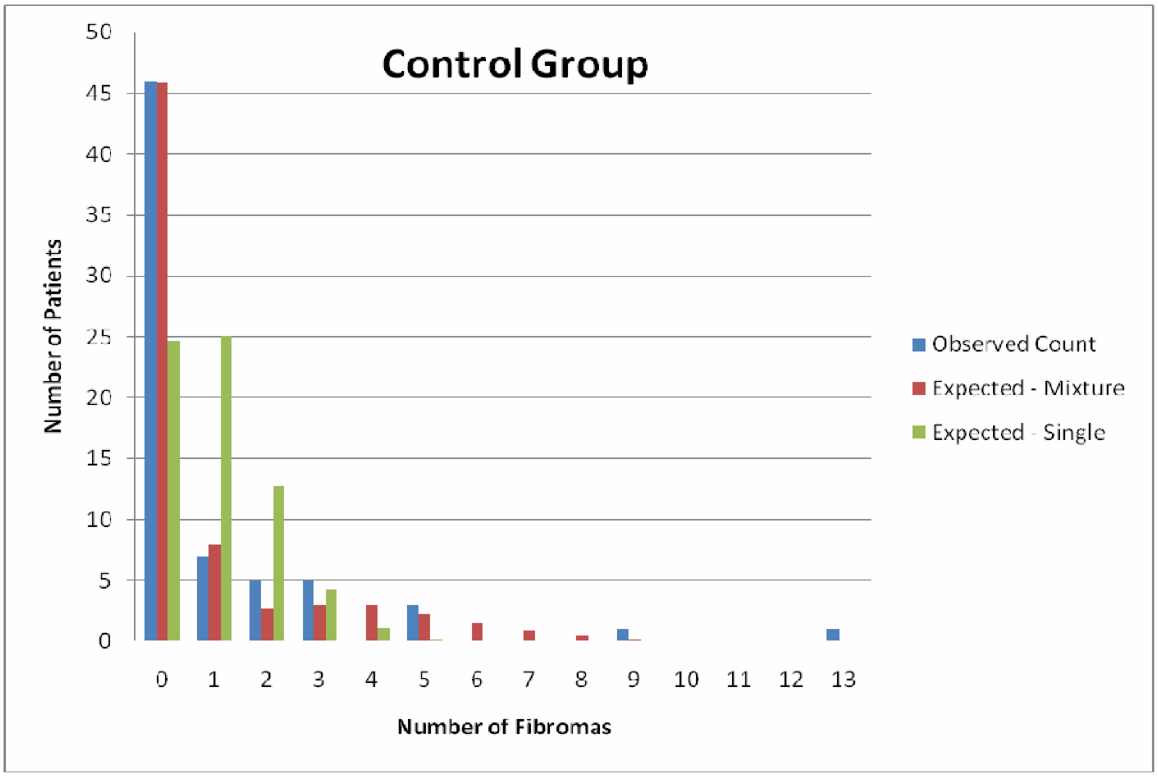
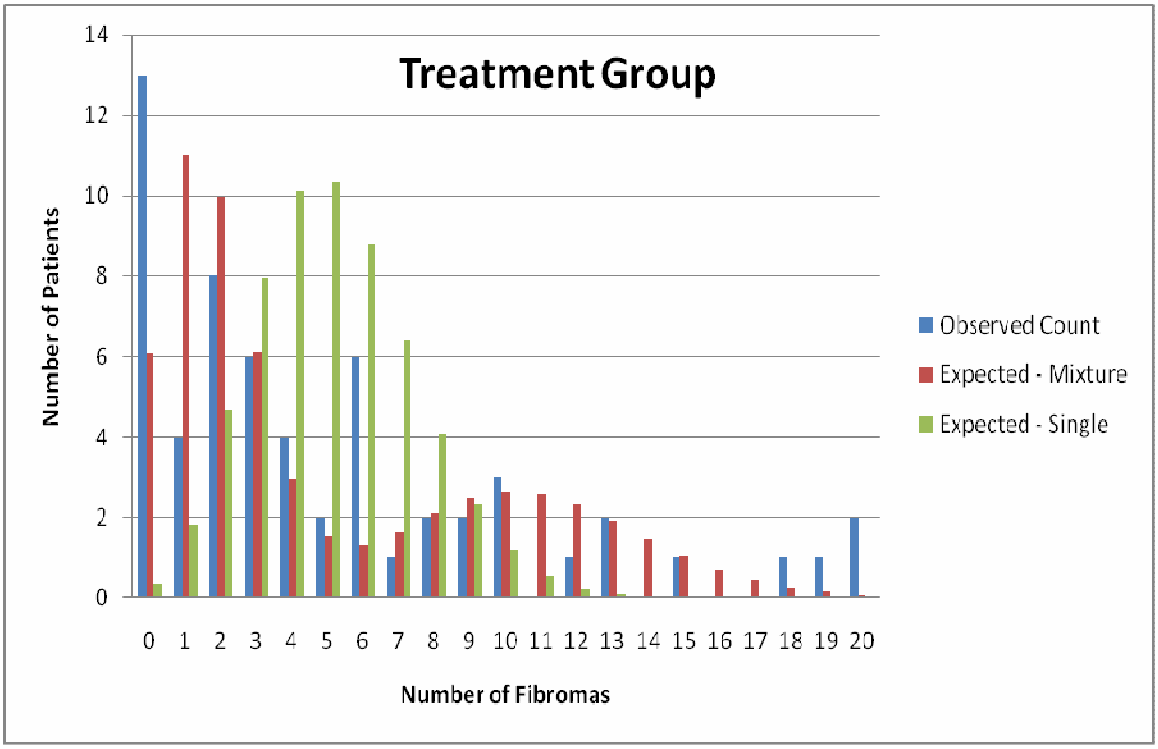


Figure 8.2 – Observed Number of Fibromas for Treatment Group (At least 30 years old) along with corresponding Expected Counts for Single Poisson Distribution and Two-Component Mixture Poisson Distribution.



We then applied our testing procedure for comparing two groups of two-component Poisson mixtures. We considered all three alternative models for the two groups. Based on the Likelihood Ratio Test statistics, there was a significant difference between the two groups for all three alternatives. The likelihood ratio test statistics were 28.83 for model H_{100}, 8.82 for model H_{010}, and 29.19 for model H_{001}.

8.2 Deviant Verbalization Data

With this data set, we compared the number of deviant verbalizations of normal controls to siblings of Schizophrenics.   For each group, there was an abundant count of zeros and the variance was significantly higher than the mean. Figures 8.3 and 8.4 are plots of the observed count of deviant verbalizations for the control group and sibling group respectively.

Figure 8.3 – Observed count of deviant verbalizations for the normal controls

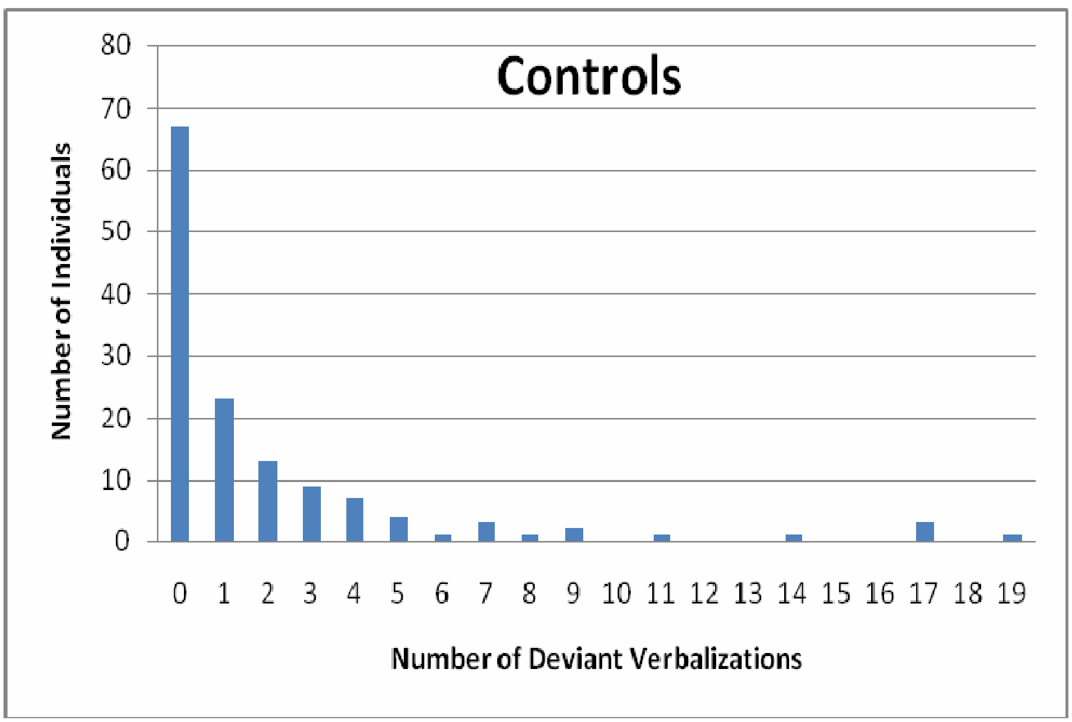
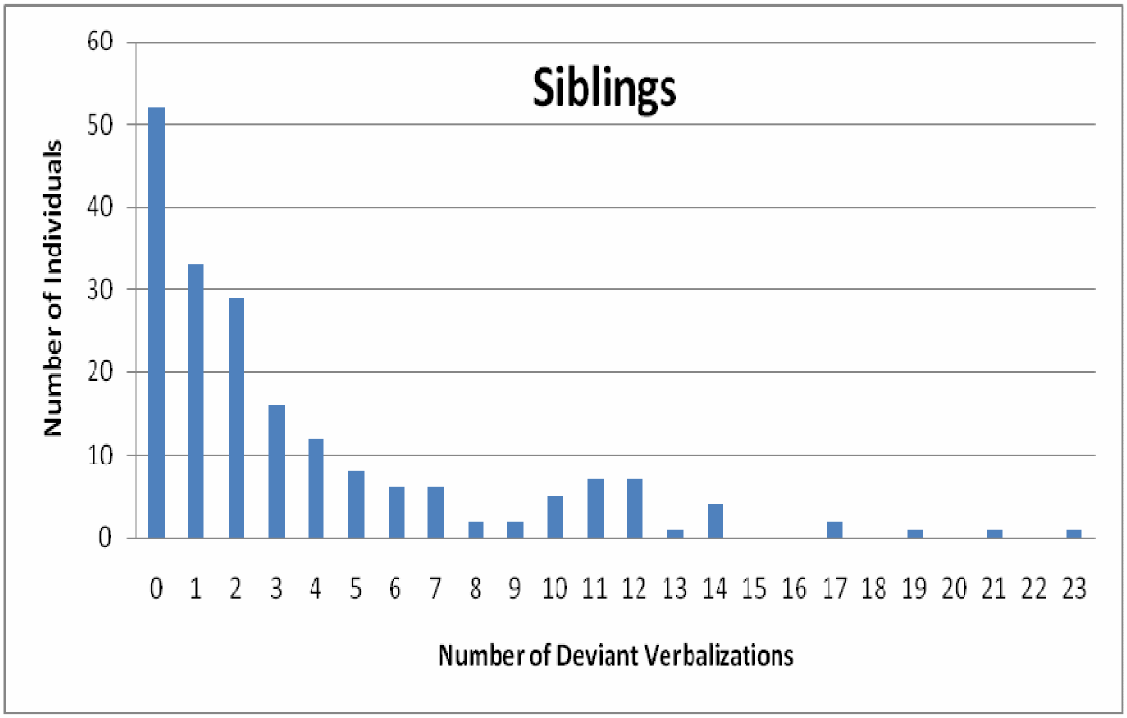


Figure 8.4– Observed count of deviant verbalizations for siblings of schizophrenics



It was believed that the siblings of Schizophrenics have a higher frequency of deviant verbalizations compared to the normal controls. Therefore, we believed if a two-component Poisson mixture existed within each group that the sibling group would have a higher mixing proportion compared to the control group. We applied our two group two-component Poisson mixture test to the data sets. We considered all three alternative models.

We found that there was a significant difference between the mixing proportion for the two groups (Model H_{100}) and first component means (Model H_{001}). The LRT test statistic for the difference between mixing proportions was 12.04. For Model H_{100}, the estimated mixing proportion for the control group was 0.12 and the mixing proportion for the sibling group was 0.28 with first and second component means of 1.12 and 9.76, respectively. The LRT statistic for a difference between first component means was 16.76. For Model H_{001}, the first component mean for the control group was estimated to be 0.80 and 1.43 for the sibling group with a common second component mean of 9.87 with a common mixing proportion of 0.21.

Chapter 9

Discussion and Conclusions

For this study, we derived a Likelihood Ratio Test comparing two groups assuming a two-component Poisson mixture existed within each group. We considered three different alternatives: (1) Difference in mixing proportions only (Model H_{100}) (2) Differences in second component means only (Model H_{010}) and (3) Differences in first component means only (Model H_{001}).

Through a simulation study the power of the LRT was compared to the Welch-Satterthwaite t-test, Wilcoxon Rank Sum test, and Adjusted Wilcoxon Rank Sum test for sample sizes of 100 and 250 per group. The power of these tests were based on the asymptotic 95th percentile critical value of the chi-squared distribution with one degree of freedom for the LRT and the asymptotic 95th percentile critical value of the standard normal distribution for the Welch-Satterthwaite t-test and Wilcoxon Rank Sum tests. We found that in a majority of the cases the LRT was significantly more powerful compared to the other tests based on McNemar's test.

One major concern was that the likelihood ratio test did not follow an asymptotic chi-squared distribution with the degrees of freedom equal to the difference of the number of parameters between the two hypotheses. Therefore, we investigated the size of the test under the null hypothesis at the 5% significance level. It appeared that the critical value used for the Wilcoxon Rank Sum tests and t-test were valid, however the size of the LRT seemed slightly inflated.

Based on the study of the size of the LRT, we decided to investigate the empirical null distribution of the LRT. Since all of the 3 alternatives differed by 1 parameter from the null hypothesis, we used simulations involving all of them. Based on these simulations, we derived the 95th percentile of the empirical null distribution for the LRT for sample sizes of 100, 500 and 1000 per group. As the sample size increased per group, the critical value approached the asymptotic 95th percentile of the chi-squared distribution with one degree of freedom.

Using the empirical critical values that we derived for the LRT, we conducted a similar power study using sample sizes of 100 per group. When comparing two groups whose mixing proportions differed, it appeared that the power of the LRT, Welch-Satterthwaite t-test and Adjusted Wilcoxon Rank Sum test were all relatively close to one another. However when a difference in one of the component means existed, the LRT was significantly more powerful than the other two tests.

Based on the study it appears that when comparing a control group and a treatment group where a two-component Poisson mixture is though to exist within each group that the LRT is more powerful at detecting a difference between the mixtures. In comparisons to the other tests, the LRT does take longer to compute due to the EM algorithm; however it seems worth the effort due to the increase in power. Just as in the past when technology limited maximum likelihood estimation, as time goes on the advancements in computers will only speed up this process.

Based on this dissertation, we considered only 3 alternatives and conducted a power study using sample sizes of 100 and 250 per group. For future work, one may want to investigate the difference in power for these tests using small sample sizes. Another topic involving smaller sample sizes that would be of interest is the empirical null distribution of the LRT. As well, the LRT could be extended to involve combinations of the alternatives that we presented in this dissertation.

Bibliography

1. Algina J., Oshima T.C., and Lin W.Y. (1994) Type I error rates for Welch's test and James second order test under normality and inequality of variance when there are two groups. Journal of Educational and Behavioral Statistics 19, 275-291.

2. Arley, N. and Buch, K.R. (1950) Introduction to the Theory of Probability and Statistics. New York: Wiley.

3. Atwood, L.D., Wilson, A.F., Elston, R.C., and Bailey-Wilson, J.E. (1992) Computational aspects of fitting a mixture of two normal distributions using maximum likelihood. Comm. Statist. Simulation Comput. 21, 769-781.

4. Baker, G.A. (1940) Maximum-likelihood estimation of the ratio of the two components of non-homogeneous populations. Tohoku Math. J., 47, 304-308.

5. Blair R.C., and Higgins J.J. (1980). A comparison of the power of Wilcoxon' Rank Sum statistic to that of Student's t statistic under various nonnormal distributions. Journal of Educational Statistics 5, 309-335.

6. Blair, R.C., Higgins, J.J., and Smitley, W.D.S. (1980). On the relative power of the U and t-tests. British Journal of Mathematical and Statistical Psychology 33 (1), 114-120.

7. Blischke, W.R. (1962). Moment estimators for the parameters of a mixture of two binomial distributions. Ann. Math. Statist., 33, 444-454.

8. Blischke, W.R. (1964). Estimating the parameters of mixtures of binomial distributions. Journal of the American Statistical Association 59, 510-528.

9. Blischke, W.R. (1965). Mixtures of discrete distributions. In Classical and Contagious Discrete Distributions (Ed. G.P. Patil), 351-372. Pergamon, New York.

10. Bohning, D., Dietz, E., Schaub, R., Schlattman, P., and Lindsay, B. (1994). The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family. Ann. Inst. Statist. Math. 46, 373-388.

11. Bohning, D. (1998) Zero-inflated Poisson models and C.A.MAN: a tutorial collection of evidence. Biometrical Journal 40 (7), 833-843.

12. Bohning, D. (1999). Computer Assisted Analysis of Mixtures & Applications in Meta-analysis, Disease Mapping & Others. CRC Press, Boca Raton, FL.

13. Bohning, D., Dietz, E., and Schlattmann, P. (1999) The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. Journal of the Royal Statistical Society A 162(2), 195-209

14. Boneau C.A. (1962). A comparison of the power of the U and t-tests. Psychological Review 69, 246-256.

15. Boos, D.D., and Brownie C. (1986). Testing for a treatment effect in the presence of nonresponders. Biometrics 42, 191-197.

16. Bradley, J.V. (1968). Distribution-free statistical test. Englewood Cliffs, N.J. : Prentice Hall

17. Campbell, M.J., Machin, D. and D'Arcangues, C. (1991) Coping with extra-Poisson variability in the analysis of factors influencing vaginal ring expulsions. Statistics in Medicine 10, 241-251.

18. Chen, H. and Chen, J. (2001) The Likelihood Ratio Test for Homogeneity in Finite Mixture Models. The Canadian Journal of Statistics 29, 201-215.

19. Cohen, A.C. (1963) "Estimation in Mixtures of Discrete Distributions," in Proceedings of the International Symposium on Discrete Distributions, Montreal, 373-378.

20. Cohen, A.C. (1965) Estimation in mixtures of discrete distributions. In Classical and Contagious Discrete Distributions (ed. G.P. Patil) 373-378. Pergamon, New York.

21. Conover, W.J., and Salsburg, D.S. (1988). Locally most powerful tests for treatment effects when only a subset of patients can be expected to respond to treatment. Biometrics 44, 189-196.

22. Coombs, W.T., Algina, J., and Oltman, D. (1996) Univariate and multivariate omnibus hypothesis tests selected to control type I error rates when population variances are not necessarily equal. Rev. Educ. Res. 66, 137-179.

23. Cox, D.B., and Hinkley, D.V. (1974). Theoretical Statistics. London: Chapman and Hall.

24. Cutler, A., and Cordero-Brana, O. (1996) Minimum Hellinger distance estimation for finite mixture models. Journal of the American Statistical Association 91, 1716-1724.

25. Day, N.E. (1969) Estimating the components of a mixture of normal distributions, Biometrika 56, 463-474.

26. Dempster, A.P., Laird, N.M., and Rubin D.R. (1977). Maximum likelihood from incomplete data (with discussion). Journal of the Royal Statistical Society B 39, 629-646.

27. Duan, T. (2005) A Power Study of a Two Sample Likelihood Ratio Test in the Presence of Mixtures. Ph.D. thesis. State University of New York at Stony Brook.

28. Duan, T., Finch, S. J., Ye, K. Q., Chase, G. A., and Mendell, N. R. (2005). Using Mixture Models to Characterize Disease Related Traits. BMC Genetics 6, S99.

29. Engelman, L., and Hartigan, J.A. (1969). Percentage points of a test for clusters. Journal of the American Statistical Association 64, 1647-1648.

30. Everitt, B.S., and Hand, D.J. (1981). Finite Mixture Distributions. London: Chapman and Hall.

31. Faddy, M.J. (1994). On variation in Poisson processes. Mathematical Scientist 19, 47-51.

32. Fisher, R.A (1935). The Design of Experiments. London: Oliver & Boyd.

33. Fong, D.Y.T. and Yip, P. (1993). An EM algorithm for a mixture model of count data. Statist. Probab. Lett., 17, 325-332.

34. Fowlkes, E.B. (1979). Some methods for studying the mixture of two normal (lognormal) distributions. Journal of the American Statistical Association 74, 561-573.

35. Good, P.I. (1979). Detection of a treatment effect when not all experimental subjects will respond to treatment. Biometrics 35, 483-489.

36. Gupta, P.L., Gupta, R.C., and Tripathi, R.C. (1996). Analysis of zero-adjusted count data. Computational Statistics and Data Analysis 23, 207-218.

37. Hartigan, J.A. (1985). A failure of likelihood asymptotics for normal mixtures. In Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. 2, Monterey: Wadsworth, 807-810.

38. Hasselbad, V. (1969) Estimation of finite mixtures from the exponential family. Journal of the American Statistical Association 64, 1459-1471.

39. Heilbron, D.C. (1989), “Generalized Linear Models for Altered Zero Probabilities and Overdispersion in Count Data,” unpublished technical report, University of California, San Francisco, Dept. of Epidemiology and Biostatistics.

40. Hodges, J.L., and Lehmann E.L. (1956). The efficiency of some nonparametric competitors of the t-test. Annals of Mathematical Statistics 27, 324-335.

41. John, S. (1970). ON analyzing mixed samples. Journal of the American Statistical Association 65, 755-760.

42. Johnson, N.L., and Kotz, S. (1969) Discrete Distributions. New York: Wiley.

43. Johnson, N., Kotz, S. and Kemp, A. W. (1992) Univariate Discrete Distributions, 2^{nd} edition. New York: Wiley.

44. Karlis, D., Xekalaki, E. (1998) Minimum Hellinger distance estimation for finite Poisson mixtures. Comput. Statist. Data Anal. 29, 81-103.

45. Karlis D., Xekalaki, E. (1999a) On testing the number of components in finite Poisson mixtures. Ann. Ins. Statist. Math. 51, 149-161.

46. Karlis, D., Xekalaki, E. (1999b) Improving the EM algorithm for mixtures. Statistics and Computing 9, 303-307.

47. Karlis, D., Xekalaki, E. (2000) A Simulation Comparison of Several Procedures for Testing the Poisson Assumption. The Statistician 49, 355-382.

48. Karlis, D., and Xekalaki, E. (2003a) A Zero Frequency Alternative Method to the Moment Method of Estimation in Finite Poisson Mixtures. Journal of Statistical Computation and Simulation 73, 409-427.

49. Karlis, D. and Xekalaki, E. (2003b) Choosing initial values for the EM algorithm for finite mixtures. Computational Statistics and Data Analysis 41, 577-590.

50. Kim, W. (2007) Application of Asymptotic LRTS Results to Mixture Problems in Genetic Research. Ph.D. thesis. State University of New York at Stony Brook.

51. Laird, N. (1978) Nonparametric maximum likelihood estimation of a mixing distribution. Journal of the American Statistical Association 73, 805-811.

52. Lambert, D. (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34, 1-14.

53. Lehmann, E.L. (1975) Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day.

54. Leroux, B.G. (1992). Consistent estimation of a mixing distribution. Ann. Statist. 20, 1350-1360.

55. Leroux, B.G. and Puterman, M.L. (1992). Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. Biometrics 48, 545-558.

56. Li, C.S., Lu, J.C., and Park J.H. (1999) Multivariate Zero-inflated Poisson models and their Applications. Technometrics 41(1), 29-38.

57. Lindsay, B. (1994) Efficiency versus robustness: the case for minimum Hellinger distance and related methods. Ann. Statist. 22, 1081-1114.

58. Lo, Y., Matthysse S., Rubin D.B., and Holzman P.S. (2002). Permutation tests for detecting and estimating mixtures in task performance within groups. Statistics in Medicine 21, 1937-1953.

59. McCullagh, P.A. and Nelder, J. (1989) Generalized Linear Models, second edition. London: Chapman & Hall.

60. McLachlan, G.J. (1988) On the choice of initial values for the EM algorithm in fitting mixture models. The Statistician 37, 417-425.

61. McLachlan, G.J., and Basford, K.E. (1988). Mixture Models: Inference and Applications to Clustering. New York: Marcel Dekker.

62. McLachlan, G.J., and Peel, D. (2000). Finite Mixture Models. New York: Wiley and Sons.

63. Mendell, N.R., Thode, H.C., and Finch, S.J. (1991). The likelihood ratio test for the two component mixture problem: power and sample size analysis. Biometrics 47, 1143-1148.

64. Mendenhall, W. and Hader, R.J. (1958) Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. Biometrika 45, 504-520.

65. Meng, X.L. and van Dyk, D. (1997) The EM algorithm – an old folk song sung to a fast new tune (with discussion). Journal of the Royal Statistical Society B 59, 511-567.

66. Moser B.K., and Stevens, G.R. (1992). Homogeneity of variance in the two-sample means test. The American Statistician 46, 19-21.

67. Moser, B.K., Stevens, G.R., and Watts, C.L. (1989). The two-sample t-test versus Satterthwaite's approximate F test. Communications in Statistics – Theory and Methods 18.

68. Muench, H. (1936). Probability distribution of protection test results. Journal of the American Statistical Association 31, 677-689.

69. Muench, H. (1938). Discrete frequency distributions arising from mixtures of several single probability values. Journal of the American Statistical Association 33, 390-398.

70. Mullahy, J. (1997) Heterogeneity, Excess Zeros, and the Structure of Count Data Models. Journal of Applied Econometrics 12, 337-350.

71. Neave, H.R., and Granger, C.W.J. (1968) A Monte Carlo study comparing various two-sample tests for differences in mean. Technometrics 10, 509-522.

72. Neuhauser, M. (2001) An adaptive location-scale tests. Biometrical Journal 43, 809-819.

73. Neuhauser, M. (2002) Two-sample tests when variances are unequal. Animal Behavior 63, 823-825.

74. Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. American Journal of Mathematics 8, 343-366.

75. Ogenstad, S. (1998). The use of generalized tests in medical research. Journal of Biopharmaceutical Statistics 8, 497-508.

76. Pearson, K. (1894). Contribution to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London A 185, 71-110.

77. Pearson, K. (1915-1917). On certain types of compound frequency distributions in which the components can be individually described by binomial series. Biometrika 11, 139-144.

78. Pitman, E.J.G. (1948). Lecture notes on non-parametric statistics. Columbia University, New York.

79. Ramsey, P. (1980). Exact Type I error rates for robustness of Student's test with unequal variances. Journal of Education Statistics 5, 337-349.

80. Rao, C.R. (1948) The utilization of multiple measurements in problems of biological classification. Journal of the Royal Statistical Society B 10, 159-203.

81. Razzaghi, M. and Kodell, R.L (2000) Risk Assessment for Quantitative Responses Using a Mixture Model. Biometrics 56 (2), 519-527.

82. Razzaghi, M. and Nanthakumar, A. (1992). On using Lehmann alternatives with nonresponders. Mathematical Biosciences 80, 69-83.

83. Razzaghi, M. and Nanthakumar, A. (1994). A distribution free test for detecting a treatment effect in the presence of nonresponders. Biometrical Journal 36, 373-384.

84. Redner, R.A., and Walker, H.F. (1984). Mixture densities, maximum likelihood, and the EM algorithm. Society for Industrial and Applied Mathematics Review 26, 195-239.

85. Rider, P.R. (1961) The method of moments applied to a mixture of two exponential distributions. Ann. Math. Statist. 32, 143-147.

86. Rider, P.R. (1962) Estimating the parameters of mixed Poisson, binomial and Weibull distributions by the method of moments. Bull Internat. Statist. Inst. 39 Part 2, 225-232.

87. Salsburg, D.S. (1986). Alternative hypotheses for the effects of drugs in small-scale clinical studies. Biometrics 42, 671-674.

88. Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. Biometrics Bulletin 2, 110-114.

89. Scheffe, H. (1956). The analysis of variance. New York: Wiley.

90. Schilling, W. (1947). A frequency distribution represented as the sum of two Poisson distributions. Journal of the American Statistical Association 42, 407-424.

91. Self, S.G., and Liang, K.Y. (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of American Statistical Association 82, 605-610.

92. Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. Ann. Statist. 4, 1200-1209.

93. Simpson, D. (1987) Minimum Hellinger distance estimation for the analysis of count data. Journal of the American Statistical Association 82, 802-807.

94. Symons, M., Grimson, R., Yuan, Y. (1983). Clustering of rare events. Biometrics 39, 193-205.

95. Tan, W.Y. and Chang, W.C. (1972). Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of two normal densities. Journal of the American Statistical Association 67, 702-708.

96. Teicher, H. (1960) On the mixture of distributions. Ann. Math. Statist. 31, 55-73.

97. Thode, H.C., Finch, S.J., and Mendell, N.R. (1988). Simulated percentage points for the null distribution of the likelihood ratio test for a mixture of two normals. Biometrics 44, 1195-1201.

98. Titterington, D.M. (1996). Mixture distributions (update). In Encyclopedia of Statistical Sciences, Vol. 5, S. Kotz, N.L. Johnson, and D. Banks (Eds.). New York: Wiley, 399-407.

99. Titterington, D.M., Smith A.F.M., and Makov, U.E. (1985). Statistical Analysis of Finite Mixture Distributions. New York: Wiley & Sons.

100. Toothaker, L.E. (1972). An empirical investigation of the permutation t-test as compared to Student t-test and the Mann-Whitney U-test. Wisconsin Research and Development Center for Cognitive Learning.

101. Wang, P., Puterman, M.L., Cockburn, I., and Le, N.D. (1996). Mixed Poisson regression models with covariate dependent rates. Biometrics 52, 381-400.

102. Welch, B.L. (1938). The significance of the difference between two means when the population variances are unequal. Biometrika 29, 350-362.

103. Wells, C.S., and Hintze, J.M. (2007). Dealing with assumptions underlying statistical tests. Psychology in the Schools 44, 495-502.

104. Wilcoxon, F. (1945). Individual Comparison by Ranking Methods. Biometrics 1, 80-83.

105. Wolfe, J.H. (1965). A computer program for the computation of maximum likelihood analysis of types. Reasearch Memo. SRM 65-12. San Diego: U.S. Naval Personnel Research Activity.

106. Wolfe, J.H. (1967). NORMIX: Computational methods for estimating the parameters of multivariate normal mixtures of distributions. Research Memo. SRM 68-2. San Diego: U.S. Naval Personnel Research Activity.

107. Wolfe, J.H. (1970). Pattern clustering by multivariate mixture analysis. Multivariate Behavioral Research 5, 329-350.

108. Wolfe, J.H. (1971). A Monte Carlo study of sampling distribution of the likelihood ratio for mixtures of multinormal distributions. Technical Bulletin STB 72-2. San Diego: U.S. Naval Personnel and Training Research Laboratory.

109. Woodward, W.A., Parr, W.C., Schucany, W.R., and Lindsey, H. (1984) A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. Journal of the American Statistical Association 79, 590-598.

110. Woodward, W., Whitney, P., and Eslinger, P. (1995) Minimum Hellinger distance estimation of mixture proportions. J. Statist. Plann. Inference 48, 303-319.

111. Xiang, L., Yau, K.K.W., Lee, A.H., and Fung, W.K. (2005) Influence diagnostics for two-component Poisson mixture regression models: applications in public health. Statistics in Medicine 24, 3053-3071.

112. Xie, M., He, B., and Goh, T.N. (2001) Zero-inflated Poisson model in statistical process control. Computational Statistics and Data Analysis 38, 191-201.

113. Zimmerman, D.W. (1996). Some properties of preliminary tests of equality of variances in the two-sample location problem. The Journal of General Psychology 123, 217-231.

114. Zimmerman, D.W. (2004). A note on preliminary tests of equality of variances. British Journal of Mathematical and Statistical Psychology 57, 173-181.

115. Zimmerman, D.W. and Zumbo, B.N. (1993). Rank transformations and the power of the Student t-test and Welch t`-test for non-normal populations. Can J Exp Psychology 47, 523-539.