PURE LIKELIHOOD-BASED METHODS FOR GENETIC ASSOCIATION STUDIES

by

Weili Li

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Biostatistics, Dalla Lana School of Public Health
University of Toronto

© Copyright 2016 by Weili Li
Abstract

Pure Likelihood-based Methods for Genetic Association Studies

Weili Li
Doctor of Philosophy
Graduate Department of Biostatistics, Dalla Lana School of Public Health
University of Toronto
2016

In this thesis, we contribute to the growing literature on the Evidential methodology for genetic association studies. The Evidential Paradigm (EP) is an alternative statistical framework to the Frequentist and Bayesian paradigms for statistical inference, which uses likelihood ratios to measure the strength of statistical evidence. The EP has favorable operational characteristics, with small and bounded error probabilities, and the framework has been applied to genetic linkage and association studies. We discuss some of the well-recognized issues associated with the Frequentist framework in assessing association evidence in genetic association studies, and we develop novel approaches under the EP to address these issues. We first address the issue of prioritizing individual rare sequence variants within an associated region in case-control association studies. Rare variants are commonly ranked by Fisher’s exact p-values, however, depending on whether one- or two-sided p-values are used, the same set of rare variants can be ranked differently. Motivated by the EP, we propose a conditional likelihood ratio-based measure, maxLRc, to rank rare variants. We show analytically that the maxLRc is based on the same underlying model as Fisher’s exact test, and it is always well-defined even when data is under separation. Through simulations, we show that the maxLRc outperforms the commonly used two-sided Fisher’s exact p-values in most simulation settings. We then address another issue in large-scale genetic association studies: under the Frequentist paradigm, any small effect size could become statistically significant with large enough sample size. We discuss an EP solution to this problem which extends the EP to accommodate composite hypotheses (EP^C) and uses the generalized likelihood ratios.
(GLRs) to measure statistical evidence. To provide theoretical justification for the extension, we analytically derive the operational characteristics for the EP\textsuperscript{C}. We show that the GLRs have small and bounded error probabilities for the Normal distribution and fixed-dimensional parametric models asymptotically. We show that the data can support either the null or the alternative hypothesis under the EP\textsuperscript{C}, allowing association results to be more clearly demarcated; and the use of composite hypotheses can guard against statistically significant but practically insignificant findings in large-scale association studies.
Dedication

This thesis is dedicated to the loving memory of my grandparents,

Shulin Li and Guilan Long.

How I wish you were here to share every joyful moment of my life.
Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor, Dr. Lisa Strug. Without her motivation and encouragement, I would not have considered pursuing a Ph.D. degree. I appreciate her knowledge and expertise in helping me frame my thesis, and giving me the perfect balance of guidance and independence in developing and completing my thesis. Her understanding, patience and encouragement pulled me out of self-doubt and desperation during the tough times in my Ph.D. pursuit. I am truly grateful for having her as my academic mentor and an inspiring role model in my life.

I am sincerely grateful to my advisory committee members, Professors Lei Sun, Jerry Lawless and Paul Corey for providing insightful comments and advice on my thesis. My sincere thanks also go to my internal examiner, Professor Michael Evans, and external examiner, Professor Subhash Lele, for their time and challenging questions.

I would also like to thank my parents for their unconditional love, encouragement and support. A special thanks to my Mom, who always has full confidence in me and gives her full support for everything I do.

My thanks also go to my wonderful friends, fellow Ph.D. students and colleagues in the Strug Lab for listening to me vent my stress and cheering me up when I was frustrated.

Lastly, a heartfelt thanks to my loving husband, Ye Li, who rode the Ph.D. emotional rollercoaster with me. I am so grateful for having him by my side, bringing me love and joy and comforting me when I was in stress and frustration. This thesis would not have been possible without his love, support and patience.
Contents

1 Introduction 1
  1.1 Background for genetic association studies . . . . . . . . . . . . . . . . . . 1
  1.2 Limitations of the Frequentist Paradigm in genetic association studies . 2
  1.3 Introduction to the Evidential Paradigm . . . . . . . . . . . . . . . . . . 5
    1.3.1 The operational characteristics of the Evidential Paradigm . . . 5
    1.3.2 Upper bound for the probability of observing misleading evidence 8
    1.3.3 Comparison of the operational characteristics between the Eviden-
        tial Paradigm and the Neyman-Pearson procedure . . . . . . . . . 11
  1.4 Extension of the Evidential Paradigm to accommodate composite hypotheses 12
  1.5 Outline of thesis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

2 Maximized Conditional Likelihood Ratio 17
  2.1 Introducing the maximum conditional likelihood ratio . . . . . . . . . . . . 17
  2.2 Rare variant prioritization performance comparison based on simulated data 22
    2.2.1 Fisher's two-sided exact p-value . . . . . . . . . . . . . . . . . . . 22
    2.2.2 Details of rare variant simulation . . . . . . . . . . . . . . . . . . 23
    2.2.3 Prioritization performance comparison . . . . . . . . . . . . . . . 24
  2.3 Prioritizing rare variants in a study of Rolandic Epilepsy . . . . . . . . . 30
  2.4 Conclusion and additional considerations . . . . . . . . . . . . . . . . . . 31

3 Generalized Likelihood Ratio for Composite Hypotheses 34
  3.1 Complementary composite hypotheses and the associated operational char-
       acteristics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
  3.2 Theorem 1: Normal distribution . . . . . . . . . . . . . . . . . . . . . . . . 37
  3.3 Theorem II: Asymptotic results for one-parameter parametric models . . 46
  3.4 Theorem III: Asymptotic results for multi-parameter parametric models . 51
  3.5 Implications for study planning . . . . . . . . . . . . . . . . . . . . . . . 55
    3.5.1 Sample size calculation for the Normal probability model . . . . 56
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6 Discussion and future work</td>
<td>60</td>
</tr>
<tr>
<td>4 Application of the GLR in Genetic Association Studies</td>
<td>62</td>
</tr>
<tr>
<td>4.1 Sample size calculation for a GWAS with a quantitative trait</td>
<td>63</td>
</tr>
<tr>
<td>4.2 Strength of the GLR in genetic association studies: a simulation study</td>
<td>68</td>
</tr>
<tr>
<td>4.2.1 Simulation design</td>
<td>69</td>
</tr>
<tr>
<td>4.2.2 Simulation results</td>
<td>70</td>
</tr>
<tr>
<td>4.3 Discussion and future work</td>
<td>72</td>
</tr>
<tr>
<td>5 Concluding Remarks &amp; Directions for Future Work</td>
<td>76</td>
</tr>
<tr>
<td>5.1 Multiple parameters of interest</td>
<td>77</td>
</tr>
<tr>
<td>5.2 Model misspecification</td>
<td>80</td>
</tr>
<tr>
<td>Bibliography</td>
<td>82</td>
</tr>
<tr>
<td>Appendix</td>
<td>87</td>
</tr>
<tr>
<td>A</td>
<td>88</td>
</tr>
<tr>
<td>A.1 Derivation of the operational characteristics for the Normal probability model</td>
<td>88</td>
</tr>
<tr>
<td>A.2 Derivation of the operational characteristics for one-parameter parametric models</td>
<td>100</td>
</tr>
<tr>
<td>A.3 Derivation of the operational characteristics for multi-parameter parametric models</td>
<td>113</td>
</tr>
<tr>
<td>A.3.1 In the presence of multiple nuisance parameters</td>
<td>120</td>
</tr>
</tbody>
</table>
List of Tables

2.1 An example of a 2x2 contingency table classified by disease status and minor allele carrier status. .......................................................... 18
2.2 Examples of 2x2 tables under complete or quasi-complete separation. . 20
2.3 Mean Kendall’s correlation coefficients, across 1000 repetitions, for rare variants generated with MAF ∈ [0.005,0.05], number of truly associated variants Q = 10, and an equal number of cases and controls. .................. 25
2.4 Mean Kendall’s correlation coefficients, across 1000 repetitions, for rare variants generated with MAF ∈ [0.005,0.05], number of truly associated variants Q = 20, and an equal number of cases and controls .............. 25
2.5 Sum of median number of truly associated variants (NT) over values of K from 1 to 50 (SumK) for rare variants generated with MAF ∈ [0.005,0.05], number of truly associated variants Q = 10, and an equal number of cases and controls. Sum of median proportion of truly associated variants (NT/K) is provided in brackets. ................................. 28
2.6 Sum of median number of truly associated variants (NT) over values of K from 1 to 50 (SumK) for rare variants generated with MAF ∈ [0.005,0.05], number of truly associated variants Q = 20, and an equal number of cases and controls. Sum of median proportion of truly associated variants (NT/K) is provided in brackets. ................................. 29
2.7 Mean of the average rankings of the collection of truly associated vari- ants, across 1000 repetitions, for rare variants generated with MAF ∈ [0.005,0.05], number of truly associated variants Q = 10, and an equal number of cases and controls. ................................. 29
2.8 Mean of the average rankings of the collection of truly associated vari- ants, across 1000 repetitions, for rare variants generated with MAF ∈ [0.005,0.05], number of truly associated variants Q = 20, and an equal number of cases and controls. ................................. 29
2.9 Prioritization of rare sequence variants from 27 Rolandic Epilepsy cases and 200 Colorectal Cancer controls; 10 rare variants top ranked by the maxLRc are shown, together with their corresponding rankings assigned by 2-sided Fisher’s exact p-value.
## List of Figures

1.1 The probability of observing strong misleading evidence for the Normal model \((M_0 = M_1 = M)\) is plotted as a function of the standardized distance between the two simple hypotheses (Royall (2000), P763).  

1.2 Probabilities of misleading and weak evidence as functions of sample size for the Normal probability distribution (Royall (2000), P764).  

1.3 Approximate likelihood functions for the difference in treatment response rate (chemotherapy - radiation therapy), with dotted red lines separate regions of inferiority, non-inferiority and superiority (reproduction of Figure 1 in Zhang and Zhang (2013) with added details).  

2.1 NT plots for rare variants simulated with \(\text{MAF} \in [0.005, 0.05]\), number of truly associated variants \(Q = 10\), and randomly assigned directions of effect. For all sample size configurations, the maxLRc uniformly selects more or an equal number of truly associated variants than Fisher’s 2-sided exact p-value across all values of \(K\).  

2.2 NT plots for rare variants simulated with \(\text{MAF} \in [0.005, 0.05]\), number of truly associated variants \(Q = 10\), and with 80% of the casual variants having deleterious effect. The maxLRc selects more or an equal number of truly associated variants than Fisher’s 2-sided exact p-value in most cases.  

2.3 NT plots for rare variants simulated with \(\text{MAF} \in [0.005, 0.05]\), number of truly associated variants \(Q = 20\), and with randomly assigned directions of effect. The maxLRc selects more or an equal number of truly associated variants than Fisher’s 2-sided exact p-value across all values of \(K\) under almost all sample size configurations.  

2.4 NT plots for rare variants simulated with \(\text{MAF} \in [0.005, 0.05]\), number of truly associated variants \(Q = 20\), and 80% of the truly associated variants having deleterious effects. The maxLRc selects more or an equal number of truly associated variants than Fisher’s 2-sided exact p-value in most cases.
3.1 The probability of misleading evidence, $M_0$, is plotted as a function of sample size ................................................................. 39
3.2 The probabilities of observing misleading, weak and strong evidence, $M_0, W_0$ and $S_0$, are plotted as functions of sample size .................. 41
3.3 The probability of misleading evidence, $M_1$, is plotted as a function of sample size ................................................................. 43
3.4 The probabilities of observing misleading, weak and strong evidence, $M_1, W_1$ and $S_1$, are plotted as functions of sample size .................. 44
3.5 $M_0, W_0$ and $S_0$ are plotted as functions of $C$ .......................... 45
3.6 $M_1, W_1$ and $S_1$ are plotted as functions of $C$ .......................... 46
3.7 $M_1, W_1$ and $S_1$ are plotted as functions of $\theta_1$ ....................... 47
3.8 The operational characteristics are plotted as functions of the alternative simple hypothesis, $\theta_1$, for sample size of $n = 10$. .................... 50
3.9 The operational characteristics are plotted as functions of the alternative simple hypothesis, $\theta_1$, for sample size of $n = 50$. .................... 51
3.10 The operational characteristics are plotted as functions of the alternative simple hypothesis, $\theta_1$, for sample size of $n = 100$. ................... 52
3.11 The probabilities of observing strong evidence in support of the simple alternative hypothesis of $\theta = \theta_1$ or the composite alternative that $\theta \in \Theta_1 = (-\infty, \theta_0 - C) \cup (\theta_0 + C, \infty)$ are plotted as functions of the sample size $n$. .................................................. 57
3.12 The probabilities of observing strong evidence, $S_0^C, S_1^C$ and $S_0^S = S_1^S$ are plotted against sample size $n$. .......................... 59

4.1 The operational characteristics are plotted against MAF .................... 67
4.2 The operational characteristics are plotted against $C$ ...................... 68
4.3 The operational characteristics are plotted against sample size .......... 69
4.4 Association results for the GLR and the likelihood ratio test with $n = 4000$ 71
4.5 Association results for the GLR and the likelihood ratio test with $n = 8000$ 73
4.6 Association results for the GLR and the likelihood ratio test with $n = 10000$ 73
Chapter 1

Introduction

1.1 Background for genetic association studies

Statistical genetics is the scientific field that focuses on developing statistical methods for understanding the genetic basis of diseases. The human genome consists of more than 3 billion base-pairs, which are the building blocks of the DNA double helix. Each base pair is formed from two complementary nucleotides. There are many types of genetic variation, and single nucleotide polymorphisms, or SNPs, are the most common type of genetic variation in the human genome. SNPs are locations across the genome where the nucleotides can be different between individuals. Consider a bi-allelic SNP with alleles a and A, the allele that is less frequently observed in the general population (with allele frequency \( \alpha < 0.5 \)) at this particular SNP is referred to as the minor allele, and its allele frequency is called the minor allele frequency (MAF). A SNP can be further classified as a common or a rare variant based on its MAF, and a SNP is called a rare variant if its MAF is less than 5%. The two alleles at a particular SNP determine its genotype, with two alleles A and a in the general population, the genotype for an individual could be AA, Aa or aa.

Genetic association studies test for correlation between a phenotype of interest and a genetic variant (such as a SNP) to identify candidate genes or genomic regions that contribute to the quantitative phenotype or disease trait. The popular genome-wide association studies (GWASs) are powerful tools for studying common genetic variants, where one examines each of the millions of genetic markers across the genome to detect association with the phenotype of interest. The phenotype under study can be a binary or categorical disease status, a quantitative trait or longitudinal measurements, etc. Depending on the assumed genetic effect of a SNP, the phenotypic values are compared
across different groups defined by the genotypes, or alleles or other genotypic configurations. For example, with a quantitative phenotype, a genotypic model compares the average quantitative trait values among the three genotype groups, AA, Aa and aa. If a dominant model for the A allele is assumed, we then compare the average trait values between individuals having the genotype AA or Aa to those with genotype aa.

1.2 Limitations of the Frequentist Paradigm in genetic association studies

The commonly used Frequentist approach to test for association between a genetic variant and a phenotype of interest is to compute a p-value for the null hypothesis of no association. Despite their widespread use, p-values have many conceptual limitations (Sellke et al. (2001); Sterne and Smith (2001); Ioannidis (2008)). Here, we discuss three of the well-recognized issues of the Frequentist framework in assessing evidence for association in genetic association studies.

The substantial drops in genotyping cost and the improved multi-study collaborations lead to huge sample sizes in GWASs. For example, in the recent GWAS conducted by the GIANT consortium (Wood et al., 2014), more than 250,000 samples were included in a study investigating genetic markers that contribute to human height. This study highlighted one well-recognized problem with the commonly used Frequentist Paradigm for assessing association: any effect size can become statistically significant with large enough sample size. This implies that with the ever-increasing sample sizes in GWASs, many markers with minimal and practically insignificant effect sizes could be detected as statistically significant, which further complicates prioritization of genetic markers for costly follow-up biological experiments. Another limitation of the Frequentist approach for genetic association analysis lies in its focus on rejecting the simple null hypothesis of ‘no association’. In a typical Frequentist analysis of association, the null hypothesis of ‘no association’ is always represented by a simple hypothesis value, for example, odds ratio $=1$, while the alternative hypothesis of ‘association’ is represented by a composite hypothesis, e.g. odds ratio $\neq 1$. Due to the asymmetric nature of the two hypotheses, a simple hypothesis versus its entire complement, one can never observe evidence in support of the null hypothesis, regardless of sample size. This usually leads to a genome of ‘weak’ or ‘inconclusive’ results, compared with a few makers that reach a genome-wide
significant threshold of association, as one often sees in a typical GWAS.

For rare genetic variants, it is well appreciated that the single marker analysis approach, which has been successful in identifying common disease-associated variants in GWASs, has limited power to detect association with rare variants (MAF $\leq 5\%$). To increase statistical power, various methods have been proposed to combine genetic information across multiple rare variants within a gene or region of interest (Li and Leal, 2009; Madsen and Browning, 2009; Morris and Zeggini, 2010; Price et al., 2010; Neale et al., 2011; Wu et al., 2011; Lee et al., 2012; Turkmen and Lin, 2014; Li and Leal, 2008). While these combined methods have been shown to be effective as a first step in detecting associated regions, the ultimate goal is localizing individual causal variants within the identified regions that directly affect disease presentation. There is limited statistical literature addressing the difficult task of fine mapping amongst rare variants within previously detected associated genes or regions. Ionita-Laza et al. (Ionita-Laza et al., 2012) proposed a sliding-window approach to scan genes or regions of interest to further detect causal rare-variant-enriched sub-regions. A similar sliding-window-based method was also published by Brisbin et al. (Brisbin et al., 2012). However, as with initial identification of associated regions of interest, fine mapped sub-regions may still include a large number of individual rare variants that require further refinement. For example, in Ionita-Laza et al. (Ionita-Laza et al., 2012), the authors applied one of the combined methods, the variable threshold approach (Price et al., 2010), in autism spectrum disorder and identified two significantly associated genes, one of which spans a 235 kb region on chromosome 2. Even with further fine mapping, the significantly associated sub-region covers a 26kb region. Therefore, in order to select the most promising variants for follow-up biological experiments, further prioritization on the individual variant level is necessary and statistical information can assist with this additional refinement.

The most commonly implemented approach to prioritizing individual variants in case-control studies is to rank them based on p-values, despite known limitations (Fisher, 1934; Cornfield, 1966; Mantel, 1977; Lindley and Scott, 1984; Royall, 1997). For rare variants specifically, Fisher’s exact p-values are often computed since the low MAF can lead to sparse and skewed data. However, the computation of Fisher’s exact p-values requires the investigator to choose whether the p-values will be one-sided or two-sided, and which two-sided method to implement. The p-value rankings for the same set of rare variants can differ substantially across these choices, due to the asymmetry of the discrete hypergeometric distribution underlying Fisher’s exact test. Therefore, two investigators
with different beliefs or knowledge of the direction of association may select different rare variants for follow-up based on statistical considerations, a problem that we encountered when deciding how to prioritize rare sequence variants from a linkage region in a study of Rolandic Epilepsy (Strug et al., 2009). Although prior information such as expert opinions, biological knowledge or findings from previous studies may improve prioritization, results heavily depend on what type of prior information is used and how this information is incorporated into the prioritization. Therefore, a prioritization method based solely on the observed data and independent of a-priori beliefs on the direction of association could provide a common starting point from which to work.

Another problem that the Frequentist framework does not offer a favorable solution to is the detection of pleiotropy. Pleiotropy is observed when one genetic locus affects multiple phenotypes. Although conceptually straightforward, statistically defining and detecting pleiotropy is challenging. For example, with two phenotypes of interest A and B, the null hypothesis of 'no pleiotropy' are represented by $H_0 : \beta_A = 0, \beta_B = 0$, where $\beta_A$ and $\beta_B$ represent the association of phenotype A and B, respectively, with a genetic marker. For one of my thesis applied projects, we set out to identify modifier genes in Cystic Fibrosis (CF) that are associated with multiple CF-related traits (Li et al., 2014). Four phenotypes, including Meconium Ileus, a quantitative phenotype measuring pediatric lung disease severity, age at first acquisition of Pseudomonas aeruginosa and prenatal pancreatic damage measured by immunoreactive trypsinogen were collected on the same set of CF patients. In this pleiotropy analysis, we combined the conditional analysis principle with the reverse regression approach (O’Reilly et al., 2012). Specifically, we reverse the role of phenotype and genotype so that the number of meconium ileus risk alleles of a SNP is regressed on meconium ileus and pancreatic disease severity at birth via ordinal logistic regression. The regression coefficient corresponding to pancreatic disease severity in this multivariate regression model captures the association between pancreatic disease severity and the SNP, after accounting for the correlation between pancreatic disease severity and meconium ileus, and association between the SNP and meconium ileus. Therefore, testing the significance of the pancreatic disease severity regression coefficient assesses the pleiotropic effect of the SNP. Although this reverse regression approach worked well for our pleiotropy project, we had to restrict the analysis to previously established meconium ileus susceptibility loci, and a generally applicable method is more desirable.

In this thesis, we develop novel statistical methods under an alternative pure likelihood-
based framework, the Evidential Paradigm, to address some of the challenges in genetic association studies, focusing on those mentioned above.

1.3 Introduction to the Evidential Paradigm

The Evidential Paradigm (EP) (Royall, 1997; Blume, 2002; Bickel, 2012; Zhang and Zhang, 2013) is an alternative to Frequentist and Bayesian paradigms for statistical inference. Theoretically grounded on the Law of Likelihood, the EP uses likelihood ratios to measure the strength of statistical evidence.

Law of Likelihood

If hypothesis A implies that the probability that a random variable X takes the value x is \( p_A(x) \), while hypothesis B implies that the probability is \( p_B(x) \), then the observation \( X=x \) is evidence supporting A over B if and only if \( p_A(x) > p_B(x) \), and the likelihood ratio, \( \frac{p_A(x)}{p_B(x)} \), measures the strength of that evidence (Hacking, 1965; Royall, 1997).

\[ P_A(x) = P(x|A) \] is the probability of observing x assuming hypothesis A is true, and \( P_B(x) = P(x|B) \) is the probability of observing x assuming hypothesis B is true. Therefore, the Law of Likelihood indicates that the hypothesis that more accurately predicts the observed data is better supported by the data, and the strength of that support, or the strength of statistical evidence, is measured by the likelihood ratio. Note that the Law of Likelihood requires the likelihood ratio to be calculated at two simple hypothesis values for the parameter of interest. That is, the probability or the likelihood of the data is completely specified by the hypothesis value. The Law of Likelihood does not provide guidance on how to measure statistical evidence when the hypotheses are not simple, for example, when the two hypotheses are both composite.

1.3.1 The operational characteristics of the Evidential Paradigm

Under the EP, statistical evidence and uncertainty are decoupled, where the size of the likelihood ratio indicates the strength of statistical evidence and the uncertainties are measured by error probabilities or the operational characteristics.
Let $\theta$ denote the single parameter of interest, with two simple hypotheses, $H_1 : \theta = \theta_0$ or $H_2 : \theta = \theta_1$. In addition, let $L(\theta_0) = L(\theta_0; x)$ and $L(\theta_1) = L(\theta_1; x)$ denote the likelihoods evaluated at $\theta_0$ and $\theta_1$, respectively. Under the EP, the likelihood ratio, $LR = \frac{L(\theta_0)}{L(\theta_1)}$, measures the strength of statistical evidence in favor of $\theta = \theta_0$ over $\theta = \theta_1$. According to the Law of Likelihood, if $LR > 1$, the data provides evidence in support of $H_1$ over $H_2$; if $LR < 1$, the data provides evidence in support of $H_2$ over $H_1$. The evidence is neutral between $H_1$ and $H_2$ if $LR = 1$.

In an experiment, one aims to find strong evidence in support of one hypothesis over another, and strong evidence is represented by a likelihood ratio of at least some threshold value of $k > 1$. That is, the observed data provide

- strong evidence in support of $\theta_0$ over $\theta_1$, if $\frac{L(\theta_0)}{L(\theta_1)} \geq k$
- strong evidence in support of $\theta_1$ over $\theta_0$, if $\frac{L(\theta_1)}{L(\theta_0)} \geq k$
- weak evidence, if $\frac{1}{k} < \frac{L(\theta_0)}{L(\theta_1)} < k$

The threshold value $k$ defines the size of the likelihood ratio at which one would consider it as strong evidence in support of one hypothesis over another. Choosing the appropriate value of $k$ is analogous to choosing the appropriate significance level in hypothesis testing. Although no single choice will work for all problems, benchmarks of $k = 8$ and 32 have been recommended (Edwards, 1984; Royall, 1997) as representing 'fairly strong' and 'strong' evidence, with $k = 32$ most commonly used in the literature (Blume, 2002; Royall, 1997). A much larger $k$ is often used in large-scale genetic studies, for example, $k = 1000$ is recommended and used in genome-wide linkage studies (Strug and Hodge, 2006b). Royall (1997) provided a simple and intuitive example to promote a quantitative understanding of the likelihood ratios. Assume there are two identical urns, one containing only white balls and the other containing an equal number of white and black balls. One urn is randomly chosen and we draw a succession of balls from it, and after each draw, returning the ball back to the urn and thoroughly mixing the contents. If three balls are drawn and all are white, then the experiment provides 'fairly strong' evidence that the urn with all white balls was chosen instead of the urn with half white and half black balls. Let us formally define the two simple hypotheses, $H_1 : \theta = 1$ and $H_2 : \theta = 0.5$, where $\theta$ denotes the proportion of white balls in the selected urn, then $L(1)/L(0.5) = 1/(0.5)^3 = 8$. Similarly, if five white balls were drawn successively, the likelihood ratio is $1/(0.5)^5 = 32$, which indicates even stronger evidence in support of $H_1$. 

Chapter 1. Introduction
versus $H_2$.

Since study designs are evaluated by their operational characteristics, it is important to have a body of literature that evaluates the operational characteristics of measuring statistical evidence directly from the likelihood ratios. Before defining the operational characteristics associated with the EP, we first discuss the three types of evidence that can be observed under the EP:

Misleading evidence - when the observations provide strong evidence in support of the incorrect hypothesis, i.e., $\frac{L(\theta_1)}{L(\theta_0)} \geq k$, where $\theta_0$ is the correct hypothesis.

Weak evidence - when the observations do not provide strong evidence in support of either hypothesis, i.e., $\frac{1}{k} < \frac{L(\theta_1)}{L(\theta_0)} < k$.

Strong evidence - when the observations provide strong evidence in support of the correct hypothesis, i.e., $\frac{L(\theta_1)}{L(\theta_0)} \leq k$, where $\theta_1$ is the correct hypothesis.

The operational characteristics for the EP are simply defined as the probabilities of observing misleading (M), weak (W) and strong (S) evidence.

When $\theta_0$ is assumed to be true:

\[
M_0 = P_0\left(\frac{L(\theta_1; x)}{L(\theta_0; x)} \geq k\right) = P_0\left(\frac{L(\theta_0; x)}{L(\theta_1; x)} \leq \frac{1}{k}\right)
\]

\[
W_0 = P_0\left(\frac{1}{k} < \frac{L(\theta_0; x)}{L(\theta_1; x)} < k\right) = P_0\left(\frac{1}{k} < \frac{L(\theta_1; x)}{L(\theta_0; x)} < k\right)
\]

\[
S_0 = P_0\left(\frac{L(\theta_0; x)}{L(\theta_1; x)} \geq k\right) = P_0\left(\frac{L(\theta_1; x)}{L(\theta_0; x)} \leq \frac{1}{k}\right)
\]

and similarly, when $\theta_1$ is assumed true:

\[
M_1 = P_1\left(\frac{L(\theta_0; x)}{L(\theta_1; x)} \geq k\right) = P_1\left(\frac{L(\theta_1; x)}{L(\theta_0; x)} \leq \frac{1}{k}\right)
\]

\[
W_1 = P_1\left(\frac{1}{k} < \frac{L(\theta_1; x)}{L(\theta_0; x)} < k\right) = P_1\left(\frac{1}{k} < \frac{L(\theta_0; x)}{L(\theta_1; x)} < k\right)
\]

\[
S_1 = P_1\left(\frac{L(\theta_1; x)}{L(\theta_0; x)} \geq k\right) = P_1\left(\frac{L(\theta_0; x)}{L(\theta_1; x)} \leq \frac{1}{k}\right)
\]
**Example 1.** Probability of misleading, weak and strong evidence: Normal distribution mean (Royall (1997), P763 - 764)

Let \( X_1, X_2, ..., X_n \) denote a set of \( n \) i.i.d random variables from the \( N(\theta, \sigma^2) \) distribution, where \( \theta \) is the unknown parameter of interest and \( \sigma^2 \) is assumed known. With two simple hypotheses \( H_1 : \theta = \theta_0 \) and \( H_2 : \theta = \theta_1 \), then the likelihood ratio measuring the support for \( H_2 \) over \( H_1 \) is given by:

\[
\frac{L(\theta_1)}{L(\theta_0)} = \exp \left\{ \frac{n(\theta_1 - \theta_0)}{\sigma^2} \left( \bar{x} - \frac{\theta_0 + \theta_1}{2} \right) \right\}
\]

Assuming \( \theta_0 \) is true, the probability of observing misleading evidence in support of the incorrect hypothesis (\( M_0 \)), the probability of observing weak evidence (\( W_0 \)) and the probability of observing strong evidence in support of the correct hypothesis (\( S_0 \)), are given by the following functional expressions:

\[
M_0 = P_0 \left( \frac{L(\theta_1)}{L(\theta_0)} \geq k \right) = \Phi \left(-\frac{\sqrt{n}\Delta}{2\sigma} - \frac{\sigma \ln k}{\sqrt{n}\Delta} \right)
\]

\[
W_0 = P_0 \left( \frac{1}{k} < \frac{L(\theta_1)}{L(\theta_0)} < k \right) = \Phi \left(-\frac{\sqrt{n}\Delta}{2\sigma} + \frac{\sigma \ln k}{\sqrt{n}\Delta} \right) - \Phi \left(-\frac{\sqrt{n}\Delta}{2\sigma} - \frac{\sigma \ln k}{\sqrt{n}\Delta} \right)
\]

\[
S_0 = P_0 \left( \frac{L(\theta_0)}{L(\theta_1)} \geq k \right) = 1 - M_0 - W_0 = \Phi \left(\frac{\sqrt{n}\Delta}{2\sigma} - \frac{\sigma \ln k}{\sqrt{n}\Delta} \right)
\]

Where \( \Phi \) is the standard Normal cumulative density function and \( \Delta = |\theta_1 - \theta_0| \) is the distance between the two simple hypothesis values. For Normal distribution, it is easy to show that \( M_1 = M_0, W_1 = W_0 \) and \( S_1 = S_0 \).

1.3.2 **Upper bound for the probability of observing misleading evidence**

It is obvious that both weak and misleading evidence are undesirable. When weak evidence is observed, the study remains inconclusive. We know when we observe weak evidence, as the data would generate a likelihood ratio between \( \frac{1}{k} \) and \( k \). However, when strong evidence is observed, it is unknown if the data is providing evidence in support of the correct hypothesis or it is misleading. Therefore, it is crucial to understand the probabilistic controls for the processes that generate statistical evidence.
1. Universal bound
For one-parameter parametric models, it has been shown (Birnbaum, 1962; Smith, 1953) that

\[ P_0 \left( \frac{f_1}{f_0} \geq k \right) \leq \frac{1}{k} \]

This result implies that the probability of observing misleading evidence can not exceed \( \frac{1}{k} \), which is referred to as the universal bound by Royall (2000). For example, when \( k = 8 \), the probability of observing strong misleading evidence cannot exceed \( 1/8 = 0.125 \).

2. Upper bound for the Normal distribution
When the parameter of interest is the mean of a normal distribution, with two simple hypotheses \( \mu = \mu_0 \) and \( \mu = \mu_1 \), Royall (2000) showed that \( M_0 = M_1 = M = \Phi(- \sqrt{\frac{n|\Delta|}{2\sigma}} - \frac{\sigma \ln \sqrt{n}}{|\Delta|\sqrt{n}}) \), where \( \Delta = |\mu_1 - \mu_0| \) is the distance between the two simple hypotheses, and \( \sigma \) is the standard deviation which is assumed known. The graph of \( M \) as a function of the standardized distance between the two hypotheses is provided in Figure 1.1 for \( k = 8 \). The probability distribution for \( M \) is referred to as the 'bump function' by Royall due to its graphic shape. Royall (2000) further showed that

\[ M \leq \Phi(- \sqrt{2\ln k}) \]

and the maximum value of \( M \) is observed at \( \Delta = \sqrt{2\ln k} \) standard errors. When \( k = 8 \), \( M \leq \Phi(- \sqrt{2\ln 8}) = 0.021 \), which is much smaller than the universal bound of 0.125. It is evident from Figure 1.1 that as \( \Delta \) increases, the probability of observing misleading evidence increases from 0 to its maximum value of 0.021, and then decreases down to 0 as \( \Delta \) further increases beyond \( \sqrt{2\ln k} \) standard errors.

3. Large-sample bound for one-parameter parametric models
Suppose now \( X_1, X_2, ..., X_n \) are i.i.d with probability distribution \( f(\cdot; \theta) \). With large sample size \( n \) and assuming \( f \) is a smooth function of the real-valued parameter \( \theta \), it was shown (Royall, 2000) that the probability of misleading evidence, when \( \theta_0 \) is true, is given by

\[ P_0 \left( \frac{f_1}{f_0} \geq k \right) \to \Phi\left(- \frac{c}{2} - \frac{\ln k}{c} \right) \]
Figure 1.1: The probability of observing strong misleading evidence for the Normal model ($M_0 = M_1 = M$) is plotted as a function of the standardized distance between the two simple hypotheses (Royall (2000), P763).

where $c = |\theta_1 - \theta_0|/\sqrt{nI(\theta_0)}$ and $I(\theta) = -E\{[\partial^2 \ln f(X; \theta)]/\partial \theta^2\}$ represents Fisher’s information. In large samples, the maximum probability of misleading evidence is approximately $\Phi(-\sqrt{2 \ln k})$, the same as for the Normal distribution case, and this maximum is achieved at $\theta_1 = \theta_0 \pm \sqrt{2 \ln k/nI(\theta_0)}$.

In practice, one rarely works with one-parameter parametric models. For example, in a genetic association analysis with a quantitative phenotype, we fit a linear regression model:

$$\text{phenotype} = \beta_0 + \beta_1 \text{SNP} + \beta_2 \text{age} + \beta_3 \text{gender} + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$ represents the error term. Although the parameter of interest here is $\beta_1$, the effect size for the SNP-phenotype association, there are multiple parameters in this model. The additional parameters, $\beta_0, \beta_2$ and $\beta_3$, are referred to as nuisance parameters, and they must be eliminated from the model in order to make inference on the parameter of interest alone from the likelihood ratio. There is no single solution to eliminating nuisance parameters. A conditional or marginal likelihood is preferred as they are true likelihood functions. When they are not available, the use of a profile likelihood offers a more generally available solution, and Royall (2000) showed that the same upper bound of $\Phi(-\sqrt{2 \ln k})$ for $M$ holds asymptotically for profile likelihoods. Moreover, if an orthogonalizing transformation of the parameters exists so that an orthogonal likelihood for the parameter of interest exists, then the profile likelihood is that orthogonal likeli-
The above results suggest that the use of likelihood ratios as a measure of statistical evidence is operationally valid and reliable, in that, (1) we do not observe misleading strong evidence with high probability, and (2) as sample size increases, the probability of observing misleading evidence decreases down to 0.

1.3.3 Comparison of the operational characteristics between the Evidential Paradigm and the Neyman-Pearson procedure

With two simple hypotheses $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$, the probability of misleading evidence, $M_0$, is analogous to the type I error rate, $\alpha$, in the sense that they both represent the probability of rejecting $H_0$ when it is true. The sum, $M_1 + W_1$, is the probability of failing to observe strong evidence in support of $H_1$ when $H_1$ is actually true, therefore, it is analogous to the type II error rate, $\beta$. There are, however, fundamental differences between these quantities. Under Neyman-Pearson, $\alpha$ is fixed at a pre-specified value, usually at 5%, and therefore, stays constant across all sample sizes. The probability of misleading evidence, on the other hand, is not fixed and it varies with sample size. Even though the probability of misleading evidence is not fixed, we know that it is naturally bounded and small, as shown previously in section 1.3.2.

In Figure 1.2, M and W for the Normal distribution case with $\Delta = \sigma$ and $k = 8$ are plotted as functions of the sample size $n$. For comparison, the type I and type II error rates corresponding to the most powerful test with $\alpha = 0.05$ are also plotted. As discussed previously, M is not fixed and it varies with sample size, however, it is bounded by its maximum of 0.021, which is much smaller than 0.05.

It is worth emphasizing that the operational characteristics are only important at the study planning stage. When planning a study, one wants to have a high chance of observing strong evidence in support of the correct hypothesis (i.e. high values of $S_1$ and/or $S_2$), which is equivalent to ensuring the probabilities of observing misleading and weak evidence are low. Once the data is collected, these probabilities become irrelevant, and only the likelihood ratio is examined as a measure of statistical evidence provided by a given body of data.
The EP framework has been applied to various studies including clinical trials (Blume, 2002; Wang and Blume, 2011; Zhang and Zhang, 2013) and genetic studies. In particular, there is a growing literature on EP methodologies and applications in genetic studies. For example, Strug and Hodge (2006a,b) derived the operational characteristics and provided guidelines for performing genetic linkage studies. Hodge et al. (2011) further discussed an EP approach for conducting multi-point linkage analysis. Strug et al. (2010) detailed sample size estimation, multiple-testing adjustment and results visualization in genetic association studies, and applied the EP to an fine-mapping study of Rolandic Epilepsy. Bickel (2012) considered an EP application in differential gene expression studies.

1.4 Extension of the Evidential Paradigm to accommodate composite hypotheses

The focus on two simple hypotheses under the Evidential Paradigm can limit its application in many study designs. Both Royall (1997) and Blume (2002) emphasized that even though two simple hypotheses need to be specified at the planning stage, one is encouraged to plot the likelihood function for the parameter of interest and examine all possible pairwise parameter values. However, in genome-wide association studies, this is not feasible due to the large number of markers under study. For other study designs, the research questions can intuitively be formulated as composite hypotheses. For example, when testing for bioequivalence in clinical trials, the null hypothesis of ‘equivalence’ is
Chapter 1. Introduction

represented by $|\theta| < 0.223$, or equivalently, $-0.223 < \theta < 0.223$, where $\theta$ represents the difference between the new treatment and the reference treatment. Therefore, the bioequivalence problem is naturally formulated as that of comparing strength of evidence in support of $\theta \in (-0.223, 0.223)$ over $\theta \notin (-0.223, 0.223)$.

The Law of Likelihood applies to simple hypotheses only, and does not provide guidance on how to measure statistical evidence for composite hypotheses. Neither Royall (1997) nor Blume (2002) was supportive of extending the Evidential Paradigm for composite hypotheses. Royall (1997) provides special examples against the extension. Blume (2002) argues that the graph of the likelihood function provides all information about a parameter of interest, and no further summarization of evidence over a composite hypothesis is necessary. In addition, Blume (2002) discusses a Bayesian approach, where the composite hypotheses are incorporated by specifying a prior distribution for the simple hypotheses. However, Blume (2002) admits that the choices for the priors are subjective, with no objective criteria to choose among them. Zhang and Zhang (2013) and Bickel (2012) are among the authors who appreciate the relevance of composite hypotheses in practical statistical applications, and support extension of the Evidential Paradigm to accommodate composite hypotheses. The authors independently proposed the General/Generalized Law of Likelihood and the Generalized Likelihood Ratio to quantify the strength of evidence in support of one composite hypothesis over another. The Generalized Law of Likelihood basically says:

Given a single parameter of interest $\theta \in \Theta$ and two composite hypotheses $\Theta' \subset \Theta$ and $\Theta'' \subset \Theta$, the strength of evidence in $X = x$, that supports $\theta \in \Theta'$ over $\theta \in \Theta''$ is measured by the Generalized Likelihood Ratio (GLR), where

$$GLR(\Theta', \Theta''; x) = \frac{\sup_{\theta' \in \Theta'} L(\theta'; x)}{\sup_{\theta'' \in \Theta''} L(\theta''; x)}$$

The Generalized Law of Likelihood basically says the strength of evidence in support of one composite hypothesis over another composite hypothesis is measured by the GLR - the ratio of the two likelihood functions, each maximized over the set of parameter values defined by the composite hypotheses. Note that in the original propositions by Bickel (2012) and Zhang and Zhang (2013), there is no restriction on the relationship of the two composite hypotheses, and they do not need to be complementary or disjoint.
Bickel (2012) provided theoretical justification for the use of the GLR, and he further studied the properties of the GLR and showed that the GLR is consistent and interpretable. That is, under regularity conditions that ensure consistency of $\hat{\theta}$, the maximum likelihood estimate of the true value $\theta$, and for any $\Theta' \subset \Theta$ and $\Theta'' \subset \Theta$ such that $\theta \in \Theta''$,

$$\lim_{n \to \infty} P_{\theta}[W(\Theta'', \Theta\setminus\Theta'') > 1] = 1 \text{ (Consistency)}$$

$$\lim_{n \to \infty} P_{\theta}[W(\Theta', \Theta'') \geq \Lambda] = 0 \text{ for all } \Lambda > 1 \text{ (Interpretability)}$$

**Example 2.** An evidential analysis with composite hypotheses (Zhang and Zhang (2013), P161-162)

A randomized clinical trial enrolled 164 children with nephroblastoma, who were randomly assigned to either chemotherapy or radiation therapy. The primary objective of the trial was to demonstrate that chemotherapy is non-inferior to radiation therapy with respect to the response rate, where non-inferiority means the response rate for chemotherapy is not lower than that for radiation therapy by more than a margin of 10%, which is considered the smallest clinically meaningful difference between the two groups. The observed response rates were 94.3% (83/88) for chemotherapy and 90.8% (69/76) for radiation therapy. Figure 1.3 presents the approximate likelihood function for the difference between the response rates in the two groups. Based on the GLL, the non-inferiority hypothesis is strongly supported with a GLR = 0.6947/0.0047 = 148. With a higher response rate in the chemotherapy group, there is even evidence supporting the superiority of chemotherapy to radiation therapy. This latter piece of evidence is rather weak, though, with a GLR = 1/0.6947 = 1.44.

Extending the EP to accommodate composite hypotheses through the use of the GLR seems to offer promising solutions to situations where composite hypotheses are more relevant and appropriate. But work remains to determine whether the GLR is operationally sound with small and bounded error probabilities.
1.5 Outline of thesis

In chapter 2, we address the issue of ranking individual rare variants within an associated region. Motivated by the EP, we develop a conditional likelihood ratio-based measure, maxLRe, for prioritizing rare variants, which does not require one to specify a direction of association or choose among multiple methods to compute exact p-values. We show analytically that the maxLRe is based on the same underlying model as Fisher’s exact test. Via simulation, we compare the prioritization performance of the maxLRe to Fisher’s exact p-value.

In chapter 3, we address the issue of evaluating association evidence for composite hypotheses. We analytically derive the operational characteristics for the composite hypotheses EP framework ($EPC$) where statistical evidence is measured by the generalized likelihood ratio. Specifically, we provide results for the special case of Normal distribution, and then generalize the results for one-parameter parametric models and multi-parameter parametric models with nuisance parameters. We show that the $EPC$ has favorable operational characteristics with small and bounded error probabilities. We
also provide general guidelines for study planning under the $EP^C$, and we discuss the difference in sample size calculations between $EP^C$ and EP with simple hypotheses.

In chapter 4, we focus on applications of the $EP^C$ in genetic association studies. We first derive the operational characteristics for a typical genetic association study with a quantitative phenotype. Then using simulated data, we highlight the advantages of assessing association evidence using the GLR in large-scale genetic association studies.

In chapter 5, we briefly discuss directions for future work.
Chapter 2

Maximized Conditional Likelihood Ratio

In this chapter, we propose an EP-motivated method to prioritize rare variants within an associated region under a case-control study design, and this work has been published in Human Heredity (Li et al., 2015). In section 2.1, we propose a likelihood ratio-based measure, the maxLRc, for ranking rare variants, and we derive theoretical results to show that the proposed measure is always well-defined, even when the data is in separation. In section 2.2, we compare prioritization performance of the maxLRc to the commonly used Fisher’s exact test p-values through simulations. In section 2.3, we apply our method to a study of Rolandic Epilepsy, and at the end, we summarize the results and discuss future directions in section 2.4.

With a case-control study design, a 2x3 contingency table can be constructed to compare genotype frequencies in cases and controls at each rare variant. Due to the low minor allele frequency (MAF) of a rare variant, the probability of observing the homozygous genotype for the minor allele is extremely low, and it is customary to collapse the data into a 2x2 table by case-control status and minor allele carrier status (Table 2.1).

2.1 Introducing the maximum conditional likelihood ratio

We propose to rank rare sequence variants using a likelihood ratio-based measure, the maxLRc. Let $Y_i$ denote the case-control status for a random sample of $i = 1, ..., N$ indi-
Table 2.1: An example of a 2x2 contingency table classified by disease status and minor allele carrier status.

<table>
<thead>
<tr>
<th></th>
<th>Carrier of minor allele</th>
<th>Non-carrier of minor allele</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case</strong></td>
<td>( t_1 )</td>
<td>( t_0 - t_1 )</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>( n_1 - t_1 )</td>
<td>( n_2 - t_0 + t_1 )</td>
</tr>
<tr>
<td></td>
<td>( n_1 )</td>
<td>( n_2 )</td>
</tr>
</tbody>
</table>

Individuals, with \( Y_i = 1 \) for cases and \( Y_i = 0 \) for controls, and \( X_i \), a dichotomous variable with \( X_i = 1 \) for carriers of the minor allele and \( X_i = 0 \) for non-carriers. Assuming a logistic regression model,

\[
\logit P(Y_i = 1) = \beta_0 + \beta_1 X_i ,
\]

the parameter of interest is \( \beta_1 \), the log odds ratio, which measures the effect of the rare variant on the case-control status. In a logistic regression model, the exposures are usually regarded as fixed quantities while the case-control status is random. Although in a typical genetic association study design, subjects are selected based on their case-control status, the logistic model provides the correct estimate of \( \beta_1 \) invariant to study design (Breslow et al., 1982). So here we consider the number of carriers and non-carriers of the minor allele, denoted by \( n_1 \) and \( n_2 \), respectively, as fixed under the logistic regression model.

For a 2x2 table, let \( t_0 = \sum_{i=1}^N Y_i \) denote the total number of cases and \( t_1 = \sum_{i=1}^N X_iY_i \) the total number of cases who carry the minor allele (Table 2.1). Then the full likelihood of the observed data is given by

\[
L(\beta_0, \beta_1; t_0, t_1) = \binom{n_1}{t_1} \left[ \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} \right]^{t_1} \left[ \frac{1}{1 + \exp(\beta_0 + \beta_1)} \right]^{n_1 - t_1} \\
\times \binom{n_2}{t_0 - t_1} \left[ \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \right]^{t_0 - t_1} \left[ \frac{1}{1 + \exp(\beta_0)} \right]^{n_2 - t_0 + t_1}
\]

To re-express the likelihood as a function of \( \beta_1 \) alone, we condition the full likelihood on the sufficient statistic for \( \beta_0 \), \( t_0 \), and the resulting conditional likelihood function has the
following form
\[
L_c(\beta_1|t_0; t_1) = \frac{(n_1)! (n_2)! (t_{0-t_1}) \exp(\beta_1 t_1)}{\sum_{\mu = \max\{0, t_0 - n_2\}}^{\min\{t_0, n_1\}} \mu (n_1)! (n_2)! (t_{0-\mu}) \exp(\beta_1 \mu)}
\]
where, \( \theta = \exp(\beta_1) \) is the odds ratio.

Note that conditioning on the sufficient statistic \( t_0 \) is equivalent to fixing the total number of cases, and in turn, the total number of controls. Since the number of carriers and non-carriers are also fixed under the model assumption, both margins of the 2x2 table are considered fixed. This is the same assumption underlying Fisher’s exact test, and therefore, the conditional likelihood has exactly the same formulation as Fisher’s non-central hypergeometric distribution.

Let \( \hat{\theta}_{MCL} \) denote the maximum conditional likelihood estimate of the odds ratio. We define
\[
\maxLR_c = \frac{L_c(\hat{\theta}_{MCL})}{L_c(1)}
\]
and we propose the simple idea of prioritizing rare sequence variants by the \( \maxLR_c \). Essentially, the \( \maxLR_c \) contrasts the conditional likelihood at the odds ratio value that is best supported by the data and the value of 1, representing no association. Fully specifying the conditional likelihood, we have
\[
\maxLR_c = \frac{L_c(\hat{\theta}_{MCL})}{L_c(1)}
= \frac{(n_1)! (n_2)! (\hat{\theta}_{MCL})^{t_1}}{\sum_{\mu = \max\{0, t_0 - n_2\}}^{\min\{t_0, n_1\}} \mu (n_1)! (n_2)! (\hat{\theta}_{MCL})^{\mu}}
\]
\[
= \hat{\theta}_{MCL}^{t_1} \sum_{\mu = \max\{0, t_0 - n_2\}}^{\min\{t_0, n_1\}} \mu (n_1)! (n_2)! (\hat{\theta}_{MCL})^{\mu}
\]
Similar to the maximum unconditional likelihood estimate of the odd ratio, \( \hat{\theta}_{MCL} \) does not always exist for 2x2 tables. In general, 2x2 contingency tables can be classified
Table 2.2: Examples of 2x2 tables under complete or quasi-complete separation.

<table>
<thead>
<tr>
<th></th>
<th>Carrier</th>
<th>Non-carrier</th>
<th></th>
<th>Carrier</th>
<th>Non-carrier</th>
<th></th>
<th>Carrier</th>
<th>Non-carrier</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Case A</td>
<td>0</td>
<td>b.</td>
<td>Case A</td>
<td>0</td>
<td>c.</td>
<td>Case A</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>C</td>
<td></td>
<td>Control</td>
<td>C</td>
<td></td>
<td>Control</td>
<td>0</td>
</tr>
<tr>
<td>d.</td>
<td>Carrier</td>
<td>0</td>
<td>e.</td>
<td>Carrier</td>
<td>A</td>
<td>f.</td>
<td>Carrier</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>B</td>
<td></td>
<td>Control</td>
<td>C</td>
<td></td>
<td>Control</td>
<td>D</td>
</tr>
</tbody>
</table>

into three cases (Albert and Anderson, 1984):

1. complete separation, where the two cell counts in the main- or off-diagonal of the 2x2 table equal zero;
2. quasi-complete separation, where only one of the four cell counts is zero;
3. overlap, in which case there is no zero cell in the table.

While the $\hat{\theta}_{MCLE}$ always exists in the overlap case, it is not well-defined when the data is in complete or quasi-complete separation, i.e. $\hat{\theta}_{MCLE} = 0$ or $\infty$; however, we show analytically below that the maxLRc remains well-defined in these two cases, and it is equal to

$$\frac{\binom{N}{t_0}}{\binom{n_1}{t_1} \binom{n_2}{t_0-t_1}}.$$

When a 2x2 table is under complete or quasi-complete separation (see Table 2.2 for examples), the maximum conditional likelihood estimate of the odd ratio, $\hat{\theta}_{MCLE}$, is equal to $\infty$ (Table 2.2 a-c) or 0 (Table 2.2 e-f). Without loss of generality, we will provide a proof here for the 2x2 table in Table 2.2 (c), where $\hat{\theta}_{MCLE} = \infty$.

**Proof**

For Table 2.2 (c), $n_1 = t_1 < t_0$, and therefore, $\min\{t_0, n_1\} = t_1$

For notational simplicity, let us further denote $\max\{0, t_0 - n_2\}$ by $m$

$$\text{maxLRc} = \hat{\theta}_{MCLE}^{t_1} \sum_{\mu = \max\{0, t_0 - n_2\}}^{\mu = \min\{t_0, n_1\}} \binom{n_1}{\mu} \binom{n_2}{t_0 - \mu} / \sum_{\mu = \max\{0, t_0 - n_2\}}^{\mu = \min\{t_0, n_1\}} \binom{n_1}{\mu} \binom{n_2}{t_0 - \mu} \hat{\theta}_{MCLE}^\mu$$

$$= \hat{\theta}_{MCLE}^{t_1} \sum_{\mu = m}^{\mu = t_1} \binom{n_1}{\mu} \binom{n_2}{t_0 - \mu} / \sum_{\mu = m}^{\mu = t_1} \binom{n_1}{\mu} \binom{n_2}{t_0 - \mu} \hat{\theta}_{MCLE}^\mu$$
The last equation holds because $m < t_1, m + 1 < t_1, ..., t_1 - 1 < t_1$, and therefore, the terms $\frac{\hat{\theta}_{M_i}^{MCLE}}{\hat{\theta}_{M_{i+1}}^{MCLE}}, ... , \frac{\hat{\theta}_{M_{i-1}}^{MCLE}}{\hat{\theta}_{M_i}^{MCLE}}$ are all equal to 0 when $\hat{\theta}_{MCLE} = \infty$.

Case 1. if $t_0 - n_2 > 0$, $\max\{0, t_0 - n_2\} = t_0 - n_2$

The numerator of Equation A

\[
= \binom{n_1}{m} \binom{n_2}{t_0 - m} + \binom{n_1}{m+1} \binom{n_2}{t_0 - m - 1} + ... + \binom{n_1}{t_1 - 1} \binom{n_2}{t_0 - t_1 + 1} + \binom{n_1}{t_1} \binom{n_2}{t_0 - t_1} 
\]

\[
= \binom{n_1}{m} \binom{n_2}{t_0 - m - 1} + ... + \binom{n_1}{t_1 - 1} \binom{n_2}{t_0 - t_1 + 1} + \binom{n_1}{t_1} \binom{n_2}{t_0 - t_1} 
\]

\[
= \binom{n_1}{m} \binom{n_2}{t_0 - m - 1} + ... + \binom{n_1}{t_1 - 1} \binom{n_2}{t_0 - t_1 + 1} + \binom{n_1}{t_1} \binom{n_2}{t_0 - t_1} \cdot 1 
\]

Equation A

The numerator of Equation A

\[
= \binom{n_1}{m} \binom{n_2}{t_0 - m - 1} + ... + \binom{n_1}{t_1 - 1} \binom{n_2}{t_0 - t_1 + 1} + \binom{n_1}{t_1} \binom{n_2}{t_0 - t_1} 
\]

Case 2. if $t_0 - n_2 < 0$, $\max\{0, t_0 - n_2\} = 0$

The numerator of Equation A

\[
= \binom{n_1}{m} \binom{n_2}{t_0 - m - 1} + ... + \binom{n_1}{t_1 - 1} \binom{n_2}{t_0 - t_1 + 1} + \binom{n_1}{t_1} \binom{n_2}{t_0 - t_1} 
\]

\[
= \binom{n_1}{m} \binom{n_2}{t_0 - m - 1} + ... + \binom{n_1}{t_1 - 1} \binom{n_2}{t_0 - t_1 + 1} + \binom{n_1}{t_1} \binom{n_2}{t_0 - t_1} \cdot 1 
\]

Equation A

The numerator of Equation A

\[
= \binom{n_1}{m} \binom{n_2}{t_0 - m - 1} + ... + \binom{n_1}{t_1 - 1} \binom{n_2}{t_0 - t_1 + 1} + \binom{n_1}{t_1} \binom{n_2}{t_0 - t_1} 
\]
Chapter 2. Maximized Conditional Likelihood Ratio

\[ \binom{N}{i_0} \]

Therefore, \( \text{maxLRc} = \frac{\binom{N}{t_0}}{\binom{n_1}{t_1} \binom{n_2}{t_0-t_1}} \cdot \left[ \frac{\binom{n_1}{t_1} \binom{n_2}{t_0-t_1}}{\binom{N}{i_0}} \right]^{-1} \)

The above result shows that in the two separation cases, the maxLRc is simply the inverse of the hypergeometric probability of the observed 2x2 table.

### 2.2 Rare variant prioritization performance comparison based on simulated data

#### 2.2.1 Fisher’s two-sided exact p-value

In the hypothesis testing framework, 1-sided tests may be appropriate and more powerful than 2-sided tests if additional information on the directions of association is known and correct. However, such information is rarely available for all causal and nearby variants in a region of interest. One strategy is to assume the same effect for all variants in a region (Li and Leal, 2008), however, such assumption can lead to substantially inferior ranking performance, as observed in our simulations. When ranking using 1-sided Fisher’s exact p-values, if the directions of effect for all causal variants are randomly assigned or fixed at 80% positive, 1-sided Fisher’s exact p-value performs substantially worse (across all three ranking performance metrics defined later in section 2.2.3) than the maxLRc and 2-sided p-value. Only when all causal variants are simulated to have positive effect and 1-sided p-values are calculated in the positive direction, do they outperform the other two methods. Therefore, we focus on comparisons of prioritization performance between the maxLRc and 2-sided Fisher’s exact p-values.

There are many strategies to compute 2-sided p-values for Fisher’s exact test, each
with its own advantages and drawbacks. Among these methods, the most popular and widely implemented method is the minimum likelihood approach. In this approach, 2-sided Fisher’s exact p-value is calculated by summing up the probabilities of all possible tables with the same margins as the observed one, whose associated probabilities are less than or equal to the probability of the observed table. We use this minimum likelihood approach here.

### 2.2.2 Details of rare variant simulation

We performed extensive simulations to empirically compare the prioritization performance between the maxLRc and Fisher’s exact p-value. We considered the number of cases to be 100, 150, 200, 250, 500 and 1000, with a control: case ratio of 1, 1.5 and 2, for a total of 18 sample size combinations. For each sample size, we simulated a total of 100 rare variants, among which, the number of truly associated rare variants, Q, was set to be 10 or 20. Each rare variant, regardless of causal or null, was generated with a MAF randomly selected from the Uniform (0.005, 0.05) distribution. For truly associated variants, we first considered the scenario where the directions of effect were randomly assigned as positive or negative, and therefore, the proportion of deleterious or protective variants was not fixed. We then considered the scenario where 80% of the truly associated variants have deleterious effects. The effect sizes, $|\beta_1|$, for the Q truly associated variants ranged from 0.41 to 2.5, corresponding to odd ratios of 1.5 to 10.5 for deleterious variants or $\frac{1}{10.5}$ to $\frac{1}{1.5}$ for protective variants. All variants were simulated under Hardy-Weinberg Equilibrium and all null variants were simulated assuming an underlying odds ratio of 1. Due to the low MAF, rare variants usually do not show strong linkage disequilibrium with each other (Pritchard, 2001; Pritchard and Cox, 2002), and therefore, all variants were simulated independently of each other. For each set of parameter values, we repeated the simulation 1000 times.
2.2.3 Prioritization performance comparison

The comparisons between the maxLRc and Fisher's exact p-values were based on three metrics:

1. Kendalls correlation between the true ranking underlying the simulation and the ranking assigned by the maxLRc or Fisher's exact p-values;

2. the number of truly associated (NT) variants among the K selected, where K is a pre-specified number of variants to be selected for follow-up, and we considered all integer values of K ranging from 1 to 50;

3. the average ranking of the collection of truly associated variants.

To evaluate (1), we only used the set of Q truly associated variants, and the true ranking was defined by the ordering of the absolute values of the underlying effect sizes, $|\beta_1|$. Here, a larger Kendalls correlation coefficient indicates better agreement with the true ranking. For (2), we considered the scenario where there are limited resources and only K variants are to be followed-up, and we compared the number of truly associated variants in the K selected according to rankings assigned by the maxLRc or Fisher's exact p-values. At each value of K considered, we plotted the median of the 1000 NT values (obtained from the 1000 repetitions) according to the maxLRc or Fisher's exact p-values (NT plot). To compare the overall ranking performance, we summed up NT over all values of K, here denoting this sum by sumK, and a larger sumK value indicates better overall performance. In (3), for each repetition, we calculated the average ranking of the Q truly associated variants assigned by each of the maxLRc and exact p-values, and we then compared the means of the 1000 averages.

We now focus on ranking performance comparisons between the maxLRc and 2-sided Fisher’s exact p-values, and present results from simulations generated with an equal number of cases and controls, although results for other control: case ratios are similar.
Table 2.3: Mean Kendall’s correlation coefficients, across 1000 repetitions, for rare variants generated with MAF ∈ [0.005,0.05], number of truly associated variants Q = 10, and an equal number of cases and controls.

<table>
<thead>
<tr>
<th>Number of Variants</th>
<th>100 (10 causal, 90 null)</th>
<th>8 deleterious, 2 protective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directions of effect for causal variants</td>
<td>maxLRc</td>
<td>Fisher’s p-value (2-sided)</td>
</tr>
<tr>
<td>Sample Size</td>
<td>maxLRc</td>
<td>Fisher’s p-value (2-sided)</td>
</tr>
<tr>
<td>100: 100</td>
<td>0.39</td>
<td>0.35</td>
</tr>
<tr>
<td>150: 150</td>
<td>0.42</td>
<td>0.40</td>
</tr>
<tr>
<td>200: 200</td>
<td>0.43</td>
<td>0.41</td>
</tr>
<tr>
<td>250: 250</td>
<td>0.43</td>
<td>0.41</td>
</tr>
<tr>
<td>500: 500</td>
<td>0.47</td>
<td>0.46</td>
</tr>
<tr>
<td>1000: 1000</td>
<td>0.50</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Table 2.4: Mean Kendall’s correlation coefficients, across 1000 repetitions, for rare variants generated with MAF ∈ [0.005,0.05], number of truly associated variants Q = 20, and an equal number of cases and controls.

<table>
<thead>
<tr>
<th>Number of Variants</th>
<th>100 (20 causal, 80 null)</th>
<th>16 deleterious, 4 protective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directions of effect for causal variants</td>
<td>maxLRc</td>
<td>Fisher’s p-value (2-sided)</td>
</tr>
<tr>
<td>Sample Size</td>
<td>maxLRc</td>
<td>Fisher’s p-value (2-sided)</td>
</tr>
<tr>
<td>100: 100</td>
<td>0.34</td>
<td>0.30</td>
</tr>
<tr>
<td>150: 150</td>
<td>0.40</td>
<td>0.37</td>
</tr>
<tr>
<td>200: 200</td>
<td>0.43</td>
<td>0.41</td>
</tr>
<tr>
<td>250: 250</td>
<td>0.44</td>
<td>0.42</td>
</tr>
<tr>
<td>500: 500</td>
<td>0.47</td>
<td>0.46</td>
</tr>
<tr>
<td>1000: 1000</td>
<td>0.50</td>
<td>0.49</td>
</tr>
</tbody>
</table>

The mean Kendall’s correlation coefficients, averaged over the 1000 repetitions, are presented in table 2.3 for the scenarios where the number of truly associated variants Q = 10, and table 2.4 for Q = 20. In all cases, rankings assigned by the maxLRc are in better agreement with the underlying true rankings, as indicated by higher mean Kendall’s correlation.

NT plots for simulations with Q = 10 are provided in Figures 2.1 and 2.2; and those with Q = 20 are provided in Figures 2.3 and 2.4. When the directions of association are randomly assigned, the maxLRc selects more or an equal number of truly associated variants than 2-sided Fisher’s exact p-value for a given K in almost all cases. When
the majority (80%) of the truly associated variants have deleterious effect, the maxLRc performs better or worse depending on the value of $K$, however, it has a better overall performance, as indicated by a larger SumK value, in all sample size configurations except when the sample size is small at $N = 200$ (see results in Tables 2.5 and 2.6).

The means of the average rankings of the collection of truly associated variants are summarized in tables 2.7 and 2.8 for $Q = 10$ and $Q = 20$, respectively. The means of the average rankings of the collection of all causal variants are smaller by the maxLRc than 2-sided Fisher’s exact p-value in all cases considered; that is, on average, the truly associated variants are always collectively ranked higher by the maxLRc. Results from simulations with control:case ratios of 1.5 and 2 are similar to those presented above.
Figure 2.2: NT plots for rare variants simulated with MAF ∈ [0.005, 0.05], number of truly associated variants Q = 10, and with 80% of the casual variants having deleterious effect. The maxLRc selects more or an equal number of truly associated variants than Fisher’s 2-sided exact p-value in most cases.

Figure 2.3: NT plots for rare variants simulated with MAF ∈ [0.005, 0.05], number of truly associated variants Q = 20, and with randomly assigned directions of effect. The maxLRc selects more or an equal number of truly associated variants than Fisher’s 2-sided exact p-value across all values of K under almost all sample size configurations.
Figure 2.4: NT plots for rare variants simulated with MAF ∈ [0.005, 0.05], number of truly associated variants Q = 20, and 80% of the truly associated variants having deleterious effects. The maxLRc selects more or an equal number of truly associated variants than Fisher’s 2-sided exact p-value in most cases.

Table 2.5: Sum of median number of truly associated variants (NT) over values of K from 1 to 50 (SumK) for rare variants generated with MAF ∈ [0.005,0.05], number of truly associated variants Q = 10, and an equal number of cases and controls. Sum of median proportion of truly associated variants (NT/K) is provided in brackets.
Table 2.6: Sum of median number of truly associated variants (NT) over values of K from 1 to 50 (SumK) for rare variants generated with MAF \in [0.005,0.05], number of truly associated variants Q = 20, and an equal number of cases and controls. Sum of median proportion of truly associated variants (NT/K) is provided in brackets.

<table>
<thead>
<tr>
<th>Number of Variants</th>
<th>Directions of effect for causal variants</th>
<th>Randomly assigned</th>
<th>16 deleterious, 4 protective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>maxLRc Fisher’s p-value (2-sided)</td>
<td>maxLRc Fisher’s p-value (2-sided)</td>
<td></td>
</tr>
<tr>
<td>100: 100</td>
<td>329 (18.32) 319 (17.95) 359 (20.19) 360 (20.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150: 150</td>
<td>362 (20.24) 352 (19.83) 381 (21.52) 380 (21.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200: 200</td>
<td>378 (21.39) 371 (21.06) 393 (22.36) 391 (22.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>250: 250</td>
<td>391 (22.20) 386 (21.93) 406 (23.10) 403 (23.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500: 500</td>
<td>431 (24.24) 426 (23.93) 439 (24.59) 438 (24.55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000:1000</td>
<td>451 (25.35) 451 (25.35) 453 (25.51) 452 (25.43)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.7: Mean of the average rankings of the collection of truly associated variants, across 1000 repetitions, for rare variants generated with MAF \in [0.005,0.05], number of truly associated variants Q = 10, and an equal number of cases and controls.

<table>
<thead>
<tr>
<th>Number of Variants</th>
<th>Directions of effect for causal variants</th>
<th>Randomly assigned</th>
<th>8 deleterious, 2 protective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>maxLRc Fisher’s p-value (2-sided)</td>
<td>maxLRc Fisher’s p-value (2-sided)</td>
<td></td>
</tr>
<tr>
<td>100: 100</td>
<td>19.32 20.63 16.71 16.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150: 150</td>
<td>16.97 18.19 14.59 14.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>250: 250</td>
<td>13.42 14.27 11.75 12.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500: 500</td>
<td>9.64 10.04 8.62 8.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000:1000</td>
<td>7.16 7.30 6.79 6.87</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.8: Mean of the average rankings of the collection of truly associated variants, across 1000 repetitions, for rare variants generated with MAF \in [0.005,0.05], number of truly associated variants Q = 20, and an equal number of cases and controls.

<table>
<thead>
<tr>
<th>Number of Variants</th>
<th>Directions of effect for causal variants</th>
<th>Randomly assigned</th>
<th>16 deleterious, 4 protective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>maxLRc Fisher’s p-value (2-sided)</td>
<td>maxLRc Fisher’s p-value (2-sided)</td>
<td></td>
</tr>
<tr>
<td>100: 100</td>
<td>22.75 23.73 19.98 20.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150: 150</td>
<td>20.33 20.97 17.82 17.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200: 200</td>
<td>17.84 18.39 16.24 16.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>250: 250</td>
<td>16.92 17.33 15.21 15.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000:1000</td>
<td>12.03 12.14 11.62 11.68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.3 Prioritizing rare variants in a study of Rolandoic Epilepsy

We illustrate how rankings assigned by the \text{maxLRe} and 2-sided Fisher’s exact \text{p-value} could differ substantially for the same set of rare variants using next-generation sequence data from a Rolandoic Epilepsy study. Rolandoic Epilepsy is the most common epilepsy syndrome in childhood (Shinnar et al., 1999), and it is linked to and associated with a 600-kilobase region on chromosome 11 (Strug et al., 2009). To prioritize individual rare variants within this region, we used next generation sequence data from 27 Rolandoic Epilepsy cases sequenced on the Illumina Genome Analyzer IIx platform, and an independent sample of 200 individuals ascertained for a study of colorectal cancer whole-genome sequenced by Complete Genomics and made available to us as a control group (Derkach et al., 2014). After standard quality control analysis, 207 rare variants with \text{MAF} \leq 5\% remained for the prioritization. We then ranked the 207 rare variants based on both the \text{maxLRe} and 2-sided Fisher’s exact \text{p-value}. In addition, we used simulations to compare the performance of the two methods under the sample size configuration as the Rolandoic Epilepsy study. Specifically, we simulated rare variants with 27 cases and 200 controls, \(Q = 10\), and randomly assigned directions of association.

The top ten ranked variants by the \text{maxLRe}, together with their corresponding rankings by 2-sided Fisher’s exact \text{p-values}, are provided in table 2.9. It is evident that the two methods prioritize this set of variants differently, disagreeing even on the top ranked variant. The ranking discrepancies can be substantial, for example, rs1806176, which is ranked 5th by the \text{maxLRe}, is only ranked the 12th by the \text{p-value} approach. Simulation results suggest that the \text{maxLRe} outperforms 2-sided Fisher’s exact \text{p-value} under the same sample size configuration, with Kendall’s correlation coefficient = 0.31 and 0.23, \(\text{SumK} = 269\) and 244, and average ranking of the collection of truly associated variants.
Table 2.9: Prioritization of rare sequence variants from 27 Rolandic Epilepsy cases and 200 Colorectal Cancer controls; 10 rare variants top ranked by the maxLRc are shown, together with their corresponding rankings assigned by 2-sided Fisher’s exact p-value

<table>
<thead>
<tr>
<th>Chr</th>
<th>Variant</th>
<th>BP</th>
<th>MAF</th>
<th>maxLRc</th>
<th>Fisher’s P</th>
<th>Ranking by maxLRc</th>
<th>Ranking by Fisher’s P</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>rs6484529</td>
<td>31724195</td>
<td>0.003</td>
<td>189</td>
<td>0.0053</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>rs180775607</td>
<td>31463255</td>
<td>0.005</td>
<td>88.08</td>
<td>0.0114</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>rs10314119</td>
<td>31605896</td>
<td>0.028</td>
<td>55.83</td>
<td>0.0052</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>rs558508</td>
<td>31800907</td>
<td>0.029</td>
<td>34.49</td>
<td>0.0081</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>rs1806176</td>
<td>31842323</td>
<td>0.017</td>
<td>28.71</td>
<td>0.0348</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>rs78174119</td>
<td>31735627</td>
<td>0.020</td>
<td>23.10</td>
<td>0.0133</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>rs4359181</td>
<td>31759404</td>
<td>0.033</td>
<td>18.21</td>
<td>0.0154</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>rs182818125</td>
<td>31388793</td>
<td>0.007</td>
<td>17.61</td>
<td>0.0265</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>rs182363098</td>
<td>31385443</td>
<td>0.007</td>
<td>17.61</td>
<td>0.0265</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>31393303</td>
<td>31393303</td>
<td>0.007</td>
<td>16.15</td>
<td>0.0290</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

= 26.37 and 29.69, for the maxLRc and 2-sided Fisher’s exact p-value, respectively.

2.4 Conclusion and additional considerations

The maxLRc does as good or outperforms 2-sided Fisher’s exact p-value in prioritizing rare variants in most cases that we considered. Across all simulation scenarios, rankings assigned by the maxLRc correlate better with the underlying true rankings, and the collection of all causal variants is always ranked higher by the maxLRc. When only a few variants are to be selected for follow-up, i.e. K is very small, the two methods perform similarly regardless of sample size; and as sample size gets large, the two methods are expected to have equivalent performance.

The difference in the two methods fundamentally lies in how they measure the strength of statistical evidence. The maxLRc is based on the distribution of the data in the observed 2x2 table, whereas p-values further incorporate the probability of more extreme tables that could have been observed. Which values are to be defined as more extreme depends on whether the investigator is conducting a one-sided or two-sided test, a choice the EP approach does not require.
In constructing the maxLRc, we chose to use conditional likelihood for several reasons. First, it allows for elimination of the nuisance parameter $\beta_0$. It also ensures that the maxLRc is always well-defined, even in the two separation cases; this property does not hold if a maximized likelihood ratio is constructed from estimated or profile likelihoods. Finally, the derivation of the conditional likelihood requires the same assumption as Fisher’s exact test, which excludes the possibility that the two methods perform differently simply due to different model assumptions.

The maxLRc is a conditional likelihood ratio. Therefore, in large samples, $2\log(\text{maxLRc})$ follows a Chi-square distribution with 1 degree of freedom asymptotically, under fairly general regularity conditions (Andersen, 1971). This implies that we could calculate asymptotic p-values based on the maxLRc despite having sparse data, and the prioritization of variants based on this asymptotic p-value would coincide with the rankings provided by maxLRc. Computing asymptotic p-values in the sparse data setting would, of course, be contrary to standard statistical practice from a hypothesis testing perspective.

Although we proposed the maxLRc for rare variant prioritization, this method is applicable to common variants as well, without requiring the genotype categories being collapsed into minor allele carrier status. Let $Y_i = 1$ or 0 represent the case-control status, and $X_i = 0, 1$ or 2 denote the number of minor alleles that subject $i$ carries, therefore, assuming an additive genetic effect. For easier representation, let us further define $n_1, n_2,$ and $n_3$ as the number of subjects carrying 0, 1 or 2 copies of the minor allele; and $r_1, r_2$ and $r_3$, the number of cases carrying 0, 1 or 2 copies of the minor allele, respectively. The sufficient statistic for $\beta_0$ is $S_0 = \sum_{c=1}^{3} r_c = \sum_{i=1}^{N} Y_i$, the total number of cases; and the sufficient statistic for $\beta_1$ is $S_1 = 0 \times r_1 + 1 \times r_2 + 2 \times r_3 = \sum_{i=1}^{N} X_iY_i$. The conditional
Chapter 2. Maximized Conditional Likelihood Ratio

The maximized conditional likelihood for the data, as a function of $\beta_1$ alone, is then

$$L_c(\beta_1|S_0; S_1) = \frac{\binom{n_1}{r_1} \binom{n_2}{r_2} \binom{n_3}{r_3} \exp(\beta_1 S_1)}{\sum_{\Gamma} \binom{n_1^*}{r_1^*} \binom{n_2^*}{r_2^*} \binom{n_3^*}{r_3^*} \exp\{\beta_1(0 \times r_1^* + 1 \times r_2^* + 2 \times r_3^*)\}}$$

where $\Gamma$ includes all possible combinations of $r^* = (r_1^*, r_2^*, r_3^*)$ such that $r_1^* + r_2^* + r_3^* = S_0$ and $0 \times r_1^* + 1 \times r_2^* + 2 \times r_3^* = S_1$.

The maxLRc is then given by

$$\text{maxLRc} = \frac{L_c(\hat{\beta}_{1,MCLE}|S_0; S_1)}{L_c(1|S_0; S_1)},$$

with $\hat{\beta}_{1,MCLE}$ representing the maximum conditional likelihood estimate of $\beta_1$.

The method of conditioning on sufficient statistics, which forms the basis of exact conditional logistic regression (Cox and Snell, 1989), can be used to eliminate multiple nuisance parameters providing a theory for including covariates in the calculation of the maxLRc. However, as the sample size and number of covariates increase, the computational burden becomes prohibitive for practical applications. In such cases, profile likelihood could be used instead of conditional likelihood to maintain computational efficiency.

In summary, the maxLRc provides reliable statistical prioritization of sequence variants and outperforms the standard method of prioritizing by Fisher’s exact p-values in the majority of settings we considered. Although the difference in some cases was minimal, using the maxLRc to prioritize variants avoids the need to make arbitrary decisions about hypothesis testing parameters such as whether to compute one- or two-sided p-values, and which two-sided p-value to compute. Moreover, the computational time is equivalent for the maxLRc and Fisher’s exact p-values, making it an attractive and easy to implement alternative. The maxLRc is applicable to both rare and common variants and can be easily implemented in R (R Core Team, 2012).
Chapter 3

Generalized Likelihood Ratio for Composite Hypotheses

Previously in Chapter 2, we considered the situation of prioritizing/ranking rare variants in a region of established association. Here we develop methods that address the issue of identifying regions of association. In this setting, the focus on two simple hypotheses under the Evidential Paradigm (EP) can limit its application. To extend the EP to accommodate composite hypotheses, Bickel (2012) and Zhang and Zhang (2013) independently proposed the Generalized Law of Likelihood (GLL) and the Generalized Likelihood Ratio (GLR), where the GLR is used to measure the strength of statistical evidence for two composite hypotheses. However, neither Bickel (2012) nor Zhang and Zhang (2013) analytically derived the operational characteristics associated with the use of the GLR. To ensure that the GLR is a valid and reliable instrument for measuring statistical evidence, it is crucial to show that it does not lead one to incorrect conclusions often, i.e. the probability of observing misleading evidence has to be bounded and small. Furthermore, it is important to study and understand the behavior of the operational characteristics for study planning. In section 3.1, we specify the composite hypotheses that are particularly relevant for genetic association studies, and we define the opera-
tional characteristics associated with the GLR. In section 3.2, we present our analytical results for the Normal probability model. In section 3.3 and 3.4, we present asymptotic results for one-parameter and multi-parameter parametric models, and discuss options for eliminating nuisance parameters. Based on these results, we provide guidelines for study planning in section 3.5, and we give concluding remarks and discuss future work in section 3.6.

3.1 Complementary composite hypotheses and the associated operational characteristics

We first rephrase the Generalized Law of Likelihood and the Generalized Likelihood Ratio for composite hypotheses. We then define the specific complementary composite hypotheses that are of the most relevance to genetic association studies.

Generalized Law of Likelihood

Given the parameter space $\Theta$ and two composite hypotheses $\Theta' \subset \Theta$ and $\Theta'' \subset \Theta$, the strength of evidence in a given body of data, $X = x$, that supports $\theta \in \Theta'$ over $\theta \in \Theta''$ is measured by the Generalized Likelihood Ratio (GLR)

$$\text{GLR} (\Theta', \Theta''; x) = \frac{\sup_{\theta' \in \Theta'} L(\theta'; x)}{\sup_{\theta'' \in \Theta''} L(\theta''; x)}$$

The GLR is simply the ratio of two likelihoods, maximized within each of the two composite parameter spaces, respectively.

In this chapter, we derive and examine the operational characteristics associated with the GLR when both the null and alternative hypotheses are composite and complementary to each other.
**Complementary composite hypotheses**

Assume there is a single or scalar parameter of interest, \( \theta \), and let \( \theta_0 \) denote the simple parameter value representing ‘no difference’. In genetic association studies, for example, this could be the parameter value representing ‘no association’. Let \( \theta_1 \) denote the simple alternative parameter value, which is usually chosen to be the minimum effect size one aims to detect. Then with a small and positive constant \( 0 < C < |\theta_0 - \theta_1| \), we define the null composite hypothesis, \( \Theta_0 \), as the interval of parameter values \( [\theta_0 - C, \theta_0 + C] \), inclusively. The alternative composite hypothesis is then defined as \( \Theta_1 = \Theta_0^c \), where \( c \) denotes the complement, that is, \( \Theta_1 = (-\infty, \theta_0 - C) \cup (\theta_0 + C, \infty) \).

By defining a composite null hypothesis as a small interval of parameter values surrounding the simple null hypothesis, we consider all \( \theta \) values that fall within \( [\theta_0 - C, \theta_0 + C] \) *not* different qualitatively from \( \theta_0 \). Note that \( \theta_0 \in \Theta_0 \) and \( \theta_1 \in \Theta_1 \) by design. It is easy to see that the width of the null hypothesis interval is given by \( 2C \), and we refer to \( C \) as the margin of the null hypothesis interval.

**Operational characteristics associated with the GLR**

Given a pre-specified constant \( k > 1 \), when \( \theta_0 \) is true

- Probability of misleading evidence, \( M_0 = P_{\theta_0} \left[ \frac{\sup_{\theta'' \in \Theta_1} L(\theta''; x)}{\sup_{\theta' \in \Theta_0} L(\theta'; x)} \geq k \right] \)
- Probability of weak evidence, \( W_0 = P_{\theta_0} \left[ \frac{1}{k} < \frac{\sup_{\theta'' \in \Theta_1} L(\theta''; x)}{\sup_{\theta' \in \Theta_0} L(\theta'; x)} < k \right] \)
- Probability of strong evidence, \( S_0 = P_{\theta_0} \left[ \frac{\sup_{\theta'' \in \Theta_0} L(\theta''; X)}{\sup_{\theta' \in \Theta_1} L(\theta'; X)} \geq k \right] \)
when $\theta_1$ is true:

Probability of misleading evidence, $M_1 = P_{\theta_1}\left[ \frac{\sup_{\theta' \in \Theta_1} L(\theta'; x)}{\sup_{\theta'' \in \Theta_0} L(\theta''; x)} \geq k \right]$

Probability of weak evidence, $W_1 = P_{\theta_1}\left[ \frac{1}{k} \frac{\sup_{\theta' \in \Theta_1} L(\theta'; x)}{\sup_{\theta'' \in \Theta_0} L(\theta''; x)} < k \right]$

Probability of strong evidence, $S_1 = P_{\theta_1}\left[ \frac{\sup_{\theta' \in \Theta_1} L(\theta'; x)}{\sup_{\theta'' \in \Theta_0} L(\theta''; x)} \geq k \right]$

We will first derive the operational characteristics, $(M_0, M_1), (W_0, W_1)$ and $(S_0, S_1)$, associated with the GLR for the Normal distribution.

## 3.2 Theorem 1: Normal distribution

Suppose $X_1, ..., X_n$ are independent and identically distributed random variables with a $N(\theta, \sigma^2)$ probability distribution, with $\sigma^2$ assumed known. Let $\theta_0$ and $\theta_1$ denote the simple null hypothesis and simple alternative hypothesis, respectively. With a small and positive constant $0 < C < |\theta_0 - \theta_1|$, we define the composite null hypothesis as $\Theta_0 = [\theta_0 - C, \theta_0 + C]$ and the composite alternative hypothesis $\Theta_1 = \Theta_0^c = (-\infty, \theta_0 - C) \cup (\theta_0 + C, \infty)$. In addition, let $k$ represent the threshold for declaring strong evidence.

$M_0 = 2 \Phi(-C \sqrt{n}/\sigma - \sqrt{2 \ln k})$

$S_0 = \begin{cases} 
2 \Phi(C \sqrt{n}/\sigma - \sqrt{2 \ln k}) - 1, & \text{if } n > \frac{2 \ln k}{(\frac{\sigma}{\theta})^2} \\
0, & \text{if } n \leq \frac{2 \ln k}{(\frac{\sigma}{\theta})^2} 
\end{cases}$

$W_0 = \begin{cases} 
2[\Phi(C \sqrt{n}/\sigma + \sqrt{2 \ln k}) - \Phi(C \sqrt{n}/\sigma - \sqrt{2 \ln k})], & \text{if } n > \frac{2 \ln k}{(\frac{\sigma}{\theta})^2} \\
2[\Phi(C \sqrt{n}/\sigma + \sqrt{2 \ln k})] - 1, & \text{if } n \leq \frac{2 \ln k}{(\frac{\sigma}{\theta})^2} 
\end{cases}$
Chapter 3. Generalized Likelihood Ratio for Composite Hypotheses

$M_1 = \begin{cases} 
\Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}], & \text{if } n > \frac{2\ln k}{(\frac{\sigma}{\theta_0})^2} \\
0, & \text{if } n \leq \frac{2\ln k}{(\frac{\sigma}{\theta_0})^2} 
\end{cases}$

$W_1 = \begin{cases} 
\Phi[-(\theta_0 + C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] - \Phi[-(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}], & \text{if } n > \frac{2\ln k}{(\frac{\sigma}{\theta_0})^2} \\
= \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}], & \text{if } n \leq \frac{2\ln k}{(\frac{\sigma}{\theta_0})^2} 
\end{cases}$

$S_1 = \Phi[-(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] + \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}]$

A proof is provided in Appendix A.1.

When $\theta_0$ is assumed true, the probabilities of observing misleading, strong and weak evidence, $M_0$, $S_0$ and $W_0$, respectively, are functions of only the pre-specified threshold value of $k$ and $\frac{C\sqrt{n}}{\sigma} = \frac{C}{(\frac{\sigma}{\theta_0})}$. That is, given $k$, the probabilities only depend on the standardized margin or width of the null hypothesis interval but not on the simple hypothesis value of $\theta_0$ or $\theta_1$. This is as expected, since whether we observe misleading, strong or weak evidence and their corresponding probabilities depend on whether, $\hat{\theta}$, the maximum likelihood estimate (MLE) of $\theta$, falls within or outside of the boundaries for the null hypothesis interval, $\theta_0 - C$ and $\theta_0 + C$. For the Normal distribution, $\hat{\theta} = \bar{X}$ follows a Normal distribution with mean $\theta_0$ and variance $\frac{\sigma^2}{n}$. Since the boundaries for the null hypothesis interval, $\theta_0 - C$ and $\theta_0 + C$, are symmetric around $\theta_0$ by design, $M_0$, $W_0$ and $S_0$ only depend on the width of the null hypothesis interval, not on its location. For example, when $C$ gets larger, the null interval gets wider, and $\hat{\theta}$ has a higher chance of falling within the null interval, resulting in a higher probability of observing stronger evidence in support of the null composite hypothesis.

For any $k > 1$ and $\frac{C}{\sigma} \neq 0$, the behavior of the probability of misleading evidence when $\theta_0$ is true, $M_0$, follows the same pattern. The probability is the highest when $n = 0$ and it
Figure 3.1: The probability of misleading evidence, $M_0$, is plotted as a function of sample size.

decreases monotonically as $n$ increases. This is illustrated in Figure 3.1, which presents, for $k = 8$, how $M_0$ varies with sample size when the margin of the null composite hypothesis interval is 0.01, 0.25, 0.5 and 1 standard deviations (i.e. $C/\sigma = 0.01, 0.25, 0.5$ and 1).

The behavior of the probability of misleading evidence under the composite hypothesis framework is very different from the one under the simple hypothesis framework, denoted by $M_0^S$ from here on. As discussed earlier in section 1.3.2, $M_0^S = \Phi\left(-\frac{\Delta\sqrt{n}}{2\sigma} - \frac{\sigma\ln k}{\sqrt{n}\Delta}\right)$, where $\Delta = |\theta_1 - \theta_0|$ is the distance between the two simple hypothesis values. $M_0^S$ starts at 0 when $n = 0$ and rises quickly with increasing sample size to its maximum of $\Phi(-\sqrt{2\ln k})$, and then decreases steadily thereafter towards 0 (see Figure 1.2 for an example).

If the margin of the null interval, $C$, is measured in standard error units, say $C = G \cdot \frac{\sigma}{\sqrt{n}}$, then $M_0$ can be re-expressed as $M_0 = 2\Phi(-G - \sqrt{2\ln k})$, where $G = \frac{C\sqrt{\pi}}{\sigma}$ is a positive constant (assuming sample size $n$ is not 0). It is then easy to see that, holding
k constant, $M_0$ is a monotonic decreasing function of $G$, and the maximum value of $M_0$ is $2\Phi(-\sqrt{2\ln k})$. This maximum only depends on the pre-specified threshold $k$ and not on the sample size $n$ nor the hypothesized parameter values $\theta_0$ and $\theta_1$. Furthermore, this maximum value or upper bound for $M_0$ is not attainable since $G$ is greater than 0 by design. With $k = 8$ and 32, $M_0$ is bounded by 0.0414 and 0.0085, respectively.

It is worth noting that the upper bound for $M_0$ with the composite hypotheses is twice the one under the simple hypothesis framework. For easy reference, we will use $EP^S$ and $EP^C$ to denote the EP framework with simple hypotheses and EP with composite hypotheses, respectively. Under $EP^S$, when the two simple hypotheses, $\theta_0$ and $\theta_1$, are specified, their distance and relation (whether $\theta_1$ is greater or smaller than $\theta_0$) are completely determined and known. While under $EP^C$, $M_0$ is the probability that the data support the alternative composite hypothesis $\theta \in \Theta_1$ while $\theta = \theta_0 \in \Theta_0$ is true. This probability depends on the restricted MLE of $\theta$ constrained to the alternative composite hypothesis ($\hat{\theta}_{\text{alternative}}$) and the restricted MLE of $\theta$ constrained to the null composite hypothesis. $\hat{\theta}_{\text{alternative}}$ can reside to the left or the right of the null interval with unknown distance from the boundaries, therefore, it is expected that there is a higher chance of observing misleading evidence under $EP^C$ than that under the $EP^S$.

The probability of observing strong evidence in support of $\theta \in \Theta_0$ when $\theta_0$ is assumed true is given by $S_0$, which is a conditional function that is described by two different curves. For example, when considered as a function of sample size $n$, $S_0$ is degenerating and equal to 0 when $n \leq \frac{2\ln k}{(\frac{\sigma}{C})^2}$; and when $n > \frac{2\ln k}{(\frac{\sigma}{C})^2}$, $S_0$ is a monotonically increasing function of $n$. The breakpoint value of $\frac{2\ln k}{(\frac{\sigma}{C})^2}$ only depends on the pre-specified values $\sigma$, $k$ and $C$. Similarly, the probability of observing weak evidence, $W_0$, is also a conditional function, with the same breakpoint value of $n = \frac{2\ln k}{(\frac{\sigma}{C})^2}$. 
Figure 3.2: The probabilities of observing misleading, weak and strong evidence, $M_0, W_0$ and $S_0$, are plotted as functions of sample size

In Figure 3.2, we plotted $M_0, W_0$ and $S_0$ against sample size $n$, with $k = 8$ and $C/\sigma = 0.5$. Essentially, when sample size is small or $n < \frac{2\ln k}{(C/\sigma)^2} = 16$, there is no chance of observing strong evidence in support of the null composite hypothesis at all, and the data almost surely provides only weak evidence. As sample size gets larger than 16, the probability of observing weak evidence decreases steadily down to 0 and the probability of strong evidence increases steadily to 1. The probability of misleading evidence stays low for all sample sizes; it starts at its maximum value of 0.01 and decreases to 0 quickly as sample size increases.

When the alternative simple hypothesis $\theta_1$ is assumed true, the operational characteristics depend on $C$ as well as the hypothesized simple parameter values $\theta_0$ and $\theta_1$. Specifically, they depend on the distance between $\theta_1$ and the boundaries of the null hypothesis interval, $\theta_0 - C$ and $\theta_0 + C$. $M_1$ is also a conditional function, and there is
no chance of observing misleading evidence in support of the null composite hypothesis when the \( n \) is less than \( \frac{2 \ln k}{\sigma^2} \). This is consistent with the functional form for \( S_0 \) since misleading evidence is also a form of strong evidence, the difference only lies in which parameter value is assumed true. Here we will show that \( M_1 \) is also small and bounded, and since \( M_1 = 0 \) when \( n \leq \frac{2 \ln k}{\sigma^2} \), we will only consider the case when \( n > \frac{2 \ln k}{\sigma^2} \).

**Case 1.** when \( \theta_1 > \theta_0 + C \), or equivalently, \( \theta_0 + C - \theta_1 < 0 \)

\[
M_1 = \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2 \ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma + \sqrt{2 \ln k}]
\]

\[
\leq \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2 \ln k}]
\]

\[
< \Phi(-\sqrt{2 \ln k})
\]

**Case 2.** when \( \theta_1 < \theta_0 - C \), or equivalently, \( -(\theta_0 - C - \theta_1) < 0 \)

\[
M_1 = \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2 \ln k}] - \{1 - \Phi[-(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2 \ln k}]\}
\]

\[
= \Phi[-(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2 \ln k}] - \{1 - \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma + \sqrt{2 \ln k}]\}
\]

\[
\leq \Phi[-(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2 \ln k}]
\]

\[
< \Phi(-\sqrt{2 \ln k})
\]

Therefore, \( M_1 \) is bounded by \( \Phi(-\sqrt{2 \ln k}) \), which again only depends on \( k \). For \( k = 8 \) and 32, \( M_1 \) is bounded by 0.0207 and 0.0042, respectively.

Note that \( M_0 \) and \( M_1 \) have different upper bounds of \( 2\Phi(-\sqrt{2 \ln k}) \) and \( \Phi(-\sqrt{2 \ln k}) \), respectively, for Normally distributed data. In contrast, under \( EP^S \), \( M_0^S = M_1^S \) for Normally distributed data, and they have the same upper bound of \( \Phi(-\sqrt{2 \ln k}) \). A plot of \( M_1 \) as a function of sample size is provided in Figure 3.3, with \( k \) fixed at 8, \( \frac{\theta_0 - \theta_1}{\sigma} \) fixed at 1, and \( \sigma \) = 0.7, 0.8 or 0.9. When \( n \leq \frac{2 \ln k}{(\sigma^2)} \), we can not observe any misleading evidence in support of the null composite hypothesis; and once \( n > \frac{2 \ln k}{(\sigma^2)} \), \( M_1 \) quickly increases to its maximum and then starts to decrease towards 0. Of note is that \( \Phi(-\sqrt{2 \ln k}) \) is a
Figure 3.3: The probability of misleading evidence, $M_1$, is plotted as a function of sample size

global upper bound for $M_1$ which only depends on $k$. Given a set of parameter values, the upper bound for $M_1$ over all sample sizes is, in general, much smaller. For example, for $k = 8, \frac{\theta_0 - \theta_1}{\sigma} = 1$ and $C/\sigma = 0.7$, the maximum value of $M_1$ is only 0.001 which is observed at $n = 11$, while the global upper bound for $M_1$ is $\Phi(-\sqrt{2 \ln 8}) = 0.0207$.

In Figure 3.4, we plotted $M_1, W_1$ and $S_1$ as functions of sample size $n$, with $k = 8, \frac{\theta_0 - \theta_1}{\sigma} = -1$, and $C/\sigma = 0.5$. $S_1$ increases monotonically with sample size and $W_1$ decreases as $n$ increases.

We can also study the operational characteristics as functions of $C$ while treating other parameter values as fixed. For the Normal distribution case, we can re-write the probabilities as:
Figure 3.4: The probabilities of observing misleading, weak and strong evidence, $M_1$, $W_1$ and $S_1$, are plotted as functions of sample size.

$M_0 = 2\Phi(-C\sqrt{n}/\sigma - \sqrt{2\ln k})$

$S_0 = \begin{cases} 
2\Phi(C\sqrt{n}/\sigma - \sqrt{2\ln k}) - 1, & \text{if } C > \frac{\sigma}{\sqrt{n}}\sqrt{2\ln k} \\
0, & \text{if } C \leq \frac{\sigma}{\sqrt{n}}\sqrt{2\ln k}
\end{cases}$

$W_0 = \begin{cases} 
2\Phi(C\sqrt{n}/\sigma + \sqrt{2\ln k}) - \Phi(C\sqrt{n}/\sigma - \sqrt{2\ln k}), & \text{if } C > \frac{\sigma}{\sqrt{n}}\sqrt{2\ln k} \\
2[\Phi(C\sqrt{n}/\sigma + \sqrt{2\ln k}) - 1], & \text{if } C \leq \frac{\sigma}{\sqrt{n}}\sqrt{2\ln k}
\end{cases}$

$M_1 = \begin{cases} 
\Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}], & \text{if } C > \frac{\sigma}{\sqrt{n}}\sqrt{2\ln k} \\
0, & \text{if } C \leq \frac{\sigma}{\sqrt{n}}\sqrt{2\ln k}
\end{cases}$

$W_1 = \begin{cases} 
\Phi[-(\theta_0 + C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] - \Phi[-(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] + \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}], & \text{if } C > \frac{\sigma}{\sqrt{n}}\sqrt{2\ln k} \\
\Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}], & \text{if } C \leq \frac{\sigma}{\sqrt{n}}\sqrt{2\ln k}
\end{cases}$

$S_1 = \Phi[-(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] + \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}]$
The above probabilities as functions of $C$ are plotted in Figure 3.5 and Figure 3.6, with $k = 8, n = 100, \theta_0 = 0, \theta_1 = 1$ and $\sigma = 1$. In these two plots, we considered values of $C$ ranging from 0.05 to 0.95. $M_0$ monotonically decreases with $C$ and $M_1$ monotonically increases with $C$, although both $M_0$ and $M_1$ are very low for all values of $C$ considered. In addition, as $C$ increases, $S_0$ increases and $S_1$ decreases. From the mathematical forms of the operational characteristics, we see that when the sample size $n$ is considered fixed, $C$ has to be at least $\frac{\sigma}{\sqrt{n}} \sqrt{2 \ln k}$ in order to have a chance of observing strong evidence in support of the null composite hypothesis. However, it is worth emphasizing that $C$ should be chosen to define a set of parameter values that are not perceived differently from $\theta_0$, and the choice should be independent of the sample size.

Lastly, we plotted $M_1, W_1$ and $S_1$ as functions of $\theta_1$ in Figure 3.7, with $k = 8, n =$
$200, \theta_0 = 0, C = 0.2$ and $\sigma = 1$. The three probabilities as functions of $\theta_1$ are symmetric around $\theta_0 = 0$. As $|\theta_1 - \theta_0|$ gets larger, $M_1$ and $W_1$ decrease, and $S_1$ increases. Note that $M_0, W_0$ and $S_0$ are independent of $\theta_1$, and therefore, are not plotted as functions of $\theta_1$.

### 3.3 Theorem II: Asymptotic results for one-parameter parametric models

In this section, we will present operational characteristics associated with the GLR for one parameter parametric models when the sample size is large.

Assume $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables.
with common density or frequency function $f(x; \theta)$ where $\theta$ is a real-valued parameter. Define $\ell(x; \theta) = \ln f(x; \theta)$ and let $\ell'(x; \theta), \ell''(x; \theta),$ and $\ell'''(x; \theta)$ denote the first three partial derivatives of $\ell(x; \theta)$ with respect to $\theta$. We will assume the following regularity conditions (Lehmann, 1983) about $f(x; \theta)$:

(R1) The parameter space $\Theta$ is an open subset of the real line.
(R2) The set $A = \{x : f(x; \theta) > 0\}$ does not depend on $\theta$.
(R3) $f(x; \theta)$ is three times continuously differentiable with respect to $\theta$ for all $x$ in A.
(R4) $E_\theta[\ell'(X_i; \theta)] = 0$ for all $\theta$ and $\text{Var}_\theta[\ell'(X_i; \theta)] = \mathcal{I}(\theta)$ where $0 < \mathcal{I}(\theta) < \infty$ for all $\theta$.
(R5) $E_\theta[\ell'''(X_i; \theta)] = -\mathcal{J}(\theta)$ where $0 < J(\theta) < \infty$ for all $\theta$.
(R6) For each $\theta$ and $\delta > 0, |\ell'''(x; t)| \leq M(x)$ for $|\theta - t| \leq \delta$ where $E_\theta[M(X_i)] < \infty$.

In addition, we assume that we can differentiate twice inside the integral $\int_A f(x; \theta)dx$, which leads to $\mathcal{I}(\theta) = J(\theta)$. 

Figure 3.7: $M_1, W_1$ and $S_1$ are plotted as functions of $\theta_1$
Chapter 3. Generalized Likelihood Ratio for Composite Hypotheses

\[ M_0 = 2\Phi(-C\sqrt{nI(\theta_0)} - \sqrt{2\ln k}) \]

\[ W_0 = \begin{cases} 
2[\Phi(C\sqrt{nI(\theta_0)} + \sqrt{2\ln k}) - \Phi(C\sqrt{nI(\theta_0)} - \sqrt{2\ln k})], & \text{if } n > \frac{2\ln k}{C^2I(\theta_0)} \\
2[\Phi(C\sqrt{nI(\theta_0)} + \sqrt{2\ln k})] - 1, & \text{if } n \leq \frac{2\ln k}{C^2I(\theta_0)} 
\end{cases} \]

\[ S_0 = \begin{cases} 
2\Phi(C\sqrt{nI(\theta_0)} - \sqrt{2\ln k}) - 1, & \text{if } n > \frac{2\ln k}{C^2I(\theta_0)} \\
0, & \text{if } n \leq \frac{2\ln k}{C^2I(\theta_0)} 
\end{cases} \]

\[ M_1 = \begin{cases} 
\Phi[(\theta_0 + C - \theta_1)\sqrt{nI(\theta_1)} + \sqrt{2\ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{nI(\theta_1)} + \sqrt{2\ln k}], & \text{if } n > \frac{2\ln k}{C^2I(\theta_1)} \\
0, & \text{if } n \leq \frac{2\ln k}{C^2I(\theta_1)} 
\end{cases} \]

\[ W_1 = \begin{cases} 
+\Phi[(\theta_0 - C - \theta_1)\sqrt{nI(\theta_1)} + \sqrt{2\ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{nI(\theta_1)} - \sqrt{2\ln k}], & \text{if } n > \frac{2\ln k}{C^2I(\theta_1)} \\
\Phi[(\theta_0 + C - \theta_1)\sqrt{nI(\theta_1)} + \sqrt{2\ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{nI(\theta_1)} - \sqrt{2\ln k}], & \text{if } n \leq \frac{2\ln k}{C^2I(\theta_1)} 
\end{cases} \]

\[ S_1 = \Phi[-(\theta_0 + C - \theta_1)\sqrt{nI(\theta_1)} - \sqrt{2\ln k}] + \Phi[(\theta_0 - C - \theta_1)\sqrt{nI(\theta_1)} - \sqrt{2\ln k}] \]

A proof is provided in Appendix A.2.

The above asymptotic results show that the operational characteristics for one-parameter parametric models have very similar functional forms to the ones we observed for the Normal probability model. The Normal distribution \( N(\theta, \sigma^2) \) with \( \sigma^2 \) assumed known is an example of a one-parameter parametric model, and its Fisher’s information is given by \( \frac{1}{\sigma^2} \), which does not depend on the unknown parameter, \( \theta \). Substituting \( I(\theta_0) \) and \( I(\theta_1) \) by \( \frac{1}{\sigma^2} \), the asymptotic results shown in Theorem II for one-parameter parametric models reduce to the ones in Theorem I for the Normal distribution, as expected. Using similar arguments as for the Normal distribution case, we also show that \( M_0 \) and \( M_1 \) are bounded by \( 2\Phi(-\sqrt{2\ln k}) \) and \( \Phi(-\sqrt{2\ln k}) \), respectively.

**Example 3.1.** The Poisson distribution
Let $X_1, X_2, ..., X_n$ denote independent and identically distributed random variables from the Poisson ($\theta$) distribution with $\theta > 0$, that is,

$$f(x; \theta) = \frac{\theta^x e^{-\theta}}{x!}, \theta > 1$$

Fisher’s information for the Poisson ($\theta$) distribution is $I(\theta) = \frac{1}{\theta}$, and the operational characteristics are now

$$M_0 = 2\Phi(-C\sqrt{\frac{n}{\theta_0}} - \sqrt{2\ln k})$$

$$W_0 = \begin{cases} 
2[\Phi(C\sqrt{\frac{n}{\theta_0}} - \sqrt{2\ln k}) - \Phi(C\sqrt{\frac{n}{\theta_1}} - \sqrt{2\ln k})], & \text{if } n > \frac{2\theta_0\ln k}{C^2} \\
2[\Phi(C\sqrt{\frac{n}{\theta_0}} + \sqrt{2\ln k})] - 1, & \text{if } n \leq \frac{2\theta_0\ln k}{C^2}
\end{cases}$$

$$S_0 = \begin{cases} 
2\Phi(C\sqrt{\frac{n}{\theta_0}} - \sqrt{2\ln k}) - 1, & \text{if } n > \frac{2\theta_0\ln k}{C^2} \\
0, & \text{if } n \leq \frac{2\theta_0\ln k}{C^2}
\end{cases}$$

$$M_1 = \begin{cases} 
\Phi[\frac{(\theta_0 + C - \theta_1)}{\sqrt{\frac{n}{\theta_1}} - \sqrt{2\ln k}}] - \Phi[\frac{(\theta_0 - C - \theta_1)}{\sqrt{\frac{n}{\theta_1}} + \sqrt{2\ln k}}], & \text{if } n > \frac{2\theta_1\ln k}{C^2} \\
0, & \text{if } n \leq \frac{2\theta_1\ln k}{C^2}
\end{cases}$$

$$W_1 = \begin{cases} 
\Phi[\frac{(\theta_0 + C - \theta_1)}{\sqrt{\frac{n}{\theta_1}} + \sqrt{2\ln k}}] - \Phi[\frac{(\theta_0 - C - \theta_1)}{\sqrt{\frac{n}{\theta_1}} - \sqrt{2\ln k}}], & \text{if } n > \frac{2\theta_1\ln k}{C^2} \\
\Phi[\frac{(\theta_0 - C - \theta_1)}{\sqrt{\frac{n}{\theta_1}} + \sqrt{2\ln k}}] - \Phi[\frac{(\theta_0 - C - \theta_1)}{\sqrt{\frac{n}{\theta_1}} - \sqrt{2\ln k}}], & \text{if } n \leq \frac{2\theta_1\ln k}{C^2}
\end{cases}$$

$$S_1 = \Phi[\frac{(\theta_0 + C - \theta_1)}{\sqrt{\frac{n}{\theta_1}} - \sqrt{2\ln k}}] + \Phi[\frac{(\theta_0 - C - \theta_1)}{\sqrt{\frac{n}{\theta_1}} - \sqrt{2\ln k}}]$$

We used simulations to empirically compute the operational characteristics for the GLR assuming the data follows the Poisson ($\theta$) distribution. Specifically, we set $k = 8$, $\theta_0 = 3$, $C = 0.6$, and considered a range of $\theta_1$ values and three samples sizes, $n = 10, 50$ and $100$. The number of repetitions was set to be 5000. We also calculated the probabilities based on the analytical results shown above, for the same set of parameter
values. In Figure 3.8, we plotted the probabilities for \( n = 10 \). In the top panel, we present the probabilities when \( \theta_0 \) is assumed true. Given the small sample size, there is no chance of observing strong evidence, and the data can only provide weak evidence. Since \( M_0, W_0 \) and \( S_0 \) do not depend on \( \theta_1 \), these probabilities do not vary with the changing \( \theta_1 \) values. In the bottom panel, we plotted the operational characteristics when \( \theta_1 \) is assumed true. All three probabilities, \( M_1, W_1 \) and \( S_1 \), behave as expected based on both the simulations and analytical results, that is, as the distance between \( \theta_1 \) and \( \theta_0 \) gets larger, \( S_1 \) increases and \( W_1 \) decreases. There are considerable difference between the simulated and analytically derived results for \( M_1, W_1 \) and \( S_1 \). This is mainly due to the small sample size, where the asymptotic assumption does not hold and there are more variability in the simulation.

In Figure 3.9 and 3.10, we present the results for \( n = 50 \) and \( n = 100 \), respectively. With the larger sample sizes, both \( S_0 \) and \( S_1 \) increase. For example, \( S_0 \) is greater than
Figure 3.9: The operational characteristics are plotted as functions of the alternative simple hypothesis, $\theta_1$, for sample size of $n = 50$.

0.8 when $n = 100$, comparing to 0 and less than 0.4 for $n = 10$ and 50, respectively. In addition, the probabilities based on the simulations and analytical derivations get closer and they become nearly identical when $n = 100$.

### 3.4 Theorem III: Asymptotic results for multi-parameter parametric models

In this section, we derive the operational characteristics associated with the GLR for multi-parameter parametric models.

The model now assumes that $X_1, X_2, ..., X_n$ are independent and identically distributed from a $(g+1)$ - parameter model $\sim f(x; \theta, \gamma_1, ..., \gamma_g)$, where $\theta$ is the single or scalar parameter of interest, and $\gamma = (\gamma_1, \gamma_2, ..., \gamma_g)$ are the $g$ nuisance parameters. Since
Chapter 3. Generalized Likelihood Ratio for Composite Hypotheses

The operational characteristics are plotted as functions of the alternative simple hypothesis, \( \theta_1 \), for sample size of \( n = 100 \).

the generalized law of likelihood applies to models with a single parameter, it is necessary to eliminate the nuisance parameters. There is not a ‘best’ approach to eliminate nuisance parameters, although the use of conditional or marginal likelihood is generally preferred when they are available. Here we employ the profile likelihood approach which offers a more generally available solution. The full likelihood of the data is given by

\[
L(\theta, \gamma) = L(\theta, \gamma_1, \ldots, \gamma_g) = \prod_{i}^{n} f(x_i; \theta, \gamma_1, \ldots, \gamma_g)
\]

and the profile likelihood for the single parameter of interest, \( \theta \), is defined as

\[
L_P(\theta) = \max_{\gamma} [L(\theta, \gamma)] = L(\theta, \hat{\gamma}(\theta))
\]

Although a profile likelihood is not a true likelihood function, Royall (2000) demonstrated that for simple hypotheses, the profile likelihood leads to favorable operational characteristics asymptotically. Here we show that the same holds for EP with composite hypotheses.
Let us denote the Fisher’s information matrix for the \( g+1 \) parameters as \( \mathcal{I} \), and \( \mathcal{I} \) is of dimension \((g + 1) \times (g + 1)\),

\[
\mathcal{I} = \begin{bmatrix}
I_{\theta\theta} & I_{\theta\gamma_1} & \cdots & I_{\theta\gamma_g} \\
I_{\gamma_1\theta} & I_{\gamma_1\gamma_1} & \cdots & I_{\gamma_1\gamma_g} \\
\vdots & \vdots & \ddots & \vdots \\
I_{\gamma_g\theta} & I_{\gamma_g\gamma_1} & \cdots & I_{\gamma_g\gamma_g}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\mathbb{E}\left(-\frac{\partial^2 \ln f(x_i; \theta, \gamma_1, \ldots, \gamma_g)}{\partial \theta^2}\right) & \mathbb{E}\left(-\frac{\partial^2 \ln f(x_i; \theta, \gamma_1, \ldots, \gamma_g)}{\partial \theta \partial \gamma_1}\right) & \cdots & \mathbb{E}\left(-\frac{\partial^2 \ln f(x_i; \theta, \gamma_1, \ldots, \gamma_g)}{\partial \theta \partial \gamma_g}\right) \\
\mathbb{E}\left(-\frac{\partial^2 \ln f(x_i; \theta, \gamma_1, \ldots, \gamma_g)}{\partial \gamma_1 \partial \theta}\right) & \mathbb{E}\left(-\frac{\partial^2 \ln f(x_i; \theta, \gamma_1, \ldots, \gamma_g)}{\partial \gamma_1^2}\right) & \cdots & \mathbb{E}\left(-\frac{\partial^2 \ln f(x_i; \theta, \gamma_1, \ldots, \gamma_g)}{\partial \gamma_1 \partial \gamma_g}\right) \\
\vdots & \vdots & \ddots & \vdots \\
\mathbb{E}\left(-\frac{\partial^2 \ln f(x_i; \theta, \gamma_1, \ldots, \gamma_g)}{\partial \gamma_g \partial \theta}\right) & \mathbb{E}\left(-\frac{\partial^2 \ln f(x_i; \theta, \gamma_1, \ldots, \gamma_g)}{\partial \gamma_g \partial \gamma_1}\right) & \cdots & \mathbb{E}\left(-\frac{\partial^2 \ln f(x_i; \theta, \gamma_1, \ldots, \gamma_g)}{\partial \gamma_g^2}\right)
\end{bmatrix}
\]

where \( \mathcal{I}_{\theta\theta} = \mathcal{I}_{\theta\theta} \), \( \mathcal{I}_{\theta\gamma} = \begin{bmatrix} I_{\theta\gamma_1} & \cdots & I_{\theta\gamma_g} \end{bmatrix} \), \( \mathcal{I}_{\gamma_\theta} = \begin{bmatrix} I_{\gamma_1\theta} \\ \vdots \\ I_{\gamma_g\theta} \end{bmatrix} \), and \( \mathcal{I}_{\gamma\gamma} = \begin{bmatrix} I_{\gamma_1\gamma_1} & \cdots & I_{\gamma_1\gamma_g} \\ \vdots & \ddots & \vdots \\ I_{\gamma_g\gamma_1} & \cdots & I_{\gamma_g\gamma_g} \end{bmatrix} \).

For notational simplicity, let us further define \( \rho_{\theta\gamma}^2 = \frac{\mathcal{I}_{\theta\gamma_\gamma}^{-1}\mathcal{I}_{\gamma_\theta}}{\mathcal{I}_{\theta\theta}} \), and note that \( \rho_{\theta\gamma}^2 \) is a scalar.

Assuming the same regularity conditions (Lehmann, 1983) as for the one-parameter
parametric models, we were able to show that

\[ M_0 = 2\Phi(-C \sqrt{n } \Theta_0 \Theta_0 (1 - \rho_{\theta_0}) - \sqrt{2 \ln k}) \]

\[ W_0 = \begin{cases} 
2[\Phi(C \sqrt{n } \Theta_0 \Theta_0 (1 - \rho_{\theta_0}) + \sqrt{2 \ln k})] & \text{if } n > \frac{2\ln k}{C^2 n \Theta_0 \Theta_0 (1 - \rho_{\theta_0})} \\
-\Phi(C \sqrt{n } \Theta_0 \Theta_0 (1 - \rho_{\theta_0}) - \sqrt{2 \ln k}) & \text{if } n \leq \frac{2\ln k}{C^2 n \Theta_0 \Theta_0 (1 - \rho_{\theta_0})}
\end{cases} \]

\[ S_0 = \begin{cases} 
\Phi(C \sqrt{n } \Theta_0 \Theta_0 (1 - \rho_{\theta_0}) - \sqrt{2 \ln k}) & \text{if } n > \frac{2\ln k}{C^2 n \Theta_0 \Theta_0 (1 - \rho_{\theta_0})} \\
-\Phi(-C \sqrt{n } \Theta_0 \Theta_0 (1 - \rho_{\theta_0}) + \sqrt{2 \ln k}) & \text{if } n \leq \frac{2\ln k}{C^2 n \Theta_0 \Theta_0 (1 - \rho_{\theta_0})}
\end{cases} \]

\[ M_1 = \begin{cases} 
\Phi[(\theta_0 + C - \theta_1) \sqrt{n } \Theta_0 \Theta_0 (1 - \rho_{\theta_0}) - \sqrt{2 \ln k}] & \text{if } n > \frac{2\ln k}{C^2 n \Theta_0 \Theta_0 (1 - \rho_{\theta_0})} \\
-\Phi[(\theta_0 + C - \theta_1) \sqrt{n } \Theta_0 \Theta_0 (1 - \rho_{\theta_0}) + \sqrt{2 \ln k}] & \text{if } n \leq \frac{2\ln k}{C^2 n \Theta_0 \Theta_0 (1 - \rho_{\theta_0})}
\end{cases} \]

\[ W_1 = \begin{cases} 
\Phi[(-\theta_0 + C - \theta_1) \sqrt{n } \Theta_0 \Theta_0 (1 - \rho_{\theta_0}) + \sqrt{2 \ln k}] & \text{if } n > \frac{2\ln k}{C^2 n \Theta_0 \Theta_0 (1 - \rho_{\theta_0})} \\
-\Phi[(-\theta_0 + C - \theta_1) \sqrt{n } \Theta_0 \Theta_0 (1 - \rho_{\theta_0}) - \sqrt{2 \ln k}] & \text{if } n \leq \frac{2\ln k}{C^2 n \Theta_0 \Theta_0 (1 - \rho_{\theta_0})}
\end{cases} \]
Chapter 3. Generalized Likelihood Ratio for Composite Hypotheses 55

\[ S_1 = \Phi \left[ - (\theta_0 + C - \theta_1) \sqrt{n I_{\theta_0,\theta_1} (1 - \rho^2_{\theta_1,\gamma})} - \sqrt{2 \ln k} \right] \\
+ \Phi \left[ (\theta_0 - C - \theta_1) \sqrt{n I_{\theta_1,\theta_1} (1 - \rho^2_{\theta_1,\gamma})} - \sqrt{2 \ln k} \right] \]

A proof is provided in A.3

When \( \theta \) and \( \gamma \) are information orthogonal (i.e. \( I_{\theta\gamma} = 0 \) and \( I_{\gamma\theta} = 0 \)), \( \rho^2_{\theta_0\gamma} = 0 \), and the above results reduce to the ones shown in section 3.3 for one-parameter parametric models. Furthermore, the same upper bounds of \( 2\Phi(-\sqrt{2\ln k}) \) and \( \Phi(-\sqrt{2\ln k}) \) hold for \( M_0 \) and \( M_1 \), respectively.

### 3.5 Implications for study planning

As we discussed earlier in the introduction, there are fundamental differences between \( M_0 \) under the \( EP^S \) and the type I error rate under the Neyman-Pearson framework, despite their similar mathematical forms. The same holds for \( M_0 \) under the \( EP^C \); it also varies with the sample size instead of being fixed at a particular value. To plan a study under the \( EP^C \), we need to ensure the probabilities of observing weak (\( W_0, W_1 \)) and misleading evidence (\( M_0, M_1 \)) are low, while the probabilities of observing strong evidence (\( S_0, S_1 \)) are high. We have shown that \( M_0 \) and \( M_1 \) are bounded by \( 2\Phi(-\sqrt{2\ln k}) \) and \( \Phi(-\sqrt{2\ln k}) \), and even with \( k = 8 \), the two probabilities are naturally low and can not exceed 0.0414 and 0.0207, respectively. Therefore, for most studies, it is sufficient to ensure that the probability of weak evidence is low, or equivalently, the probability of strong evidence is high.

When planning a study under the \( EP^C \), one needs to first specify the two simple parameter values: \( \theta_0 \) - the simple null hypothesis value, and \( \theta_1 \) - the simple alternative
Chapter 3. Generalized Likelihood Ratio for Composite Hypotheses 56

hypothesis value. These two values can be selected in the same way as if one plans a study under the Neyman-Pearson or $E P^S$. Then the threshold value $k$ for declaring strong evidence needs to be chosen to ensure that the maximum probability of observing misleading evidence is low. Under $E P^C$, one additional parameter value, $C$, which is used to define the two composite hypotheses, also needs to be specified. The additional planning parameter, $C$, should generally be chosen to define a null hypothesis interval where all parameter values within the null interval are perceived not different qualitatively from the simple null hypothesis, $\theta_0$. With these parameters specified, we can then use the results for the operational characteristics derived earlier in this chapter to determine the minimum sample size required to ensure that $\min\{S_0, S_1\} > \Lambda$, where $\Lambda$ represents the minimum probability of strong evidence desired.

3.5.1 Sample size calculation for the Normal probability model

Strug et. al. (Strug et al., 2007) formulated study planning under $E P^S$ for Normal response data. For the simple case of comparing two hypothesized means of a sample of independent Normally distributed observations, the authors derived a relationship between the error probabilities under the $E P^S$ and the Neyman-Pearson framework and showed that a larger sample size is generally required under $E P^S$ in order to have the same power and probability of strong evidence. Here, for the simple case of comparing two Normal means, we compare sample size calculations between $E P^S$ and $E P^C$. Since the probabilities of misleading evidence, $M_0$ and $M_1$, vary with parameter values and sample size under both $E P^S$ and $E P^C$, we can not set these probabilities at the same level for the comparison. Here, we simply use the same threshold value of $k = 8$ for both methods, and compare the minimum sample size needed in order to have 80% chance of observing strong evidence. Without loss of generality, we consider $\theta_0 = 0, \theta_1 = 0.8$ and
Chapter 3. Generalized Likelihood Ratio for Composite Hypotheses

Figure 3.11: The probabilities of observing strong evidence in support of the simple alternative hypothesis of \( \theta = \theta_1 \) or the composite alternative that \( \theta \in \Theta_1 = (-\infty, \theta_0 - C) \cup (\theta_0 + C, \infty) \) are plotted as functions of the sample size \( n \).

\[ \sigma = 1. \]

We first consider the scenario where one is only interested in observing strong evidence in support of the alternative hypothesis, either the simple alternative hypothesis of \( \theta = \theta_1 \) or the composite alternative that \( \theta \in \Theta_1 = (-\infty, \theta_0 - C) \cup (\theta_0 + C, \infty) \). Under \( EP^S \), the probability of observing strong evidence in support of \( \theta_1 \) when \( \theta_1 \) is true is \( S^S_1 = \Phi(\frac{\Delta \sqrt{n}}{2\sigma} - \frac{\sigma \ln k}{\Delta \sqrt{n}}) = \Phi(\frac{0.8\sqrt{n}}{2\sigma} - \frac{8}{0.8\sqrt{n}}) \). Setting \( S_1 \geq 80\% \), it is straightforward to show that \( n \) has to be at least 15. Under \( EP^C \), we need to specify one additional parameter value, \( C \), which defines the null and alternative hypothesis intervals, and \( S^C_1 = \Phi[-(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] + \Phi((\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] \). Here we consider 4 different values of \( C = 0.05, 0.2, 0.3 \) and 0.4 for the comparison. The proba-
bilities of strong evidence, \( S_1^S \) and \( S_1^C \), are plotted as functions of sample size in Figure 3.11, for \( C = 0.05, 0.2, 0.3 \) and 0.4, respectively. When \( C \) is very small at 0.05, the null composite hypothesis is a very tight interval around the simple null hypothesis, and the minimum sample sizes required to have an 80% chance of observing strong evidence are exactly the same at \( n = 15 \) under both frameworks. As \( C \) increases, the null interval under \( EP^C \) becomes wider, and a larger sample size is needed in order to have the same high probability of observing strong evidence in support of \( \Theta_1 \). For example, when \( C = 0.4 \), we need a sample size of at least 52 in order to have an 80% chance of observing strong evidence in support of \( \Theta_1 \), a much larger sample size than 15.

If one is also interested in detecting strong evidence in support of the null composite hypothesis \( \Theta_0 \), \( S_0 \) also needs to be taken into account. Recall that for Normally distributed data, \( M_0^S = M_1^S, W_0^S = W_1^S \) and \( S_0^S = S_1^S \) under \( EP^S \), however, the same symmetry does not hold under \( EP^C \). Here, we further require \( S_0^C \geq 80\% \). We plotted \( S_0^C, S_1^C \) and \( S_0^S = S_1^S \) in Figure 3.12 for the same values of \( C \). When \( C = 0.05 \), it is not possible to observe strong evidence in support of \( \Theta_0 = [-0.05, 0.05] \) for any sample size considered. As \( C \) increases or the null hypothesis interval gets wider, the probability of observing strong evidence in support of the null composite hypothesis gets larger, and in turn, the minimum sample size required to have \( S_0^C \geq 80\% \) gets smaller. When \( C = 0.4 \), a sample size of \( n = 69 \) is required in order to have 80% chance of observing strong evidence in support of \( \Theta_0 \) or \( \Theta_1 \).

Holding sample size \( n \) constant, there is clearly a trade-off between \( S_0^C \) and \( S_1^C \) with respect to \( C \). That is, \( S_0^C \) is an increasing function of \( C \) and \( S_1^C \) is a decreasing function of \( C \). Therefore, the only way to increase both \( S_0^C \) and \( S_1^C \) is to increase the sample size.
Figure 3.12: The probabilities of observing strong evidence, $S^C_0$, $S^C_1$ and $S^S_0 = S^S_1$ are plotted against sample size $n$. 
3.6 Discussion and future work

In summary, we analytically derived the operational characteristics for the Evidential Paradigm with complementary composite hypotheses, which provided theoretical justification for using the generalized likelihood ratio to measure statistical evidence in support of one composite hypothesis over the other. Our results demonstrated that the extended EP framework with complementary composite hypotheses is operationally sound and does not lead one to incorrect conclusions often.

In this chapter, we often presented the operational characteristics as functions of the sample size n, while holding all other parameters fixed. But it is obvious that given the derived results, we can study the behavior of any one parameter while holding the rest fixed. For example, as a function of the parameter C, the probability of observing strong evidence in support of the null composite hypothesis, $S_0$, decreases when C gets smaller. In the extreme case where $C \to 0$, the composite interval null hypothesis reduces approximately to the point null hypothesis, and $S_0 \approx P_0\left(\frac{L(\theta_0)}{L(\theta)} \geq k\right) = P_0\left(-2 \ln \frac{L(\theta_0)}{L(\theta)} \leq -2 \ln k\right) = P_0(\chi^2_1 < -\ln k) = 0.$ This is essentially a likelihood ratio test under the Frequentist framework of testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, which could never support or ‘accept’ the point null hypothesis of $\theta = \theta_0$.

Here we focused on composite hypotheses that are complementary to each other due to their particular relevance to genetic association studies; however, applications of this particular set-up of complementary composite hypotheses are not limited to genetic association studies. For example, it can be applied to generic research questions that are commonly framed as a hypothesis testing problem in the form of $\theta = \theta_0$ versus $\theta \neq \theta_0$. In the original extension proposed by Bickel (2012) and Zhang and Zhang (2013), there are no restrictions on the type of or the relationship between the composite hypotheses, and the two hypotheses do not need to be complementary to each other. Take the clinical
trial example presented earlier in Chapter 1 for instance, the two composite hypotheses, ’inferiority’ versus ’non-inferiority’, are not complementary. It is of great interest to study and understand how the GLR behaves for two generic composite hypotheses so that the extended EP framework is applicable to more study designs. However, analytical derivations of the operational characteristics for genetic composite hypotheses are challenging, and simulation studies may be needed to assess the validity of using the GLR in such cases.

Except for the Normal probability model, the analytical results we presented in this chapter rely on the asymptotic assumptions. In practice though, it is often difficult to determine whether a particular dataset under study meets these assumptions. It is, therefore, important to know how robust the derived results are to violations of these assumptions, for example, when the sample size is small or the likelihood function is highly skewed. As future work, we will empirically assess the robustness of operational characteristics for the $EP^{C}$ through extensive simulations.
Chapter 4

Application of the Generalized Likelihood Ratio in Genetic Association Studies

In this chapter, we focus on applications of the composite hypothesis EP framework in genetic associated studies. In section 4.1, we derive formulas for sample size calculation for a typical genetic association study with a quantitative phenotype. In section 4.2, we demonstrate the difference between a Frequentist method and the GLR in a simulated large-scale genetic association study and highlight the advantages of using the GLR in such studies.

The sample size used in genetic studies is becoming larger and larger. One problem with the Frequentist paradigm is that given large enough sample size, any minimal effect size could become statistically significant. Even though these markers are statistically significant, they may have minimal effect on the phenotype, and it is not feasible to have sufficient data to replicate these findings. Therefore, it might be beneficial to ‘filter out’ these markers that are practically insignificant.
4.1 Sample size calculation for a GWAS with a quantitative trait

In section 3.4, we provided analytical results for the operational characteristics associated with the use of the GLR in multi-parameter parametric models. The probabilities are shown as functions of the elements of the Fisher information matrix. Here we provide the details for sample size calculations for a GWAS with a Normally distributed phenotype.

Given a random sample of size n, let $Y_i$ and $X_i$ denote the phenotypic and genotypic value for individual i, where $i = 1,..., n$. Then we can use a simple linear regression to model the relationship between $Y_i$ and $X_i$,

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

where $\epsilon_i \sim N(0, \sigma^2)$ and $\sigma^2$ is assumed known.

Here, we have one parameter of interest, $\beta$, which measures the effect of $X$ on $Y$, and one nuisance parameter $\alpha$. For genetic association studies, the simple null hypothesis value for $\beta$ is usually chosen to be $\beta_0 = 0$, representing no association, and the simple alternative hypothesis value, $\beta_1$, is often chosen to be the minimum effect size that one aims to detect. The null composite hypothesis can then be specified as a small interval $[-C, C]$ that contains the simple null hypothesis value of 0, where $C$ is a small but positive constant, therefore, all parameter values within $[-C, C]$ are considered not different from 0. In this case, $C$ has to be less than $|\beta_1|$ by design.

It is straightforward to show that Fisher’s information matrix can be re-expressed in terms of the genotype data $X$, the sample size $n$, and the known variance $\sigma^2$. 
\[
I_{\beta,\alpha} = \begin{bmatrix}
I_{\beta\beta} & I_{\beta\alpha} \\
I_{\alpha\beta} & I_{\alpha\alpha}
\end{bmatrix} = \begin{bmatrix}
\frac{\sum X_i^2}{\sigma^2} & \frac{\sum X_i}{\sigma^2} \\
\frac{\sum X_i}{\sigma^2} & \frac{n}{\sigma^2}
\end{bmatrix}
\]

Let us further define

\[SSX = \sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2\]

In the context of genetic association studies, \(\frac{1}{n-1}SSX\) can be interpreted as the sample variance of the genetic marker. Here we assume the genetic marker X has an additive effect on the phenotype Y, that is, each additional minor allele increases or decreases the value of Y by \(\beta\). Without loss of generality, we also assume the coding for the genetic marker is X = 0, 1 or 2, corresponding to carrying 0, 1 or 2 copies of the minor allele. Denoting the minor allele frequency (MAF) for the marker as \(f\), the total sample size as \(n\), then the number of samples, \(n_0, n_1\) and \(n_2\), carrying 0, 1 or 2 copies of the minor allele (i.e. \(X = 0, 1\) or 2), respectively, can be approximated by

\[n_0 = n(1 - f)^2\]
\[n_1 = 2nf(1 - f)\]
\[n_2 = nf^2\]

Then \(\bar{X}\) can also be expressed in terms of the MAF, \(f\), as

\[\bar{X} = \frac{n_0 + n_1 + n_2}{n_0 + n_1 + n_2} = \frac{n_1 + 2n_2}{n} = 2f\]

and similarly,

\[SSX = \sum (X_i - \bar{X})^2 = n_0(0 - \bar{X})^2 + n_1(1 - \bar{X})^2 + n_2(2 - \bar{X})^2 = 2nf(1 - f)\]

Then using results for the multi-parameter parametric models (section 3.4; Theorem III), we can show that
$\sqrt{nI_{\beta_0\beta_0}(1 - \rho_{\beta_0\alpha_0}^2)}$

$= \sqrt{nI_{\beta_1\beta_1}(1 - \rho_{\beta_1\alpha_1}^2)}$

$= \sqrt{\sum X_i^2/\sigma^2 - (\sum X_i)^2/n\sigma^2}$

$= \sqrt{SSX/\sigma^2}$

$= \sqrt{2nf(1-f)/\sigma^2}$

With these results, the operational characteristics now depend on the following parameter values:

- $\beta_0$ and $\beta_1$ - the two simple hypothesis values for the parameter of interest $\beta$
- $C$ - the parameter value that defines the null and alternative composite hypotheses
- $\sigma^2$ - the variance of the error term, which is assumed known
- $f$ - minor allele frequency of the genetic marker
- $k$ - the pre-specified threshold for declaring strong evidence

The operational characteristics, as functions of sample size, now simplify to

$M_0 = 2\Phi(-C\sqrt{2nf(1-f)/\sigma^2}) - \sqrt{2\ln k}$

$W_0 = \begin{cases} 
2[\Phi(C\sqrt{2nf(1-f)/\sigma^2}) + \sqrt{2\ln k}] - \Phi(C\sqrt{2nf(1-f)/\sigma^2}) - \sqrt{2\ln k}], & \text{if } n > \frac{\ln k}{(\frac{\sigma^2}{f})^{1-f}} \\
2[\Phi(C\sqrt{2nf(1-f)/\sigma^2}) + \sqrt{2\ln k})] - 1, & \text{if } n \leq \frac{\ln k}{(\frac{\sigma^2}{f})^{1-f}}
\end{cases}$
Let us first explore the relationship between the operational characteristics and the MAF. Note that, as a function of MAF, each of $W_0, S_0, M_1$ and $W_1$ is described by two different functions, depending on the cut-off value $f^*$, where $f^*(1-f^*) = \frac{\ln k}{(\frac{k}{r})^2 f(1-f)}$. Since $f^*$ denotes the MAF and is no greater than 0.5, there is a unique solution to the equation. In Figure 4.1, we plot the operational characteristics as functions of the MAF, while other parameters are fixed at $\beta_0 = 0, \beta_1 = 1, \sigma^2 = 1, C = 0.3, k = 8$ and $n = 200$. With this sets of parameter values, the cut-off value for $f$ is 0.13. In general, the probabilities of observing misleading evidence, $M_0$ and $M_1$, are both very small and close to 0 across all ranges of MAF. The probability of observing weak evidence decreases and the probability of observing strong evidence increases as MAF increases from 0.05 to 0.5. When $\beta_1 = 1$ is assumed true, $W_1$ is small ($< 0.2$) even when MAF is low, and it quickly approaches 0 as MAF increases. $S_1$ is greater than 0.8 at MAF=0.05, and approaches 1 as MAF increases.
increases. On the other hand, when $\beta_1 = 0$ is assumed true, the data can only provide weak evidence (i.e. $W_0 \approx 1$) when MAF is $< 0.13$. As MAF increases beyond 0.13, $W_0$ starts to decrease and the probability of observing strong evidence in support of the composite null hypothesis ($S_0$) starts to increase.

In Figure 4.2, the operational characteristics are plotted as functions of $C$, while other parameters are fixed at $\beta_0 = 0, \beta_1 = 1, \sigma^2 = 1, f = 0.1, k = 8$ and $n = 200$. Here we consider all values of $C$ ranging from 0.1 to 0.9 ($C$ has to be less than $\beta_1 = 1$ by design). The cut-off value for $C$, where $W_0, S_0, M_1$ and $W_1$ are described by two different functions, is given by $C^* = +\sqrt{\frac{\sigma^2 \ln k}{nf(1-f)}}$. With the chosen parameter values, $C^* = 0.25$. When $C$ is smaller than 0.25, i.e. the null hypothesis interval is only as large as $[-0.25, 0.25]$, and the data can only provide weak evidence ($W_0 \approx 1$) in support of $\beta \in [-C, C]$. As $C$ gets larger and the null interval gets wider, $W_0$ decreases to 0 and $S_0$ increases to 1.

Lastly, we plotted the operational characteristics as functions of the sample size in Figure 4.3, while fixing other parameters at $\beta_0 = 0, \beta_1 = 1, \sigma^2 = 1, C = 0.3, k = 8$ and
0.2 0.4 0.6 0.8
0.0 0.2 0.4 0.6 0.8 1.0
M0
W0
S0
\beta_0 = 0, \beta_1 = 1, \sigma = 1, \text{MAF} = 0.1, N = 200
C = 0.25
Probability
0.2 0.4 0.6 0.8
0.0 0.2 0.4 0.6 0.8 1.0
M1
W1
S1
\beta_0 = 0, \beta_1 = 1, \sigma = 1, \text{MAF} = 0.1, N = 200
C = 0.25
Probability

Figure 4.2: The operational characteristics are plotted against C

$f = 0.1$. As expected, both $S_0$ and $S_1$ increase with n, although there is no chance of observing strong evidence in support of the composite hypothesis that $\beta \in [-C, C]$ until n is greater than 257. With the given parameter values, in order to have $S_1 \geq 80\%$, we can calculate that n needs to be at least 95; and if we further require $S_0 \geq 80\%$, a much larger sample size of n = 681 is required.

4.2 Strength of the GLR in genetic association studies: a simulation study

Using a simple simulated genetic association study with a quantitative trait, we demonstrate the differences in the assessment of association evidence between the GLR and a commonly used Frequentist method, the likelihood ratio test. This simulation study also highlights some of the advantages of using the GLR in large-scale genetic association studies.
4.2.1 Simulation design

For the simulation study, we used PLINK (Purcell et al., 2007) to simulate a population of 10000 individuals and a small region with 100 SNPs. Assuming an additive genetic effect, the data were simulated under a simple linear regression model,

\[ Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \]

where \( Y_i \) represents the quantitative phenotype, \( X_i \) the genetic marker with an additive effect coding, and \( \epsilon_i \sim N(0, 1) \). Specifically, we assumed 10 markers were independently associated with the quantitative phenotype, with varying effect sizes, \( \beta_1 \), ranging from 0.10 to 0.22. The remaining 90 markers were simulated under the null, with \( \beta_1 = 0 \). A reference allele frequency was randomly generated from the uniform \([0.05, 0.95]\) distribution and assigned to each marker.

We created two datasets by randomly sampling \( n = 4000 \) or \( 8000 \) individuals from the population with 10000 individuals. We first analyzed the two datasets as well as the entire population using the GLR with \( k = 32 \). For comparison, we also performed the
Frequentist likelihood ratio test with a significance threshold of $\alpha = \frac{0.05}{100} = 0.005$, implementing a Bonferroni multiple-testing correction. When calculating the GLR, we set $C = 0.15$, which implies that any $\beta_1$ estimate that falls within [-0.15, 0.15] are considered not different qualitatively from 0. That is, a genetic marker does not have a practically significant effect on the quantitative phenotype if carrying an additional minor allele of this marker does not increase or decrease the mean of the phenotype by 15% of the residual variance ($\sigma^2 = 1$). In Figures 4.4 - 4.6, we present the results for $n = 4000$, 8000 and 10000, respectively, and in each figure, we plot the association results from the GLR framework in the top panel and those from the likelihood ratio test in the bottom panel. It is worth noting that our aim here is not performing a sequential analysis, but simply demonstrating the differences in the two analytical approaches with different sample sizes.

### 4.2.2 Simulation results

In the top panel of Figure 4.4, we plotted the indices of the 100 SNPs on the x-axis and log10 of the GLR values on the y-axis. For easy tracking, we always plotted the 10 truly associated markers at the beginning, with indices of 1 to 10. Since we chose $k = 32$ as the threshold for declaring strong evidence, we also plotted log10(32) and log10(1/32), or equivalently -log10(32), as the dotted red and blue lines, respectively. Markers with $\text{GLR} \geq 32$ (or, equivalently, log10(GLR) $\geq 1.51$) are plotted in red, and at these markers, the data provide strong evidence in support of the alternative composite hypothesis that the underlying effect sizes are statistically and practically different from 0 (i.e. $\beta_1 \notin [-0.15, 0.15]$). Markers with $\text{GLR} \leq \frac{1}{32}$ (or -log10(GLR) $\leq -1.51$) are plotted in blue, and at these markers, the data provide strong evidence in support of the null composite hypothesis that the underlying effect sizes are minimal and fall within [-0.15, 0.15]. The gray points that fall between the dotted red and blue lines represent markers with $\frac{1}{32} < \text{GLR} < 32$, and therefore, indicate weak evidence. In the bottom panel of
Figure 4.4: Association results for the GLR and the likelihood ratio test with $n = 4000$

Figure 4.4, the indices of the 100 markers are plotted on the x-axis, and on the y-axis, we plotted the likelihood ratio test p-value on the -log10 scale. The higher the points are on the plot, the smaller the corresponding p-values are. The dotted red line was set at -log10$(0.05/100)$ = 3.30, corresponding to the Bonferroni multiple-testing corrected p-value threshold. This is essentially a ‘Manhattan’ plot with 100 markers.

At $n = 4000$ (Figure 4.4), 2 out of the 10 markers with underlying $\beta_1 \neq 0$ provided strong evidence in support of the alternative composite hypothesis. In additional, 4 out of these 10 markers provided strong support for the null composite hypothesis. This observation is not contradicting our simulation, as the underlying effect sizes for these 4 markers are small and their maximum likelihood estimates for the effect size are less than 0.15. For the remaining 90 markers simulated with $\beta_1 = 0$, the data provided strong evidence in support of the null composite hypothesis for almost all of them. Under the
Frequentist likelihood ratio test, 6 markers were statistically significant, providing evidence that the underlying effect sizes for these markers are different from the point null of $\beta_1 = 0$. However, no conclusion can be made for all the remaining markers.

At $n = 8000$ (Figure 4.5), more markers are detected as significant under both methods. When we finally analyzed the entire population with $n = 10000$ (Figure 4.6), all markers are completely demarcated under the $EPC$ as providing strong evidence in support of the null or the alternative composite hypothesis. While under the Frequentist method, all 10 markers with underlying $\beta_1 \neq 0$ were detected as statistically significant, regardless of the minimal effect sizes for some of the markers. The 90 markers simulated with underlying $\beta_1 = 0$ could never provide evidence in support of the point null hypothesis, even with such large sample size. This is because under the Frequentist framework, one can never ‘accept’ the point null hypothesis. So the major difference between the two frameworks is that, under the $EPC$, the 4 markers with minimal underlying effect sizes did not provide strong evidence in support of the alternative composite hypothesis; when the sample size gets larger, they actually provided strong evidence in support of the null composite hypothesis. While under the Frequentist method, they eventually became statistically significant, rejecting the simple null hypothesis of $\beta_1 = 0$.

4.3 Discussion and future work

We see from the simulation study that under the Frequentist framework, any small effect size can become statistically significant as sample size increases. This implies that with the ever-growing sample size in GWAS, a large number of markers with minimal and practically insignificant effect sizes will become statistically significant, which makes prioritization for replication and follow-up studies difficult. The use of GLR under $EPC$,
Chapter 4. Application of the GLR in Genetic Association Studies

Figure 4.5: Association results for the GLR and the likelihood ratio test with $n = 8000$

Figure 4.6: Association results for the GLR and the likelihood ratio test with $n = 10000$
on the other hand, can guard against these practically insignificant findings. Another fundamental difference between the Frequentist method and the GLR is that the Frequentist methods focus on rejecting the null hypothesis by design, and can never generate evidence to support the null hypothesis; therefore, one can never conclude that there is no association between the phenotype and a genetic marker, regardless of sample size. The GLR, in contrast, allows the data to provide strong evidence in support of either hypothesis, which improves fine mapping of the locus.

Given the large number of genetic markers tested in a typical GWAS, it is important to address the issue of multiple comparisons. Under the Evidential Paradigm, the measure of statistical evidence, the likelihood ratio, is decoupled from the error probability, the probability of observing misleading evidence. Therefore, it is the probability of observing misleading evidence that one needs to control, and the likelihood ratio itself is never adjusted regardless of the number of comparisons made. To account for multiple comparisons, it is common to consider the family-wise error rate (FWER), which is the probability of rejecting the null hypothesis of no association for at least one SNP, assuming all of the J SNPs under study are not associated. Under the Frequentist framework, one typically fixes the FWER at a pre-specified level, e.g., $\alpha = 0.05$, in contrast, the probabilities of misleading evidence, $M_0$ and $M_1$, vary with sample size and other parameters. Let us focus on $M_0$, the probability of observing strong evidence in support of the incorrect hypothesis of association when there is indeed no association at a single SNP, then $\text{FWER} \leq J \times M_0$, and the equality holds only when all J tests are independent. Given a fixed number of tests/SNPs, to reduce FWER, we only need to reduce $M_0$. It is easy to see from the results in Chapter 3 that $M_0$ can be decreased by increasing the sample size $n$, the threshold value $k$, the minor allele frequency of the SNP or the null hypothesis margin $C$. Strug et al. (2010) has argued that increasing the threshold $k$ is counterproductive, as this only minimally reduces $M_0$ but dramatically increases the
probability of observing weak evidence. The authors (Strug et al., 2010) showed that the 
most desirable and appropriate approach to lower FWER is to increase the sample size, 
by either planning a larger initial study or implementing a two-stage replication design. 
These recommendations were initially proposed for the EP with the simple versus simple 
hypothesis set-up, and as future work, we will investigate whether the same principles 
apply to the $EP^C$. 
Chapter 5

Concluding Remarks & Directions for Future Work

In summary, we have accomplished our goals of using likelihood-based methods to address common statistical issues in genetic association studies. We developed a conditional likelihood ratio-based measure, the maxLRc, for prioritizing rare sequence variants in an associated region. Analytical and simulation results show that the maxLRc is always well-defined and it outperforms the commonly used approach of ranking based on exact p-values in most simulated settings. In addition, we analytically derived the operational characteristics for the generalized likelihood ratio, providing theoretical grounding for extending the Evidential Paradigm to accommodate composite hypotheses \((EPC)\). We provided guidelines for study planning under the \(EPC\), and we demonstrated that data can support either the null or the alternative hypothesis under the \(EPC\), allowing association results to be more clearly demarcated. In addition, the use of a composite instead of a point null hypothesis can guard against statistically significant but practically insignificant findings in large-scale association studies.

In the next two sections, we discuss two related statistical issues that remain to be
addressed as future work.

5.1 Multiple parameters of interest

Although in this thesis so far, we have focused on parametric models with one parameter of interest, Royall (2000) showed that the nice operational characteristics hold for models with multiple parameters of interest. Specifically, let $\theta$ denote a fixed-dimensional vector parameter, with $\theta_1$ and $\theta_2$ represent the two simple vector hypotheses. For notational simplicity, let us write $\theta_2 = \theta_1 + (n\mathcal{I}(\theta_1))^{-1/2}c$, where $c$ is a vector of the same dimension as $\theta$, and $\mathcal{I}(\theta)$ is the information matrix, with $\mathcal{I}_{ij}(\theta) = -E\{[\partial^2 \ln f(X, \theta)]/\partial \theta_i \partial \theta_j\}$. Then the probability of observing misleading evidence is given by $\Phi(-\|c\|/2 - \ln k/\|c\|)$, with the maximum of $\Phi(-\sqrt{2\ln k})$. The option of having multiple parameters of interest could offer a favorable Evidential solution to the problem of statistically detecting pleiotropy in genetic association studies. Pleiotropy is the phenomenon where one genetic marker affects multiple phenotypes, therefore, simultaneous inference on multiple parameters in a model is required in such studies.

As one of my thesis projects, we set out to assess pleiotropy of modifier genes in Cystic Fibrosis (CF) (Li et al., 2014). CF, caused by mutations in the Cystic Fibrosis transmembrane conductance regulator (CFTR) gene, affects multiple organs, including the lungs, liver, pancreas and intestines. However, modifier genes contribute to variable disease severity across affected organs, even in individuals with the same CFTR genotype. We sought to determine whether SNPs that were previously identified as contributing to Meconium Ileus in CF, are pleiotropic for other early-affecting CF co-morbidities, including lung disease severity in the pediatric population, age at first acquisition of P. aeruginosa infection and early exocrine pancreatic disease. In this project, we faced two
major statistical challenges, (1) how to account for possible correlations in the phenotypes, and (2) how to statistically define pleiotropy under the commonly used Frequentist framework. To address (1), we first tested for phenotypic correlations between each pair of the phenotypes of interest. In the absence of phenotypic correlation, we performed the traditional single-phenotype association analysis with a SNP. Evidence of association in this case is also evidence of pleiotropy, because each SNP is independently associated with Meconium Ileus and the other phenotype. In the presence of phenotypic correlation, we used an alternative analytic approach to account for both the phenotypic correlation and the established Meconium Ileus-SNP association. We proposed a method analogous to the standard regional analysis of multiple SNPs surrounding a genome-wide significant SNP; in this case, strong association evidence at a nearby SNP does not necessarily imply locus heterogeneity. Instead, conditional analysis is often used to establish evidence for multiple independent loci within a region, after accounting for the inherent correlation between SNPs due to linkage disequilibrium (Gharavi et al., 2011). In this pleiotropy analysis, we combined the conditional analysis principle with the reverse regression approach (O’Reilly et al., 2012). Specifically, we reversed the role of phenotype and genotype so that the number of Meconium Ileus risk alleles of a SNP is regressed on Meconium Ileus and, for example, pancreatic disease severity at birth via ordinal logistic regression. The regression coefficient corresponding to pancreatic disease severity in this regression model captures the association between pancreatic disease severity and the SNP, after accounting for the correlation between pancreatic disease severity and Meconium Ileus, and association between the SNP and Meconium Ileus. Therefore, testing the significance of the pancreatic disease severity regression coefficient assessed the pleiotropic effect of the SNP.

The second challenge of statistically defining pleiotropy fundamentally lies in the focus on the null hypothesis under the Frequentist framework. Assuming a regression model
SNP = \beta_0 + \beta_1 X_1 + \beta_2 X_2,

where \(X_1\) and \(X_2\) represent two quantitative phenotypes, then the null hypothesis of ‘no pleiotropy’ is often translated into \(H_0 : \beta_1 = 0\) and \(\beta_2 = 0\), since \(\beta_1\) and \(\beta_2\) measure the association between the SNP and phenotype 1 and 2, respectively. However, rejecting the null hypothesis only implies that the genetic marker is associated with at least one of the two phenotypes, which does not guarantee pleiotropy. To address this issue, we restricted our analysis to SNPs with previously established association with Meconium Ileus, therefore, a significant association with another phenotype, while properly accounting for phenotypic correlations, indicates pleiotropy. Although this approach worked well for our project, it is not an optimal method for assessing pleiotropy. For example, it requires the analysis to be restricted to genetic markers with previously established association with a primary phenotype, in our case, Meconium Ileus. With more than two phenotypes of interest, additional sequential conditional analyses are needed.

The composite EP framework with multiple parameters of interest could offer a favorable solution to the problem of statistically assessing pleiotropy. Without loss of generality, let us assume the same model with two quantitative phenotypes of interest,

\[
SNP = \beta_0 + \beta_1 X_1 + \beta_2 X_2
\]

with \(\beta_1\) and \(\beta_2\) representing their association effect sizes with the SNP. Define a null composite hypothesis region, \(B_0 : \beta_1 \in [-C, C] \text{ and } \beta_2 \in [-C, C]\), where C is a small and positive constant, then \(B_0\) is a square with center 0 and side length \(2C\). \(B_0\) defines a region that represents ‘no pleiotropy’, since for all \(\beta_1 \in B_0\) and \(\beta_2 \in B_0\), \(-C \leq \beta_1 \leq C\) and \(-C \leq \beta_2 \leq C\). Define an alternative composite hypothesis region as \(B_1 = B_0^\complement\), the complement of the null composite hypothesis region, then all parameter values \((\beta_1, \beta_2) \in B_1\) indicate pleiotropy. Therefore, the generalized likelihood ratio

\[
\frac{\sup_{(\beta_1', \beta_2') \in B_1} L_p(\beta_1', \beta_2')} {\sup_{(\beta_1', \beta_2') \in B_0} L_p(\beta_1', \beta_2')},
\]
measures the strength of evidence in support of ‘pleiotropy’ over ‘no pleiotropy’. In future work, we will derive and examine the operational characteristics associated with this generalized likelihood ratio for inferring pleiotropy in genetic association studies.

5.2 Model misspecification

In all the analytical derivations and results we have shown so far, we assumed that the working model \( f \) is the correct model. What happens if the working model \( f \) is not the true model for the data? Can the likelihood ratios and the generalized likelihood ratios constructed from \( f \) still be used to interpret the strength of statistical evidence?

Let us denote the working model or working distribution for \( X \) by \( f(\cdot; \theta) \), and the true probability distribution for \( X \) by \( g(\cdot) \). Let \( \theta_g \) denote the value of \( \theta \) that maximizes \( E_g[\ln\{f(\cdot; \theta)\}] \), then \( f(\cdot; \theta_g) \) is the density function that is closest to the true density \( g(\cdot) \), in the sense that it minimizes the Kullback-Leibler divergence between \( f(\cdot; \theta) \) and \( g(\cdot) \).

Royall and Tsou (2003) designated \( \theta_g \) as the object of inference when the true distribution is assumed to be \( g(\cdot) \), because \( P\{L(\theta_g)/L(\theta) \to \infty \text{ as } n \to \infty\} = 1 \), for any value \( \theta \neq \theta_g \). That is, the evidence will eventually support \( \theta_g \) over \( \theta \) by an arbitrarily large factor.

Taking the Poisson(\( \theta \)) distribution for example from Royall and Tsou (2003):

\[
f(x; \theta) = \theta^x \exp(-\theta)/x!,
\]

and the parameter of interest or object of interest is the mean, \( E_g(X) \). \( E_g[\ln\{f(X; \theta)\}] = E_g(X)\ln(\theta) - \theta - E_g\{\ln(X!)\}, \) which is maximized at \( \theta_g = E_g(X) \).

In this case, the object of inference is the object of interest.

Royall and Tsou (2003) defined two key conditions that a likelihood function must satisfy in order for it to be considered a robust likelihood function.
(1) the object of inference, $\theta_g$, is the object of interest;

(2) In large samples, the probability of misleading evidence is approximated by the bump function, with maximum value of $\Phi[-\sqrt{2\ln(k)}]$. 

The poisson example we discussed above satisfies condition (1), however, not condition (2), and the maximum probability of misleading evidence is $\Phi[-\sqrt{2\ln(k)}\theta_g/\text{var}_g(X)]$, which can exceed $\Phi[-\sqrt{2\ln(k)}]$. Therefore, when $E(X)$ is the object of interest, the poisson($\theta$) likelihood function is not robust.

Let $a = E_g(-\partial^2\ln\{f(X; \theta_g)/\partial \theta^2\})$ and $b = E_g(\partial[\ln\{f(X; \theta_g)\}]/\partial \theta)^2$, Royall and Tsou (2003) proposed to make a likelihood function robust to model misspecification by raising the working likelihood function to the power of $a/b$. In practice, the unknown quantities $a$ and $b$ can be replaced by their consistent estimates $\hat{a} = \sum(-\partial^2[\ln\{f(X_i; \hat{\theta})\}/\partial \theta^2])/n$ and $\hat{b} = \sum(\partial[\ln\{f(X_i; \hat{\theta})\}]/\partial \theta)^2/n$, which only depend on the working model $f$, not the unknown true model $g$. Royall and Tsou showed that the robustified likelihood function $L_r(\theta) = L(\theta)^{\hat{a}/\hat{b}}$ satisfies conditions (1) and (2). The two properties also hold when a profile likelihood is used instead of a real likelihood, with a different robustifying factor (see Royall and Tsou (2003) for more details).

As future work, we will examine whether the same robustifying factors apply to the GLR under $E^PC$, and we will propose new methods to correct for model misspecification if the same robustifying factors do not hold for the GLR.
**Bibliography**


Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., NHLBI GO Exome Sequencing Project - ESP Lung Project Team, Christiani, D. C.,


Appendix A

A.1 Derivation of the operational characteristics for the Normal probability model

Suppose $X_1, ..., X_n$ are independent and identically distributed random variables with a $N(\theta, \sigma^2)$ probability distribution, with $\sigma^2$ assumed known. Let $\theta_0$ and $\theta_1$ denote the simple null and alternative hypothesis, respectively. With a small and positive constant $0 < C < |\theta_0 - \theta_1|$, we define the composite null hypothesis $\Theta_0 = [\theta_0 - C, \theta_0 + C]$, and the composite alternative hypothesis $\Theta_1 = \Theta_0^c = (-\infty, \theta_0 - C) \cup (\theta_0 + C, \infty)$.

Probability of misleading evidence ($M_0$) when $\theta_0$ is assumed true: 0mu

\[
M_0 = P_0 \left\{ \sup_{\theta \in \Theta_1} \frac{L(\theta^*; x)}{L(\theta_0; x)} \geq k \right\}, \text{ for some } k > 1
\]

\[
= P_0 \left( \frac{L(\bar{X}_n)}{L(\theta_0 + C)} \geq k | \bar{X}_n > \theta_0 + C \right) \cdot P_0(\bar{X}_n > \theta_0 + C) \tag{1}
\]

\[
+ P_0 \left( \frac{L(\theta_0 + C)}{L(X_n)} \geq k | \theta_0 \leq \bar{X}_n \leq \theta_0 + C \right) \cdot P_0(\theta_0 \leq \bar{X}_n \leq \theta_0 + C) \tag{2}
\]

\[
+ P_0 \left( \frac{L(\theta_0 - C)}{L(X_n)} \geq k | \theta_0 - C \leq \bar{X}_n \leq \theta_0 \right) \cdot P_0(\theta_0 - C \leq \bar{X}_n \leq \theta_0) \tag{3}
\]

\[
+ P_0 \left( \frac{L(\bar{X}_n)}{L(\theta_0 - C)} \geq k | \bar{X}_n < \theta_0 - C \right) \cdot P_0(\bar{X}_n < \theta_0 - C) \tag{4}
\]

88
Appendix A.

\begin{align*}
(1) &= P_0 \left( \frac{L(X_n)}{L(\theta_0 + C)} \geq k | X_n > \theta_0 + C \right) \cdot P_0 (X_n > \theta_0 + C) \\
&= P_0 \left\{ \exp \left[ \frac{-\sum (X_i - \bar{X}_n)^2}{2\sigma^2} \right] \geq k | X_n > \theta_0 + C \right\} \cdot P_0 (X_n > \theta_0 + C) \\
&= P_0 \left\{ \exp \left[ \frac{n(\bar{X}_n - \theta_0 - C)^2}{2\sigma^2} \right] \geq k | X_n > \theta_0 + C \right\} \cdot P_0 (X_n > \theta_0 + C) \\
&= P_0 \left[ (\bar{X}_n - \theta_0 - C)^2 \geq 2\sigma^2 / n | X_n > \theta_0 + C \right] \cdot P_0 (X_n > \theta_0 + C) \\
&= [ P_0 (\bar{X}_n - \theta_0 - C \geq \sqrt{2\sigma^2 / n} | X_n > \theta_0 + C) \\
&+ P_0 (\bar{X}_n - \theta_0 - C \leq -\sqrt{2\sigma^2 / n} | X_n > \theta_0 + C) ] \cdot P_0 (X_n > \theta_0 + C) \\
&= P_0 (\bar{X}_n - \theta_0 \geq \sqrt{2\sigma^2 / n} + C | X_n > \theta_0 + C) \cdot P_0 (X_n > \theta_0 + C) \\
&= P_0 (Z \geq \sqrt{\frac{2}{n}} + C \sqrt{n}/\sigma | Z \geq C \sqrt{n}/\sigma) \cdot P_0 (Z \geq C \sqrt{n}/\sigma) \\
&= P_0 (Z \geq \sqrt{\frac{2}{n}} + C \sqrt{n}/\sigma) \\
&= \Phi(-\sqrt{\frac{2}{n}} - C \sqrt{n}/\sigma)
\end{align*}

\begin{align*}
(2) &= P_0 \left( \frac{L(\theta_0 + C)}{L(X_n)} \geq k | \theta_0 \leq \bar{X}_n \leq \theta_0 + C \right) \cdot P_0 (\theta_0 \leq \bar{X}_n \leq \theta_0 + C) \\
&= 0, \text{ since } \frac{L(\theta_0 + C)}{L(X_n)} \leq 1 \text{ when } \theta_0 \leq \bar{X}_n \leq \theta_0 + C
\end{align*}

\begin{align*}
(3) &= P_0 \left( \frac{L(\theta_0 - C)}{L(X_n)} \geq k | \theta_0 - C \leq \bar{X}_n \leq \theta_0 \right) \cdot P_0 (\theta_0 - C \leq \bar{X}_n \leq \theta_0) \\
&= 0, \text{ since } \frac{L(\theta_0 - C)}{L(X_n)} \leq 1 \text{ when } \theta_0 - C \leq \bar{X}_n \leq \theta_0
\end{align*}

\begin{align*}
(4) &= P_0 \left( \frac{L(\bar{X}_n)}{L(\theta_0 - C)} \geq k | (\bar{X}_n < \theta_0 - C) \right) \cdot P_0 (\bar{X}_n < \theta_0 - C) \\
&= P_0 \left\{ \exp \left[ \frac{-\sum (X_i - \bar{X}_n)^2}{2\sigma^2} \right] \geq k | \bar{X}_n < \theta_0 - C \right\} \cdot P_0 (\bar{X}_n < \theta_0 - C) \\
&= P_0 \left\{ \exp \left[ \frac{n(\bar{X}_n - \theta_0 + C)^2}{2\sigma^2} \right] \geq k | \bar{X}_n < \theta_0 - C \right\} \cdot P_0 (\bar{X}_n < \theta_0 - C) \\
&= P_0 [(\bar{X}_n - \theta_0 + C)^2 \geq 2\sigma^2 / n | \bar{X}_n < \theta_0 - C] \cdot P_0 (\bar{X}_n < \theta_0 - C)
\end{align*}
\[ = [P_0(\bar{X}_n - \theta_0 + C \geq \sqrt{2\sigma^2/n} | \bar{X}_n < \theta_0 - C) \]
\[ + P_0(\bar{X}_n - \theta_0 + C \leq -\sqrt{2\sigma^2/n} | \bar{X}_n < \theta_0 - C) \cdot P_0(\bar{X}_n < \theta_0 - C) \]
\[ = P_0(\bar{X}_n - \theta_0 \leq -\sqrt{2\sigma^2/n} - C | \bar{X}_n - \theta_0 < -C) \cdot (\bar{X}_n - \theta_0 < -C) \]
\[ = P_0(Z \leq -\sqrt{2\ln k} - C \sqrt{n}/\sigma | Z < -C \sqrt{n}/\sigma) \cdot P_0(Z < -C \sqrt{n}/\sigma) \]
\[ = P_0(Z \leq -\sqrt{2\ln k} - C \sqrt{n}/\sigma) \]
\[ = \Phi(-\sqrt{2} - C \sqrt{n}/\sigma) \]

Therefore, \( M_0 = (1) + (2) + (3) + (4) = \Phi(-C \sqrt{n}/\sigma - \sqrt{2}) + 0 + 0 + \Phi(-C \sqrt{n}/\sigma - \sqrt{2}) = 2\Phi(-C \sqrt{n}/\sigma - \sqrt{2}). \)

Probability of Strong Evidence (\( S_0 \)) when \( \theta_0 \) is assumed true:

\[ S_0 = P_0\{ \sup_{\theta' \in \Theta_0} \frac{L(\theta'; X)}{L(\theta_0; X)} \geq k \} \]
\[ = P_0\{ \sup_{\theta' \in \Theta_0} \frac{L(\theta'; X)}{L(\theta_0; X)} \leq \frac{1}{k}, \text{ for some } k > 1 \} \]
\[ = P_0\left( \frac{L(\theta_0 + C)}{L(\bar{X}_n)} \geq k | \bar{X}_n \geq \theta_0 + C \right) \cdot P_0(\bar{X}_n \geq \theta_0 + C) \] \( (5) \)
\[ + P_0\left( \frac{L(\bar{X}_n)}{L(\theta_0 + C)} \geq k | \theta_0 < \bar{X}_n < \theta_0 + C \right) \cdot P_0(\theta_0 < \bar{X}_n < \theta_0 + C) \] \( (6) \)
\[ + P_0\left( \frac{L(\bar{X}_n)}{L(\theta_0 - C)} \geq k | \theta_0 - C < \bar{X}_n < \theta_0 \right) \cdot P_0(\theta_0 - C < \bar{X}_n < \theta_0) \] \( (7) \)
\[ + P_0\left( \frac{L(\theta_0 - C)}{L(\bar{X}_n)} \geq k | \bar{X}_n \leq \theta_0 - C \right) \cdot P_0(\bar{X}_n \leq \theta_0 - C) \] \( (8) \)

\( (5) = 0, \text{ since } \frac{L(\theta_0 + C)}{L(\bar{X}_n)} \leq 1 \text{ when } \bar{X}_n \geq \theta_0 + C \)

\( (6) = P_0\left( \frac{L(\bar{X}_n)}{L(\theta_0 + C)} \geq k | \theta_0 < \bar{X}_n < \theta_0 + C \right) \cdot P_0(\theta_0 < \bar{X}_n < \theta_0 + C) \)
\begin{align*}
= & P_0[(\bar{X}_n - \theta_0 - C)^2 \geq 2 \ln k \sigma^2/n | \theta_0 < \bar{X}_n < \theta_0 + C] \cdot P_0(\theta_0 < \bar{X}_n < \theta_0 + C) \\
= & [P_0(\bar{X}_n - \theta_0 \geq C + \sqrt{2 \ln k \sigma^2/n} | 0 < \bar{X}_n - \theta_0 < C) \\
& + P_0(\bar{X}_n - \theta_0 \leq -C - \sqrt{2 \ln k \sigma^2/n} | 0 < \bar{X}_n - \theta_0 < C)] \cdot P_0(0 < \bar{X}_n - \theta_0 < C) \\
= & P_0(\bar{X}_n - \theta_0 \leq C - \sqrt{2 \ln k \sigma^2/n} | 0 < \bar{X}_n - \theta_0 < C) \cdot P_0(0 < \bar{X}_n - \theta_0 < C) \\
= & P_0(Z \leq C \sqrt{n}/\sigma - \sqrt{2 \ln k} | 0 < Z < C \sqrt{n}/\sigma) \cdot P_0(0 < Z < C \sqrt{n}/\sigma) \\
= & \begin{cases} \\
\Phi(C \sqrt{n}/\sigma - \sqrt{2}) - \Phi(0), & \text{if } C \sqrt{n}/\sigma - \sqrt{2 \ln k} > 0 \\
0, & \text{if } C \sqrt{n}/\sigma - \sqrt{2 \ln k} \leq 0 \\
\Phi(C \sqrt{n}/\sigma - \sqrt{2}) - \Phi(0), & \text{if } n > \frac{2 \sigma^2 \ln k}{c^2} \\
0, & \text{if } n \leq \frac{2 \sigma^2 \ln k}{c^2} \\
\end{cases}
\end{align*}

(7) \begin{align*}
= & P_0\left(\frac{L(\bar{X}_n)}{L(\theta_0 - C)} \geq k | \theta_0 - C < \bar{X}_n < \theta_0 \right) \cdot P_0(\theta_0 - C < \bar{X}_n < \theta_0) \\
= & P_0[(\bar{X}_n - \theta_0 + C)^2 \geq 2 \ln k \sigma^2/n | \theta_0 - C < \bar{X}_n < \theta_0] \cdot P_0(\theta_0 - C < \bar{X}_n < \theta_0) \\
= & P_0(\bar{X}_n - \theta_0 \geq -C + \sqrt{2 \ln k \sigma^2/n}) - \Phi(0) \cdot P_0(\theta_0 - C < \bar{X}_n < \theta_0 < 0) \\
= & P_0(Z \geq -C \sqrt{n}/\sigma + \sqrt{2 \ln k} | -C \sqrt{n}/\sigma < Z < 0) \cdot P_0(\theta_0 - C < \bar{X}_n < \theta_0 < 0) \\
= & \begin{cases} \\
\Phi(0) - \Phi(-C \sqrt{n}/\sigma + \sqrt{2 \ln k}), & \text{if } -C \sqrt{n}/\sigma + \sqrt{2 \ln k} < 0 \\
0, & \text{if } -C \sqrt{n}/\sigma + \sqrt{2 \ln k} \geq 0 \\
\Phi(0) - \Phi(-C \sqrt{n}/\sigma + \sqrt{2 \ln k}), & \text{if } n > \frac{2 \sigma^2 \ln k}{c^2} \\
0, & \text{if } n \leq \frac{2 \sigma^2 \ln k}{c^2} \\
\end{cases}
\end{align*}

(8) = 0, since \( \frac{L(\theta_0 - C)}{L(\bar{X}_n)} \leq 1 \) when \( \bar{X}_n < \theta_0 - C \)
Therefore, \( S_0 = \begin{cases} 
2\Phi(C\sqrt{n}/\sigma - \sqrt{2}) - 1, & \text{if } n > \frac{2\sigma^2 \ln k}{C^4} \\
0, & \text{if } n \leq \frac{2\sigma^2 \ln k}{C^4} 
\end{cases} \)

Probability of weak evidence (\( W_0 \)) when \( \theta_0 \) is assumed true:

\[
W_0 = P_0 \left\{ \frac{1}{k} < \frac{\sup_{\theta' \in \Theta} L(\theta''; x)}{\sup_{\theta'' \in \Theta} L(\theta''; x)} < k \right\}, \text{for some } k > 1
\]

\[
= P_0 \left( \frac{1}{k} < \frac{L(\bar{X}_n)}{L(\theta_0 + C)} < k | \bar{X}_n > \theta_0 + C \right) \cdot P_0(\bar{X}_n > \theta_0 + C) (9)
\]

\[
+ P_0 \left( \frac{1}{k} < \frac{L(\theta_0 + C)}{L(\bar{X}_n)} < k | \theta_0 < \bar{X}_n < \theta_0 + C \right) \cdot P_0(\theta_0 < \bar{X}_n < \theta_0 + C) (10)
\]

\[
+ P_0 \left( \frac{1}{k} < \frac{L(\theta_0 - C)}{L(\bar{X}_n)} < k | \theta_0 < \bar{X}_n < \theta_0 \right) \cdot P_0(\theta_0 < \bar{X}_n < \theta_0) (11)
\]

\[
+ P_0 \left( \frac{1}{k} < \frac{L(\bar{X}_n)}{L(\theta_0 - C)} < k | \bar{X}_n < \theta_0 - C \right) \cdot P_0(\bar{X}_n < \theta_0 - C) (12)
\]

\[
(9) = P_0 \left( \frac{1}{k} < \frac{L(\bar{X}_n)}{L(\theta_0 + C)} < k | \bar{X}_n > \theta_0 + C \right) \cdot P_0(\bar{X}_n > \theta_0 + C)
\]

\[
= P_0[(\bar{X}_n - \theta_0 - C)^2 < 2\ln k \sigma^2/n | \bar{X}_n > \theta_0 + C] \cdot P_0(\bar{X}_n > \theta_0 + C)
\]

\[
= P_0(C < \bar{X}_n - \theta_0 < C + \sqrt{2\ln k \sigma^2/n})
\]

\[
= P_0(C \sqrt{n}/\sigma < Z < C \sqrt{n}/\sigma + \sqrt{2\ln k})
\]

\[
= \Phi(C \sqrt{n}/\sigma + \sqrt{2\ln k}) - \Phi(C \sqrt{n}/\sigma)
\]

\[
(10) = P_0 \left( \frac{1}{k} < \frac{L(\theta_0 + C)}{L(\bar{X}_n)} < k | \theta_0 < \bar{X}_n < \theta_0 + C \right) \cdot P_0(\theta_0 < \bar{X}_n < \theta_0 + C)
\]

\[
= P_0 \left( \frac{1}{k} < \frac{L(\bar{X}_n)}{L(\theta_0 + C)} < k | \theta_0 < \bar{X}_n < \theta_0 + C \right) \cdot P_0(\theta_0 < \bar{X}_n < \theta_0 + C)
\]

\[
= P_0[(\bar{X}_n - \theta_0 - C)^2 < 2\ln k \sigma^2/n | \theta_0 < \bar{X}_n < \theta_0 + C] \cdot P_0(\theta_0 < \bar{X}_n < \theta_0 + C)
\]
\( P_0(C - \sqrt{2 \ln k \sigma^2/n} < \bar{X}_n - \theta_0 < C + \sqrt{2 \ln k \sigma^2/n} | 0 \leq \bar{X}_n - \theta_0 \leq C) \cdot P_0(0 \leq \bar{X}_n - \theta_0 \leq C) \\
= P_0(C \sqrt{n}/\sigma - \sqrt{2 \ln k} < Z < C \sqrt{n}/\sigma + \sqrt{2 \ln k} | 0 \leq Z \leq C \sqrt{n}/\sigma) \cdot P_0(0 \leq Z \leq C \sqrt{n}/\sigma) \\
= \begin{cases} \\
\Phi(C \sqrt{n}/\sigma) - \Phi(C \sqrt{n}/\sigma - \sqrt{2 \ln k}), & \text{if } n > \frac{2 \sigma^2 \ln k}{C^2} \\
\Phi(C \sqrt{n}/\sigma) - \Phi(0), & \text{if } n \leq \frac{2 \sigma^2 \ln k}{C^2} \\
\end{cases} \\
\)

(11) \( P_0 \left( \frac{1}{k} \leq \frac{L(\theta_0 - C)}{L(\bar{X}_n)} < k | \theta_0 - C \leq \bar{X}_n < \theta_0 \right) \cdot P_0(\theta_0 - C \leq \bar{X}_n < \theta_0) \\
= P_0 \left( \frac{1}{k} \leq \frac{L(\bar{X}_n)}{L(\theta_0 - C)} < k | \theta_0 - C \leq \bar{X}_n < \theta_0 \right) \cdot P_0(\theta_0 - C \leq \bar{X}_n < \theta_0) \\
= P_0((\bar{X}_n - \theta_0 + C)^2 < 2 \ln k \sigma^2/n | \theta_0 - C \leq \bar{X}_n < \theta_0) \cdot P_0(\theta_0 - C \leq \bar{X}_n < \theta_0) \\
= P_0(-C - \sqrt{2 \ln k \sigma^2/n} < \bar{X}_n - \theta_0 < -C + \sqrt{2 \ln k \sigma^2/n} | -C \leq \bar{X}_n - \theta_0 \leq 0) \cdot P_0(-C \leq \bar{X}_n - \theta_0 \leq 0) \\
= P_0(-C \sqrt{n}/\sigma - \sqrt{2 \ln k} < -C \sqrt{n}/\sigma + \sqrt{2 \ln k} | -C \sqrt{n}/\sigma \leq Z \leq 0) \cdot P_0(-C \sqrt{n}/\sigma \leq Z \leq 0) \\
= \begin{cases} \\
\Phi(-C \sqrt{n}/\sigma + \sqrt{2 \ln k}) - \Phi(-C \sqrt{n}/\sigma), & \text{if } n > \frac{2 \sigma^2 \ln k}{C^2} \\
\Phi(0) - \Phi(-C \sqrt{n}/\sigma), & \text{if } n \leq \frac{2 \sigma^2 \ln k}{C^2} \\
\end{cases} \\
\)

(12) \( P_0 \left( \frac{1}{k} \leq \frac{L(\bar{X}_n)}{L(\theta_0 - C)} < k | X_n < \theta_0 - C \right) \cdot P_0(X_n < \theta_0 - C) \\
= P_0((\bar{X}_n - \theta_0 + C)^2 < 2 \ln k \sigma^2/n | X_n < \theta_0 - C) \cdot P_0(X_n < \theta_0 - C) \\
= P_0(-C - \sqrt{2 \ln k \sigma^2/n} < X_n - \theta_0 < -C + \sqrt{2 \ln k \sigma^2/n} | X_n - \theta_0 < -C) \cdot P_0(X_n - \theta_0 < -C) \\
= P_0(-C \sqrt{n}/\sigma - \sqrt{2 \ln k} < -C \sqrt{n}/\sigma + \sqrt{2 \ln k} | Z < -C \sqrt{n}/\sigma) \cdot P_0(Z < -C \sqrt{n}/\sigma) \\
= P_0(-C \sqrt{n}/\sigma - \sqrt{2 \ln k} < Z < -C \sqrt{n}/\sigma) \\
= \Phi(-C \sqrt{n}/\sigma) - \Phi(-C \sqrt{n}/\sigma - \sqrt{2 \ln k}) \)
If $C\sqrt{n}/\sigma - \sqrt{2 \ln k} > 0$

$$W_0 = [\Phi(C\sqrt{n}/\sigma + \sqrt{2 \ln k}) - \Phi(C\sqrt{n}/\sigma)] + [\Phi(C\sqrt{n}/\sigma) - \Phi(C\sqrt{n}/\sigma - \sqrt{2 \ln k})]$$

$$+ [\Phi(-C\sqrt{n}/\sigma + \sqrt{2 \ln k}) - \Phi(-C\sqrt{n}/\sigma)] + [\Phi(-C\sqrt{n}/\sigma) - \Phi(-C\sqrt{n}/\sigma - \sqrt{2 \ln k})]$$

$$= [\Phi(C\sqrt{n}/\sigma + \sqrt{2 \ln k}) - \Phi(-(C\sqrt{n}/\sigma + \sqrt{2 \ln k}))]$$

$$- [\Phi(C\sqrt{n}/\sigma - \sqrt{2 \ln k}) - \Phi(-(C\sqrt{n}/\sigma - \sqrt{2 \ln k}))]$$

$$= 2(\Phi(C\sqrt{n}/\sigma + \sqrt{2 \ln k}) - \Phi(C\sqrt{n}/\sigma - \sqrt{2 \ln k}))$$

If $C\sqrt{n}/\sigma - \sqrt{2 \ln k} \leq 0$

$$W_0 = [\Phi(C\sqrt{n}/\sigma + \sqrt{2 \ln k}) - \Phi(C\sqrt{n}/\sigma)] + [\Phi(C\sqrt{n}/\sigma) - \Phi(0)]$$

$$+ [\Phi(0) - \Phi(-C\sqrt{n}/\sigma)] + [\Phi(-C\sqrt{n}/\sigma) - \Phi(-C\sqrt{n}/\sigma - \sqrt{2 \ln k})]$$

$$= \Phi(C\sqrt{n}/\sigma + \sqrt{2 \ln k}) - \Phi(-(C\sqrt{n}/\sigma + \sqrt{2 \ln k}))$$

$$= 2(\Phi(C\sqrt{n}/\sigma + \sqrt{2 \ln k}) - \Phi(0))$$

$$= 2(\Phi(C\sqrt{n}/\sigma + \sqrt{2 \ln k}) - 1)$$

Therefore, $W_0 = \begin{cases} 
2[\Phi(C\sqrt{n}/\sigma + \sqrt{2 \ln k}) - \Phi(C\sqrt{n}/\sigma - \sqrt{2 \ln k})], & \text{if } n > \frac{2\sigma^2 \ln k}{C^2} \\
2[\Phi(C\sqrt{n}/\sigma + \sqrt{2 \ln k})] - 1, & \text{if } n \leq \frac{2\sigma^2 \ln k}{C^2}
\end{cases}$

Then as expected, when $n > \frac{2\sigma^2 \ln k}{C^2}$,

$$M_0 + W_0 + S_0 = 2\Phi(-C\sqrt{n}/\sigma - \sqrt{2}) + 2[\Phi(C\sqrt{n}/\sigma + \sqrt{2}) - 1] + 2[\Phi(C\sqrt{n}/\sigma + \sqrt{2 \ln k}) - \Phi(C\sqrt{n}/\sigma - \sqrt{2 \ln k})] = 1,$$

and when $n \leq \frac{2\sigma^2 \ln k}{C^2}$,

$$M_0 + W_0 + S_0 = 2\Phi(-C\sqrt{n}/\sigma - \sqrt{2}) + 0 + 2[\Phi(C\sqrt{n}/\sigma + \sqrt{2 \ln k}) - 1] = 1.$$
When $\theta_1$ is assumed true, we derive the corresponding operational characteristics, $M_1$, $W_1$ and $S_1$ below.

Probability of observing misleading evidence ($M_1$) when $\theta_1$ is assumed true:

\[
M_1 = P_1\left( \sup_{\theta' \in \Theta_0} L(\theta'; x) \geq k, \sup_{\theta'' \in \Theta_1} L(\theta''; x) \geq k \right), \text{ for some } k > 1
\]

\[
= P_1\left( \frac{L(\theta_0 + C)}{L(\bar{X}_n)} \geq k|\bar{X}_n \geq \theta_0 + C \right) \cdot P_1(\bar{X}_n \geq \theta_0 + C) (1)
\]

\[
+ P_1\left( \frac{L(\bar{X}_n)}{L(\theta_0 + C)} \geq k|\theta_0 < \bar{X}_n < \theta_0 + C \right) \cdot P_1(\theta_0 < \bar{X}_n < \theta_0 + C) (2)
\]

\[
+ P_1\left( \frac{L(\bar{X}_n)}{L(\theta_0 - C)} \geq k|\theta_0 - C < \bar{X}_n < \theta_0 \right) \cdot P_1(\theta_0 - C < \bar{X}_n < \theta_0) (3)
\]

\[
+ P_1\left( \frac{L(\theta_0 - C)}{L(\bar{X}_n)} \geq k|\bar{X}_n \leq \theta_0 - C \right) \cdot P_1(\bar{X}_n \leq \theta_0 - C) (4)
\]

(1) $= 0$, since $\frac{L(\theta_0 + C)}{L(\bar{X}_n)} \leq 1$ when $\bar{X}_n \geq \theta_0 + C$

(2) $= P_1\left( \frac{L(\bar{X}_n)}{L(\theta_0 + C)} \geq k|\theta_0 < \bar{X}_n < \theta_0 + C \right) \cdot P_1(\theta_0 < \bar{X}_n < \theta_0 + C)$

\[
= P_1(\bar{X}_n < \theta_0 + C - \sqrt{2\sigma^2 \ln k/\bar{n}}|\theta_0 < \bar{X}_n < \theta_0 + C) \cdot P_1(\theta_0 < \bar{X}_n < \theta_0 + C)
\]

\[
= P_1(\bar{X}_n - \theta_1 \leq \theta_0 + C - \theta_1 - \sqrt{2\sigma^2 \ln k/\bar{n}}|\theta_0 - \theta_1 < \bar{X}_n - \theta_1 < \theta_0 + C - \theta_1)
\]

\[
\times P_1(\theta_0 - \theta_1 < \bar{X}_n - \theta_1 - < \theta_0 + C - \theta_1)
\]

\[
= P_1\left( Z \leq (\theta_0 + C - \theta_1) \sqrt{n}/\sigma - \sqrt{2\ln k}(\theta_0 - \theta_1) \sqrt{n}/\sigma < Z < (\theta_0 + C - \theta_1) \sqrt{n}/\sigma \right)
\]

\[
\times P_1(\theta_0 - \theta_1 < \sqrt{n}/\sigma < Z < (\theta_0 + C - \theta_1) \sqrt{n}/\sigma)
\]

\[
= \begin{cases} 
\Phi\left((\theta_0 + C - \theta_1) \sqrt{n}/\sigma - \sqrt{2\ln k}\right) - \Phi\left((\theta_0 - \theta_1) \sqrt{n}/\sigma\right), & \text{if } n > \frac{2\sigma^2 \ln k}{C^2} \\
0, & \text{if } n \leq \frac{2\sigma^2 \ln k}{C^2}
\end{cases}
\]

(3) $= P_1\left( \frac{L(\bar{X}_n)}{L(\theta_0 - C)} \geq k|\theta_0 - C < \bar{X}_n < \theta_0 \right) \cdot P_1(\theta_0 - C < \bar{X}_n < \theta_0)$
\[ P_1(\bar{X}_n - \theta_1 \geq \theta_0 - C - \theta_1 + \sqrt{2\sigma^2 \ln k/n}|\theta_0 - C - \theta_1 < \bar{X}_n - \theta_1 < \theta_0 - \theta_1) \]
\[ \times P_1(\theta_0 - C - \theta_1 < \bar{X}_n - \theta_1 < \theta_0 - \theta_1) \]
\[ = P_1[Z \geq (\theta_0 - C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}|(\theta_0 - C - \theta_1)\sqrt{n}/\sigma < Z < (\theta_0 - \theta_1)\sqrt{n}/\sigma] \]
\[ \times P_1[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma < Z < (\theta_0 - \theta_1)\sqrt{n}/\sigma] \]
\[ = \begin{cases} 
\Phi[(\theta_0 - \theta_1)\sqrt{n}/\sigma] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}], & \text{if } n > \frac{2\sigma^2 \ln k}{C^2} \\
0, & \text{if } n \leq \frac{2\sigma^2 \ln k}{C^2}
\end{cases} \]

(4) = 0, since \( \frac{L(\theta_0 - C)}{L(\bar{X}_n)} \leq 1 \) when \( \bar{X}_n \leq \theta_0 - C \)

Therefore, \( M_1 = \begin{cases} 
\Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}], & \text{if } n > \frac{2\sigma^2 \ln k}{C^2} \\
-\Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}], & \text{if } n \leq \frac{2\sigma^2 \ln k}{C^2}
\end{cases} \)

Upper bound for \( M_1 \):

Since \( M_1 = 0 \) when \( n \leq \frac{2\sigma^2 \ln k}{C^2} \), we will focus on the case when \( n > \frac{2\sigma^2 \ln k}{C^2} \).

(1) When \( \theta_1 > \theta_0 + C \), or equivalently, \( \theta_0 + C - \theta_1 < 0 \),

\[ M_1 = \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] - \Phi[(\theta_0 - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] \]
\[ \leq \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] \]
\[ < \Phi(-\sqrt{2\ln k}) \]
(2) When \( \theta_1 < \theta_0 - C \), or equivalently, \(-(\theta_0 - C - \theta_1) < 0\),

\[
M_1 = \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] - \{1 - \Phi[\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}]\)
\[
= \Phi[\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] - \{1 - \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}]\}
\]

\[
\leq \Phi[\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}]
\]

\[
< \Phi(-\sqrt{2\ln k})
\]

Probability of observing strong evidence \((S_1)\) when \( \theta_1 \) is assumed true:

\[
S_1 = P_1(\sup_{\theta' \in \Theta_1} \frac{L(\theta''; x)}{L(\theta''; x)} \geq k | X_n > \theta_0 + C) \cdot P_1(X_n > \theta_0 + C)
\]

\[
= P_1(\frac{L(\bar{X}_n)}{L(\theta_0 + C)} \geq k | X_n > \theta_0 + C) \cdot P_1(X_n > \theta_0 + C)
\]

(5) \[
= P_1(\frac{L(\bar{X}_n)}{L(\theta_0 + C)} \geq k | X_n > \theta_0 + C) \cdot P_1(\bar{X}_n > \theta_0 + C)
\]

(5) \[
= P_1(X_n > \theta_0 + C + \sqrt{2\sigma^2 \ln k/n} | X_n > \theta_0 + C) \cdot P_1(X_n > \theta_0 + C)
\]

(5) \[
= P_1[Z > (\theta_0 + C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}]
\]

(5) \[
= \Phi[\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}]
\]

(6) = 0, since \( \frac{L(\theta_0 + C)}{L(X_n)} \leq 1 \) when \( \theta_0 \leq X_n \leq \theta_0 + C \)

(7) = 0, since \( \frac{L(\theta_0 - C)}{L(X_n)} \leq 1 \) when \( \theta_0 - C \leq X_n \leq \theta_0 \)
(8) \[ P_1\left( \frac{L(\bar{X}_n)}{L(\theta_0 - C)} \geq k \left\vert \bar{X}_n < \theta_0 - C \right. \right) \cdot P_1(\bar{X}_n < \theta_0 - C) \]
\[ = P_1(\bar{X}_n < \theta_0 - C - \sqrt{2\sigma^2 \ln k/n}) \]
\[ = P_1(Z < (\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}) \]
\[ = \Phi[((\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] \]

Therefore, \[ S_1 = \Phi[\neg((\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] + \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}] \]

Probability of observing weak evidence (W1) when \( \theta_1 \) is assumed true:

\[ W_1 = P_1\left\{ \frac{1}{k} \leq \frac{\sup_{\theta'' \in \Theta_1} L(\theta''; x)}{\sup_{\theta' \in \Theta_0} L(\theta'; x)} < k \right\} \]
\[ = P_1(\frac{1}{k} < \frac{L(\bar{X}_n)}{L(\theta_0 + C)} < k \left\vert \bar{X}_n > \theta_0 + C \right. \right) \cdot P_1(\bar{X}_n > \theta_0 + C) \] (9)
\[ = P_1(\frac{1}{k} < \frac{L(\theta_0 + C)}{L(\bar{X}_n)} < k \left\vert \theta_0 \leq \bar{X}_n \leq \theta_0 + C \right. \right) \cdot P_1(\theta_0 \leq \bar{X}_n \leq \theta_0 + C) \] (10)
\[ = P_1(\frac{1}{k} < \frac{L(\theta_0 - C)}{L(\bar{X}_n)} < k \left\vert \theta_0 - C \leq \bar{X}_n \leq \theta_0 \right. \right) \cdot P_1(\theta_0 - C \leq \bar{X}_n \leq \theta_0) \] (11)
\[ = P_1(\frac{1}{k} < \frac{L(\bar{X}_n)}{L(\theta_0 - C)} < k \left\vert \bar{X}_n < \theta_0 - C \right. \right) \cdot P_1(\bar{X}_n < \theta_0 - C) \] (12)

(9) \[ = P_1(\frac{1}{k} < \frac{L(\bar{X}_n)}{L(\theta_0 + C)} < k \left\vert \bar{X}_n > \theta_0 + C \right. \right) \cdot P_1(\bar{X}_n > \theta_0 + C) \]
\[ = P_1((\theta_0 + C - \theta_1)\sqrt{n}/\sigma < Z < (\theta_0 + C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}) \]
\[ = \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] - \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma] \]

(10) \[ = P_1(\frac{1}{k} < \frac{L(\theta_0 + C)}{L(\bar{X}_n)} < k \left\vert \theta_0 \leq \bar{X}_n \leq \theta_0 + C \right. \right) \cdot P_1(\theta_0 \leq \bar{X}_n \leq \theta_0 + C) \]
\[ = P_1(\theta_0 + C - \sqrt{2\sigma^2 \ln k/n} \leq \bar{X}_n \leq \theta_0 + C + \sqrt{2\sigma^2 \ln k/n}) \cdot P_1(\theta_0 \leq \bar{X}_n \leq \theta_0 + C) \]
\[ \times P_1(\theta_0 \leq \bar{X}_n \leq \theta_0 + C) \]
= \begin{cases} 
\Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma] - \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}], & \text{if } n > \frac{2\sigma^2\ln k}{C^2} \\
\Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma] - \Phi[(\theta_0 - \theta_1)\sqrt{n}/\sigma], & \text{if } n \leq \frac{2\sigma^2\ln k}{C^2} 
\end{cases}

(11) = P_1 \left( \frac{1}{k} < \frac{L(\theta_0 - C)}{L(X_n)} < k | \theta_0 - C \leq \bar{X}_n \leq \theta_0 \right) \cdot P_1(\theta_0 - C \leq \bar{X}_n \leq \theta_0)

= P_1(\theta_0 - C - \sqrt{2\sigma^2 \ln k}/n < \bar{X}_n < \theta_0 - C + \sqrt{2\sigma^2 \ln k}/n | \theta_0 - C \leq \bar{X}_n \leq \theta_0)

\times P_1(\theta_0 - C \leq \bar{X}_n \leq \theta_0)

= \begin{cases} 
\Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma], & \text{if } n > \frac{2\sigma^2\ln k}{C^2} \\
\Phi[(\theta_0 - \theta_1)\sqrt{n}/\sigma] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma], & \text{if } n \leq \frac{2\sigma^2\ln k}{C^2} 
\end{cases}

(12) = P_1 \left( \frac{1}{k} < \frac{L(\bar{X}_n)}{L(\theta_0 - C)} < k | \bar{X}_n \leq \theta_0 - C \right) \cdot P_1(\bar{X}_n < \theta_0 - C)

= P_1[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k} < \bar{X}_n < \theta_0 - C - \theta_1)\sqrt{n}/\sigma]

= \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}]

W_1 = \begin{cases} 
\Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] - \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}], & \text{if } n > \frac{2\sigma^2\ln k}{C^2} \\
\Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}], & \text{if } n \leq \frac{2\sigma^2\ln k}{C^2} 
\end{cases}

= \begin{cases} 
\Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}], & \text{if } n > \frac{2\sigma^2\ln k}{C^2} \\
\Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma + \sqrt{2\ln k}] - \Phi[(\theta_0 + C - \theta_1)\sqrt{n}/\sigma - \sqrt{2\ln k}], & \text{if } n \leq \frac{2\sigma^2\ln k}{C^2} 
\end{cases}
A.2 Derivation of the operational characteristics for one-parameter parametric models

For the following derivation, we will assume that $X_1, X_2, ..., X_n$ are i.i.d random variables with common density or frequency function $f(x; \theta)$ where $\theta$ is a real-valued parameter.

Define $\ell(x; \theta) = \ln f(x; \theta)$ and let $\ell'(x; \theta), \ell''(x; \theta), \text{ and } \ell'''(x; \theta)$ be the first three partial derivatives of $\ell(x; \theta)$ with respect to $\theta$. We will assume the following regularity conditions about $f(x; \theta)$:

(R1) The parameter space $\Theta$ is an open subset of the real line.

(R2) The set $A = \{x : f(x; \theta) > 0\}$ does not depend on $\theta$.

(R3) $f(x; \theta)$ is three times continuously differentiable with respect to $\theta$ for all $x$ in $A$.

(R4) $E_{\theta}[\ell'(X_i; \theta)] = 0$ for all $\theta$ and $Var_{\theta}[\ell'(X_i; \theta)] = I(\theta)$ where $0 < I(\theta) < \infty$ for all $\theta$.

(R5) $E_{\theta}[\ell''(X_i; \theta)] = -J(\theta)$ where $0 < J(\theta) < \infty$ for all $\theta$.

(R6) For each $\theta$ and $\delta > 0$, $|\ell'''(x; t)| \leq M(x)$ for $|\theta - t| \leq \delta$ where $E_{\theta}[M(X_i)] < \infty$.

In addition, we assume that we can differentiate twice inside the integral $\int_A f(x; \theta)dx$, which leads to $I(\theta) = J(\theta)$.

Probability of misleading evidence ($M_0$) when $\theta_0$ is assumed true:

$$M_0 = P_0 \left[ \sup_{\theta' \in \Theta} \frac{L(\theta'; x)}{L(\theta_0 + C)} \geq k \right]$$

$$= P_0 \left[ \frac{L(\hat{\theta})}{L(\theta_0 + C)} \geq k \frac{\theta \geq \theta_0 + C}{\theta \geq \theta_0} \cdot P_0(\theta \geq \theta_0 + C) \right]$$

$$+ P_0 \left[ \frac{L(\theta_0 + C)}{L(\hat{\theta})} \geq k \frac{\theta \geq \theta_0 + C}{\theta \geq \theta_0} \cdot P_0(0 \leq \theta \leq \theta_0 + C) \right]$$

$$+ P_0 \left[ \frac{L(\theta_0 - C)}{L(\hat{\theta})} \geq k \frac{\theta \leq \theta_0 - C}{\theta \leq \theta_0} \cdot P_0(\theta_0 - C \leq \theta \leq \theta_0) \right]$$

$$+ P_0 \left[ \frac{L(\hat{\theta})}{L(\theta_0 - C)} \geq k \frac{\theta \leq \theta_0 - C}{\theta \leq \theta_0} \cdot P_0(\theta \leq \theta_0 - C) \right]$$
Appendix A.

It has been previously shown that
\[ 2 \ln \frac{L(\hat{\theta})}{L(\theta_0)} \xrightarrow{n \to \infty} \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{1}{n \mathcal{I}(\theta_0)}}} \sim \chi^2_{(1)}, \]
where by \( \frac{D}{n \to \infty} \), we denote convergence in distribution, \( \chi^2_{(1)} \), Chi-square distribution with 1 degree of freedom, and \( Z = N(0, 1) \), standard normal distribution with mean 0 and standard deviation 1.

Since \( \hat{\theta} \geq \theta_0 + C \) implies
\[ \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{1}{n \mathcal{I}(\theta_0)}}} \geq C \sqrt{\frac{n}{\mathcal{I}(\theta_0)}} \]
\[ 2 \ln \frac{L(\hat{\theta})}{L(\theta_0)} \geq 0 \]
\[ \frac{D}{n \to \infty} \rightarrow Z^2_* \]
where \( Z_* \) is a truncated normal (TN) distribution with mean \( C \sqrt{n / \mathcal{I}(\theta_0)} \), or simply \( TN(0, 1, C \sqrt{n / \mathcal{I}(\theta_0)}, \infty) \).

Now we will show that
\[ 2 \ln \frac{L(\hat{\theta})}{L(\theta_0)} \xrightarrow{n \to \infty} \hat{\theta} \geq \theta_0 + C \xrightarrow{n \to \infty} Z^2_* \]

\[ 2 \ln \frac{L(\hat{\theta})}{L(\theta_0 + C)} = 2 \ell(\theta_0) - 2 \ell(\theta_0 + C) \]
\[ = 2 \ell(\theta_0) - 2 \left[ \ell(\theta_0) + C \ell'(\theta_0) + \frac{1}{2} C^2 \ell''(\theta_0) + \frac{1}{6} C^3 \ell'''(\theta_0)^* \right] \]
\[ = -2C \ell'(\theta_0) - C^2 \ell''(\theta_0) - \frac{1}{3} C^3 \ell'''(\theta_0)^* \]
\[ = -2C \sqrt{n \mathcal{I}(\theta_0)} \ell'(\theta_0) - C^2 \ell''(\theta_0) - \frac{1}{3} C^3 \ell'''(\theta_0)^* \]
where \( \theta_0^* \) lies between \( \theta_0 \) and \( \theta_0 + C \).

Taylor series expand \( \ell'(\hat{\theta}) \) at \( \theta_0 \), we get
\[ \frac{\ell'(\hat{\theta})}{\sqrt{nI(\theta_0)}} = \frac{\ell'(\theta_0)}{\sqrt{nI(\theta_0)}} + \frac{\ell''(\theta_0)(\hat{\theta} - \theta_0)}{\sqrt{nI(\theta_0)}} + \frac{\ell'''(\theta_0)(\hat{\theta} - \theta_0)^2}{2\sqrt{nI(\theta_0)}}, \]

where \( \theta_n^{**} \) lies between \( \theta \) and \( \hat{\theta} \).

Since \( \ell'(\hat{\theta}) = 0 \) by the definition of MLE,
\[ \frac{\ell'(\theta_0)}{\sqrt{nI(\theta_0)}} = -\frac{\ell''(\theta_0)(\hat{\theta} - \theta_0)}{\sqrt{nI(\theta_0)}} - \frac{\ell'''(\theta_0)(\hat{\theta} - \theta_0)^2}{2\sqrt{nI(\theta_0)}}, \]

\[ 2 \ln \frac{L(\theta_0)}{L(\theta_0 + C)} = -2C\sqrt{nI(\theta_0)}[\frac{\ell''(\theta_0)(\hat{\theta} - \theta_0)}{\sqrt{nI(\theta_0)}} - \frac{\ell'''(\theta_0)(\hat{\theta} - \theta_0)^2}{2\sqrt{nI(\theta_0)}}] - C^2\ell''(\theta_0) - \frac{1}{3}C^3\ell'''(\theta_0) \]

As \( n \to \infty \), \( \ell''(\theta_0) \to -nI(\theta_0) \), \( \frac{\hat{\theta} - \theta_0}{\sqrt{nI(\theta_0)}} \to D \), and \( \frac{\ell'''(\theta_0)(\hat{\theta} - \theta_0)^2}{2\sqrt{nI(\theta_0)}} \to 0 \) by A6.

Let's define \( \Delta = \frac{C}{\sqrt{nI(\theta_0)}} = C\sqrt{nI(\theta_0)} \), then \( \Delta \) is the margin or error measured in the units of the standard error, and \( C^3\ell'''(\theta_0) = \frac{\Delta^2\ell'''(\theta_0)}{n} \cdot \frac{1}{I(\theta_0)} \)

Since \( \Delta = O(p(\sqrt{n})) \), \( \Delta^2 = O(p(n)) \), and \( \ell'''(\theta_0) \) is bounded by A6, \( C^3\ell'''(\theta_0) \to 0 \) as \( n \to \infty \).

Therefore, \( 2 \ln \frac{L(\theta_0)}{L(\theta_0 + C)} = -2C\sqrt{nI(\theta_0)}Z + C^2nI(\theta_0) \)

Since \( \hat{\theta} \geq \theta_0 \) and \( C \) implies \( \frac{\hat{\theta} - \theta_0}{\sqrt{nI(\theta_0)}} \leq C\sqrt{nI(\theta_0)} \), \( 2 \ln \frac{L(\theta_0)}{L(\theta_0 + C)} \hat{\theta} \geq \theta_0 + C = -2C\sqrt{nI(\theta_0)}Z + C^2nI(\theta_0) \), where \( Z \sim TN(0, 1, C\sqrt{nI(\theta_0)}, \infty) \).

1(a) \( = P_0[2 \ln \frac{L(\hat{\theta})}{L(\theta_0)} + \frac{L(\theta_0)}{L(\theta_0 + C)} - 2 \ln k \geq 0 | \hat{\theta} \geq \theta_0 + C] \)

\( = P_0(Z^2 - 2C\sqrt{nI(\theta_0)}Z + C^2nI(\theta_0) - 2 \ln k \geq 0) \)

\( = P_0(Z \geq C\sqrt{nI(\theta_0)} + \sqrt{2 \ln k}) + P_0(Z \leq C\sqrt{nI(\theta_0)} - \sqrt{2 \ln k}) \)

\( = P_0(Z \geq C\sqrt{nI(\theta_0)} + \sqrt{2 \ln k}) + 0 \)

\( = \frac{P_0(Z \geq C\sqrt{nI(\theta_0)})}{P_0(Z \geq C\sqrt{nI(\theta_0)})} \)
1(b) = \( P_0(\hat{\theta} \geq \theta_0 + C) \)
\[ = P_0\left( \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{1}{nL(\theta_0)}}} \geq C \sqrt{nI(\theta_0)} \right) \]
\[ = P_0(\theta \geq C \sqrt{nI(\theta_0)}) \]

(1) = 1(a) \cdot 1(b)
\[ = \frac{P_0(Z \geq C \sqrt{nI(\theta_0)} + \sqrt{2 \ln k})}{P_0(Z \geq C \sqrt{nI(\theta_0)})} \cdot P_0(Z \geq C \sqrt{nI(\theta_0)}) \]
\[ = P_0(Z \geq C \sqrt{nI(\theta_0)} + \sqrt{2 \ln k}) \]
\[ = \Phi(-C \sqrt{nI(\theta_0)} - \sqrt{2 \ln k}) \]

(2) = 0 by the definition of MLE

(3) = 0 by the definition of MLE

4(a) = \( P_0\left[ \frac{L(\hat{\theta})}{L(\theta - C)} \geq k | \hat{\theta} \leq \theta_0 - C \right] \)
\[ = P_0[2 \ln \frac{L(\hat{\theta})}{L(\theta_0)} + 2 \ln \frac{L(\theta_0)}{L(\theta_0 - C)} - 2 \ln k \geq 0 | \hat{\theta} \leq \theta_0 - C] \]

Since \( \hat{\theta} \leq \theta_0 - C \implies \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{1}{nI(\theta_0)}}} \leq -C \sqrt{nI(\theta_0)} \),
\[ 2 \ln \frac{L(\hat{\theta})}{L(\theta_0)} \leq 0 \]
\[ = 0 \]
where \( Z_{**} = \mathcal{N}(0, 1, -\infty, -C \sqrt{nI(\theta_0)}) \)

\[ 2 \ln \frac{L(\theta_0)}{L(\theta_0 - C)} \]
\[ = 2\ell(\theta_0) - 2\ell(\theta_0 - C) \]
\[ = 2\ell(\theta_0) - 2[\ell(\theta_0) - C \ell'(\theta_0) + \frac{1}{2} C^2 \ell''(\theta_0) - \frac{1}{6} C^3 \ell'''(\theta_n^*)] \]
\[ = 2C\ell' (\theta_0) - C^2\ell'' (\theta_0) + \frac{1}{3} \ell''' (\theta_0^*) \]
\[ = 2C \sqrt{n\mathcal{I} (\theta_0)} \frac{\ell' (\theta_0)}{\sqrt{n\mathcal{I} (\theta_0)}} - C^2\ell'' (\theta_0) + \frac{1}{3} \ell''' (\theta_0^*) \]
, where \( \theta_n^* \) lies between \( \theta_0 - C \) and \( \theta_0 \).

Since \( \hat{\theta} \leq \theta_0 - C \) implies \( \frac{\hat{\theta} - \theta_0}{\sqrt{n\mathcal{I} (\theta_0)}} \leq -C \sqrt{n\mathcal{I} (\theta_0)} \),
\[ 2 \ln \frac{L(\theta_0)}{L(\theta_0 - C)} | \hat{\theta} \leq \theta_0 - C \xrightarrow{D} \xrightarrow{n \to \infty} TN(0, 1, -\infty, -C \sqrt{n\mathcal{I} (\theta_0)}) \]

\[ 4(a) = P_0 \left[ \frac{L(\hat{\theta})}{L(\theta - C)} \geq k | \hat{\theta} \leq \theta_0 - C \right] = P_0 [Z_{**}^2 + 2C \sqrt{\mathcal{I} (\theta_0 + 0)} Z_{**}^2 + C^2 \mathcal{I} (\theta_0) - 2 \ln k \geq 0] \]
\[ = P_0 [Z_{**} \geq -C \sqrt{\mathcal{I} (\theta_0)} + \sqrt{2 \ln k}] + P_0 [Z_{**} \leq -C \sqrt{\mathcal{I} (\theta_0)} - \sqrt{2 \ln k}] \]
\[ = 0 + P_0 (Z_{**} \leq -C \sqrt{\mathcal{I} (\theta_0)} - \sqrt{2 \ln k}) \]
\[ = \frac{P_0 (Z \leq -C \sqrt{\mathcal{I} (\theta_0)} - \sqrt{2 \ln k})}{P_0 (Z \leq -C \sqrt{\mathcal{I} (\theta_0)})} \]

\[ 4(b) = P_0 (\hat{\theta} \leq \theta_0 - C) = P_0 \left( \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{1}{n\mathcal{I} (\theta_0)}}} \leq -C \sqrt{\mathcal{I} (\theta_0)} \right) = P_0 (Z \leq -C \sqrt{\mathcal{I} (\theta_0)}) \]

\[ 4 = 4(a) \cdot 4(b) \]
\[ = \frac{P_0 (Z \leq -C \sqrt{\mathcal{I} (\theta_0)} - \sqrt{2 \ln k})}{P_0 (Z \leq -C \sqrt{\mathcal{I} (\theta_0)})} \cdot P_0 (Z \leq -C \sqrt{\mathcal{I} (\theta_0)}) \]
\[ = P_0 (Z \leq -C \sqrt{\mathcal{I} (\theta_0)} - \sqrt{2 \ln k}) \]
\[ = \Phi[-C \sqrt{\mathcal{I} (\theta_0)} - \sqrt{2 \ln k}] \]

\[ M_0 = (1) + (2) + (3) + (4) = \Phi(-C \sqrt{\mathcal{I} (\theta_0)} - \sqrt{2 \ln k}) + 0 + 0 + \Phi(-C \sqrt{\mathcal{I} (\theta_0)} - \sqrt{2 \ln k}) = 2\Phi(-C \sqrt{\mathcal{I} (\theta_0)} - \sqrt{2 \ln k}) \]

Probability of observing strong \( (S_0) \) evidence when \( \theta_0 \) is assumed true:
For any $\Theta_0 \subset \Theta, \Theta_1 = R \setminus \Theta_0$ and $k > 1$, the probability of observing strong evidence in $X = x$ that supports $\theta \in \Theta_0$ over $\theta \in \Theta_1$ at level $k$ with respect to some $\theta_0 \in \Theta_0$ is $S_0 = P_0\left[ \sup_{\theta'^{\prime} \in \Theta_0} L(\theta'^{\prime}; x) \geq k \right]$

$$S_0 = P_0\left[ \sup_{\theta'^{\prime} \in \Theta_0} \frac{L(\theta'^{\prime}; x)}{L(\theta_s'; x)} \geq k \right]$$

$$= P_0\left[ \frac{L(\hat{\theta}_0 + C)}{L(\theta)} \geq k | \hat{\theta} \geq \theta_0 + C \right] \cdot P_0(\hat{\theta} \geq \theta_0 + C) \tag{5}$$

$$+ P_0\left[ \frac{L(\hat{\theta})}{L(\theta_0 + C)} \geq k | \theta_0 < \hat{\theta} \leq \theta_0 + C \right] \cdot P_0(\theta_0 < \hat{\theta} \leq \theta_0 + C) \tag{6(a)}$$

$$+ P_0\left[ \frac{L(\hat{\theta})}{L(\theta_0 - C)} \geq k | \theta_0 - C \leq \hat{\theta} \leq \theta_0 \right] \cdot \frac{P_0(\theta_0 - C \leq \hat{\theta} \leq \theta_0)}{P_0(\theta_0 - C \leq \hat{\theta} \leq \theta_0)} \tag{7(a)}$$

$$+ P_0\left[ \frac{L(\theta_0 - C)}{L(\theta)} \geq k | \hat{\theta} \leq \theta_0 - C \right] \cdot P_0(\hat{\theta} \leq \theta_0 - C) \tag{8}$$

(5) = 0 by the definition of MLE.

$$6(a) = P_0\left[ \frac{L(\hat{\theta})}{L(\theta_0 + C)} \geq k | \theta_0 \leq \hat{\theta} \leq \theta_0 + C \right]$$

$$= P_0\left[ 2 \ln \left( \frac{L(\hat{\theta})}{L(\theta_0)} \right) + 2 \ln \frac{L(\theta_0)}{L(\theta_0 + C)} - 2 \ln k \geq 0 | \theta_0 \leq \hat{\theta} \leq \theta_0 + C \right]$$

Since $\theta_0 \leq \hat{\theta} \leq \theta_0 + C$ implies $0 \leq \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{1}{n^2(\theta_0)}}} \leq C \sqrt{n \mathcal{I}(\theta_0)}$,

$$2 \ln \frac{L(\hat{\theta})}{L(\theta_0)} | \theta_0 \leq \hat{\theta} \leq \theta_0 + C \xrightarrow{n \to \infty} Z_2^*, \text{ and } 2 \ln \frac{L(\theta_0)}{L(\theta_0 + C)} | \theta_0 \leq \hat{\theta} \leq \theta_0 + C \xrightarrow{n \to \infty} -2C \sqrt{n \mathcal{I}(\theta_0)} Z_2^* + C^2 n \mathcal{I}(\theta_0),$$

where $Z_2^* = TN(0, 1, 0, C \sqrt{n \mathcal{I}(\theta_0)})$.

$$P_0\left[ 2 \ln \left( \frac{L(\hat{\theta})}{L(\theta_0)} \right) + 2 \ln \frac{L(\theta_0)}{L(\theta_0 + C)} - 2 \ln k \geq 0 | \theta_0 \leq \hat{\theta} \leq \theta_0 + C \right]$$
\[ P_0 \left[ Z_*^2 - 2C \sqrt{nI(\theta_0)} Z_* + C^2 nI(\theta_0) - 2 \ln \lambda \geq 0 \right] \]
\[ = P_0(\theta_0 \leq \hat{\theta} \leq \theta_0 + C) = P_0(0 \leq Z \leq C \sqrt{nI(\theta_0)}) \]
\[ = \begin{cases} P(0 \leq Z \leq C \sqrt{nI(\theta_0)} - 2 \ln k), & \text{if } C \sqrt{nI(\theta_0)} - 2 \ln k > 0 \\ 0, & \text{if } C \sqrt{nI(\theta_0)} - 2 \ln k \leq 0 \end{cases} \]

\[ 6(b) = P_0(\theta_0 \leq \hat{\theta} \leq \theta_0 + C) = P_0(0 \leq Z \leq C \sqrt{nI(\theta_0)}) \]

\[ 7(a) = P_0 \left[ \frac{L(\hat{\theta})}{L(\theta_0 - C)} \geq k | \theta_0 - C \leq \hat{\theta} \leq \theta_0 \right] \]
\[ = P_0 \left[ 2 \ln \frac{L(\hat{\theta})}{L(\theta_0)} + 2 \ln \frac{L(\theta_0)}{L(\theta_0 - C)} - 2 \ln k \geq 0 | \theta_0 - C \leq \hat{\theta} \leq \theta_0 \right] \]

Since \( \theta_0 - C \leq \hat{\theta} \leq \theta_0 \) implies \( -C \sqrt{nI(\theta_0)} \leq \frac{\hat{\theta} - \theta_0}{\sqrt{nI(\theta_0)}} \leq 0, 2 \ln \frac{L(\hat{\theta})}{L(\theta_0)} | \theta_0 - C \leq \hat{\theta} \leq \theta_0 \overset{D}{\rightarrow} 2C \sqrt{nI(\theta_0)} Z_* + C^2 nI(\theta_0), \)

\[ P_0 \left[ 2 \ln \frac{L(\hat{\theta})}{L(\theta_0)} + 2 \ln \frac{L(\theta_0)}{L(\theta_0 - C)} - 2 \ln k \geq 0 | \theta_0 - C \leq \hat{\theta} \leq \theta_0 \right] \]
\[ = P_0 \left[ Z_*^2 + 2C \sqrt{nI(\theta_0)} Z_* + C^2 nI(\theta_0) - 2 \ln \lambda \geq 0 \right] \]
\[ = P_0(\theta_0 \leq \hat{\theta} \leq \theta_0 + C) = P_0(0 \leq Z \leq C \sqrt{nI(\theta_0)}) \]
\[ = \begin{cases} \frac{P(0 \leq Z \leq C \sqrt{nI(\theta_0)} - 2 \ln k \leq 0)}{P(0 \leq Z \leq C \sqrt{nI(\theta_0)} - 2 \ln k \leq Z \leq 0)}, & \text{if } -C \sqrt{nI(\theta_0)} + \sqrt{2 \ln k} \leq 0 \end{cases} \]
\[ 7(b) = P_0(\theta_0 - C \leq \hat{\theta} \leq \theta_0) = P_0(-C \sqrt{n I(\theta_0)} \leq Z \leq 0) \]

\[
\begin{aligned}
(7) = \begin{cases} 
P(-C \sqrt{n I(\theta_0)} + \sqrt{2 \ln k} \leq Z \leq 0) = \Phi(0) - \Phi(-C \sqrt{n I(\theta_0)} + \sqrt{2 \ln k}), & \text{if } n > \frac{2 \ln k}{C^2 I(\theta_0)} \\
0, & \text{if } n < \frac{2 \ln k}{C^2 I(\theta_0)}
\end{cases}
\end{aligned}
\]

\[ (8) = 0 \text{ by the definition of MLE.} \]

\[ S_0 = (5) + (6) + (7) + (8) \]

\[
\begin{aligned}
= \begin{cases} 
\Phi(C \sqrt{n I(\theta_0)} - \sqrt{2 \ln k}) - \Phi(-C \sqrt{n I(\theta_0)} + \sqrt{2 \ln k}), & \text{if } n > \frac{2 \ln k}{C^2 I(\theta_0)} \\
0, & \text{if } n < \frac{2 \ln k}{C^2 I(\theta_0)}
\end{cases}
\end{aligned}
\]

\[
\begin{aligned}
= \begin{cases} 
2 \Phi(C \sqrt{n I(\theta_0)} - \sqrt{2 \ln k}) - 1, & \text{if } n > \frac{2 \ln k}{C^2 I(\theta_0)} \\
0, & \text{if } n < \frac{2 \ln k}{C^2 I(\theta_0)}
\end{cases}
\end{aligned}
\]

Probability of observing misleading evidence \((M_1)\) when \(\theta_1\) is assumed true:

\[ M_1 = P_1 \left( \sup_{\theta' \in \Theta_1} \frac{L(\theta'; x)}{L(\theta_0; x)} \leq \frac{1}{k} \right) \]

\[ = P_1 \left( \sup_{\theta' \in \Theta_0} \frac{L(\theta'; x)}{L(\theta_0; x)} \geq k \right) \]

\[ = P_1 \left[ \frac{L(\theta_0 + C)}{L(\hat{\theta})} \geq k | \hat{\theta} > \theta_0 + C \right] \cdot P_1(\hat{\theta} > \theta_0 + C) \]

\[ = P_1 \left[ \frac{L(\hat{\theta})}{L(\theta_0 + C)} \geq k | \theta_0 \leq \hat{\theta} \leq \theta_0 + C \right] \cdot P_1(\theta_0 \leq \hat{\theta} \leq \theta_0 + C) \]

\[ \geq \frac{10(a)}{10(b)} \]

---

\[ 10(a) \]

\[ 10(b) \]
\[
L(\hat{\theta}) \\
\begin{align*}
\frac{L(\hat{\theta})}{L(\theta_0 - C)} & \geq k |\theta_0 - C \leq \hat{\theta} \leq \theta_0 | \cdot P_1(\theta_0 - C \leq \hat{\theta} \leq \theta_0) \\
& \geq k |\theta_0 - \theta_0 - C | \cdot P_1(\theta_0 - C < \hat{\theta} - \theta_0) \\
\end{align*}
\]

\[
L_1(\theta) = P_1(\frac{L(\hat{\theta})}{L(\theta_0 - C)} \geq k |\theta_0 - C \leq \hat{\theta} \leq \theta_0 |) \\
L_2(\theta) = P_1(\frac{L(\hat{\theta})}{L(\theta_0 - C)} \geq k |\theta_0 - C < \hat{\theta} - \theta_0 |)
\]

\[
(9) = 0 \text{ by the definition of MLE}
\]

\[
10(a) = P_1(\frac{L(\hat{\theta})}{L(\theta_0 + C)} \geq k |\theta_0 \leq \hat{\theta} \leq \theta_0 + C |)
\]

\[
10(a) = P_1[2 \ln \frac{L(\hat{\theta})}{L(\theta_1)} + 2 \ln \frac{L(\theta_1)}{L(\theta_0 + C)} - 2 \ln k \geq 0 |\theta_0 \leq \hat{\theta} \leq \theta_0 + C |)
\]

Since \(\theta_0 \leq \hat{\theta} \leq \theta_0 + C\) implies \((\theta_0 - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} \leq \frac{\hat{\theta} - \theta_1}{\sqrt{n\mathcal{I}(\theta_1)}} \leq (\theta_0 + C - \theta_1) \sqrt{n\mathcal{I}(\theta_1)}\),

\[
2 \ln \frac{L(\hat{\theta})}{L(\theta_1)} |\theta_0 \leq \hat{\theta} \leq \theta_0 + C \xrightarrow{n \to \infty} Z^2, \text{ and } 2 \ln \frac{L(\theta_1)}{L(\theta_0 + C)} |\theta_0 \leq \hat{\theta} \leq \theta_0 + C \xrightarrow{n \to \infty} -2(\theta_0 + C - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} Z + (\theta_0 + C - \theta_1)^2 n\mathcal{I}(\theta_1), \text{ where } Z = T \mathcal{N}(0, 1, (\theta_0 - \theta_1) \sqrt{n\mathcal{I}(\theta_1)}, (\theta_0 + C - \theta_1) \sqrt{n\mathcal{I}(\theta_1)}).
\]

Therefore, \(10(a) = P_1(Z^2 - 2(\theta_0 + C - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} Z + (\theta_0 + C - \theta_1)^2 n\mathcal{I}(\theta_1) - 2 \ln k \geq 0) = P_1(Z > (\theta_0 + C - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} + \sqrt{2 \ln k}) + P_1(Z < (\theta_0 + C - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} - \sqrt{2 \ln k})
\]

If \((\theta_0 + C - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} - \sqrt{2 \ln k} \geq (\theta_0 - \theta_1) \sqrt{n\mathcal{I}(\theta_1)}, \text{ i.e. } n \geq \frac{2 \ln k}{C \sqrt{n\mathcal{I}(\theta_1)}}\),

\[
10(a) = P_1[(\theta_0 - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} < Z < (\theta_0 + C - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} - \sqrt{2 \ln k}] \]

\[
= \frac{P_1[(\theta_0 - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} < Z < (\theta_0 + C - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} - \sqrt{2 \ln k}]}{P_1[(\theta_0 - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} < Z < (\theta_0 + C - \theta_1) \sqrt{n\mathcal{I}(\theta_1)}]}
\]

If \((\theta_0 + C - \theta_1) \sqrt{n\mathcal{I}(\theta_1)} - \sqrt{2 \ln k} < (\theta_0 - \theta_1) \sqrt{n\mathcal{I}(\theta_1)}, \text{ i.e. } n < \frac{2 \ln k}{C \sqrt{n\mathcal{I}(\theta_1)}}\)

\[
10(a) = 0
\]
$10(b) = P_1(\theta_0 < \hat{\theta} < \theta_0 + C) = P_1[(\theta_0 - \theta_1)\sqrt{n I(\theta_1)} < Z < (\theta_0 + C - \theta_1)\sqrt{n I(\theta_1)}]$

$10 = 10(a) \cdot 10(b)$

$= P_1[(\theta_0 - \theta_1)\sqrt{n I(\theta_1)} < Z < (\theta_0 + C - \theta_1)\sqrt{n I(\theta_1)} - \sqrt{2 \ln k}]$

$= \Phi[(\theta_0 + C - \theta_1)\sqrt{n I(\theta_1)} - \sqrt{2 \ln k}] - \Phi[(\theta_0 - \theta_1)\sqrt{n I(\theta_1)}]$

$11(a) = P_1\left[\frac{L(\hat{\theta})}{L(\theta_0 - C)} \geq k | \theta_0 - C \leq \hat{\theta} \leq \theta_0\right]$

$= P_1[2 \ln \frac{L(\hat{\theta})}{L(\theta_1)} + 2 \ln \frac{L(\theta_1)}{L(\theta_0 - C)} - 2 \ln k \geq 0 | \theta_0 - C \leq \hat{\theta} \leq \theta_0]$

Since $\theta_0 - C \leq \hat{\theta} \leq \theta_0$ implies $(\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} \leq \frac{\theta_0 - \theta_1}{\sqrt{n I(\theta_1)}} \leq (\theta_0 - \theta_1)\sqrt{n I(\theta_1)}$, $2 \ln \frac{L(\hat{\theta})}{L(\theta_1)} | \theta_0 - C \leq \hat{\theta} \leq \theta_0 \xrightarrow{D_{n \to \infty}} Z_{**}^2$, and $2 \ln \frac{L(\theta_1)}{L(\theta_0 - C)} | \theta_0 - C \leq \hat{\theta} \leq \theta_0 \xrightarrow{D_{n \to \infty}} 2(\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} Z_{**} + (\theta_0 - C - \theta_1)^2 n I(\theta_1)$, where $Z_{**} = T N(0, 1, (\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)}), (\theta_0 - \theta_1)\sqrt{n I(\theta_1)}$.

Therefore, $P_1[2 \ln \frac{L(\hat{\theta})}{L(\theta_1)} + 2 \ln \frac{L(\theta_1)}{L(\theta_0 - C)} - 2 \ln k \geq 0 | \theta_0 - C \leq \hat{\theta} \leq \theta_0] = P_1[Z_{**}^2 + 2(\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} Z_{**} + (\theta_0 - C - \theta_1)^2 n I(\theta_1) - 2 \ln k \geq 0] = P_1[Z_{**} \geq (\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} + \sqrt{2 \ln k}] + P_1[Z_{**} \leq (\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} - \sqrt{2 \ln k}] = P_1[Z_{**} \geq (\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} + \sqrt{2 \ln k}] + 0$

If $(\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} + \sqrt{2 \ln k} \leq (\theta_0 - \theta_1)\sqrt{n I(\theta_1)}$, i.e. $n \geq \frac{2 \ln k}{C^2 I(\theta_1)}$,

$11(a) = P_1[(\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} + \sqrt{2 \ln k} < Z_{**} < (\theta_0 - \theta_1)\sqrt{n I(\theta_1)}]$  

$= \frac{P_1[(\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} + \sqrt{2 \ln k} < Z < (\theta_0 - \theta_1)\sqrt{n I(\theta_1)}]}{P_1[(\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} < Z < (\theta_0 - \theta_1)\sqrt{n I(\theta_1)}]}$

If $(\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} + \sqrt{2 \ln k} > (\theta_0 - \theta_1)\sqrt{n I(\theta_1)}$, i.e. $n < \frac{2 \ln k}{C^2 I(\theta_1)}$.

$11(a) = 0.$

$11(b) = P_1(\theta_0 - C \leq \hat{\theta} \leq \theta_0) = P_1[(\theta_0 - C - \theta_1)\sqrt{n I(\theta_1)} \leq Z \leq (\theta_0 - \theta_1)\sqrt{n I(\theta_1)}]$
\[
S_1 = P_1 \left( \sup_{\theta^* \in \Theta_1} \frac{L(\theta^*; x)}{L(\theta_0; x)} \geq k \right) 
= P_1 \left[ \frac{L(\hat{\theta})}{L(\theta_0 + c)} \geq k | \hat{\theta} \geq \theta_0 + C \right] \cdot P_1(\hat{\theta} \geq \theta_0 + C) \tag{13(a)} 
+ P_1 \left[ \frac{L(\theta_0 + C)}{L(\hat{\theta})} \geq k | \theta_0 \leq \hat{\theta} \leq \theta_0 + C \right] \cdot P_1(\theta_0 \leq \hat{\theta} \leq \theta_0 + C) \tag{13(b)} 
+ P_1 \left[ \frac{L(\theta_0 - C)}{L(\hat{\theta})} \geq k | \hat{\theta} \leq \theta_0 - C \right] \cdot P_1(\hat{\theta} \leq \theta_0 - C) \tag{15} 
+ P_1 \left[ \frac{L(\theta_0 - C)}{L(\theta_0 - C)} \geq k | \theta_0 - C \leq \hat{\theta} \leq 0 \right] \cdot P_1(\theta_0 - C \leq \hat{\theta} \leq 0) \tag{16}
\]

Since \( \hat{\theta} \geq \theta_0 + C \) implies \( \frac{(\hat{\theta} - \theta_1)}{\sqrt{n}I(\theta_1)} \rightarrow \frac{(\theta_0 + C - \theta_1)}{\sqrt{n}I(\theta_1)} \),

\[
2 \log \frac{L(\hat{\theta})}{L(\theta_0)} | \hat{\theta} \geq \theta_0 + C \rightarrow D \overset{D_{n \to \infty}}{\rightarrow} Z^2 \text{ and } 2 \log \frac{L(\theta_0)}{L(\theta_0 + C)} | \theta_0 \rightarrow D \overset{D_{n \to \infty}}{\rightarrow} 0 \text{ as } \theta_0 \rightarrow C - \theta_1, \]

where \( Z = T N(0, 1, (\theta_0 + C - \theta_1)^2 n I(\theta_1), \infty) \),

\[
13(a) = P_1 \left[ 2 \log \frac{L(\hat{\theta})}{L(\theta_0)} + 2 \log \frac{L(\theta_0)}{L(\theta_0 + C)} - 2 \log k \geq 0 \right] \rightarrow 0 \text{ as } \theta_0 \rightarrow C - \theta_1 \]

Hence,

\[
13(a) = P_1 \left[ Z^2 \geq -2(\theta_0 + C - \theta_1) \sqrt{n I(\theta_1)} + (\theta_0 + C - \theta_1)^2 n I(\theta_1) - 2 \log k \geq 0 \right]
\]
Appendix A.

\[
\begin{align*}
&P_1[Z^2\geq (\theta_0 + C - \theta_1) \sqrt{nI(\theta_1)} + \sqrt{2\ln k}] + P_1[Z^2\leq (\theta_0 + C - \theta_1) \sqrt{nI(\theta_1)} - \sqrt{2\ln k}] \\
&= P_1[Z^2\geq (\theta_0 + C - \theta_1) \sqrt{nI(\theta_1)} + \sqrt{2\ln k}] + 0 \\
&= P_1[Z\geq (\theta_0 + C - \theta_1) \sqrt{nI(\theta_1)} + \sqrt{2\ln k}] \\
&\quad / P_1[Z\geq (\theta_0 + C - \theta_1) \sqrt{nI(\theta_1)}] \\
13(b) &= P_1[\hat{\theta}\geq \theta_0 + C] = P_1[Z\geq (\theta_0 + C - \theta_1) \sqrt{nI(\theta_1)}] \\
13 &= P_1[Z\geq (\theta_0 + C - \theta_1) \sqrt{nI(\theta_1)} + \sqrt{2\ln k}] = \Phi[(\theta_0 + C - \theta_1) \sqrt{nI(\theta_1)} - \sqrt{2\ln k}] \\
14 &= 0 \text{ by the definition of MLE} \\
15 &= 0 \text{ by the definition of MLE} \\
16(a) &= P_1[2\ln \frac{L(\hat{\theta})}{L(\theta_1)} + 2\ln \frac{L(\theta_1)}{L(\theta_0 - C)} - 2\ln k \geq 0 | \hat{\theta}\leq \theta_0 - C] \\
&\text{Since } \hat{\theta}\leq \theta_0 - C \text{ implies } \frac{\hat{\theta} - \theta_0}{\sqrt{nI(\theta_1)}} \leq (\theta_0 - C - \theta_1) \sqrt{nI(\theta_1)}, \\
&2\ln \frac{L(\hat{\theta})}{L(\theta_1)} | \hat{\theta}\leq \theta_0 - C \xrightarrow{D_{n\to\infty}} Z^2_{**}; \text{ and } 2\ln \frac{L(\theta_1)}{L(\theta_0 - C)} | \hat{\theta}\leq \theta_0 - C \xrightarrow{D_{n\to\infty}} -2(\theta_0 - C - \theta_1) \sqrt{nI(\theta_1)} Z^2_{**} + \\
&(\theta_0 - C - \theta_1)^2 nI(\theta_1), \text{ where } Z^2_{**} = TN(0, 1, -\infty, (\theta_0 - C - \theta_1) \sqrt{nI(\theta_1)}) \\
16(a) &= P_1[Z^2_{**} - 2(\theta_0 - C - \theta_1) \sqrt{nI(\theta_1)} Z^2_{**} + nI(\theta_1)(\theta_0 - C - \theta_1)^2 - 2\ln k \geq 0] \\
&= P_1[Z_{**}\geq (\theta_0 - C - \theta_1) \sqrt{nI(\theta_1)} + \sqrt{2\ln k}] + P_1[Z_{**}\leq (\theta_0 - C - \theta_1) \sqrt{nI(\theta_1)} - \sqrt{2\ln k}] \\
&= 0 + P_1[Z_{**}\leq (\theta_0 - C - \theta_1) \sqrt{nI(\theta_1)} - \sqrt{2\ln k}] \\
&= P_1[Z\leq (\theta_0 - C - \theta_1) \sqrt{nI(\theta_1)} - \sqrt{2\ln k}] \\
&\quad / P_1[Z\leq (\theta_0 - C - \theta_1) \sqrt{nI(\theta_1)}] \\
16(b) &= P_1[\hat{\theta} < \theta_0 - C] = P_1[Z < (\theta_0 - C - \theta_1) \sqrt{nI(\theta_1)}] \\
16 &= 16(a) \cdot 16(b) = \Phi[(\theta_0 - C - \theta_1) \sqrt{nI(\theta_1)} - \sqrt{2\ln k}]
\[ S_1 = (13) + (14) + (15) + (16) \]

\[ = \Phi[-(\theta_0 + C - \theta_1)\sqrt{n\mathcal{I}(\theta_1)} - \sqrt{2 \ln k}] + \Phi[(\theta_0 - C - \theta_1)\sqrt{n\mathcal{I}(\theta_1)} - \sqrt{2 \ln k}] \]

\[ W_1 = 1 - M_1 - S_1 \]

\[ = \begin{cases} 
\Phi[-(\theta_0 + C - \theta_1)\sqrt{n\mathcal{I}(\theta_1)} + \sqrt{2 \ln k}] - \Phi[-(\theta_0 + C - \theta_1)\sqrt{n\mathcal{I}(\theta_1)} - \sqrt{2 \ln k}] \\
+ \Phi[(\theta_0 - C - \theta_1)\sqrt{n\mathcal{I}(\theta_1)} + \sqrt{2 \ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n\mathcal{I}(\theta_1)} - \sqrt{2 \ln k}], & \text{if } n \geq \frac{2 \ln k}{c^2 \mathcal{I}(\theta_1)} \\
\Phi[(\theta_0 + C - \theta_1)\sqrt{n\mathcal{I}(\theta_1)} + \sqrt{2 \ln k}] - \Phi[(\theta_0 - C - \theta_1)\sqrt{n\mathcal{I}(\theta_1)} - \sqrt{2 \ln k}], & \text{if } n < \frac{2 \ln k}{c^2 \mathcal{I}(\theta_1)} 
\end{cases} \]
A.3 Derivation of the operational characteristics for multi-parameter parametric models

Here, we derive the operational characteristics associated with the GLR for multi-parameter parametric models, with the presence of nuisance parameters.

Let $\theta$ represent the parameter of interest and $\gamma$ the nuisance parameter, then the full likelihood is given by $L(\theta, \gamma) = \prod_i^n f(x_i; \theta, \gamma)$.

We define the profile likelihood function for $\theta$, as $L_P(\theta) = \max_{\gamma}(L(\theta, \gamma)) = L(\theta, \hat{\gamma}(\theta))$. In the following derivation, we consider the case where both $\theta$ and $\gamma$ are scalars.

Let $\hat{\theta}_P = \arg \max L_P(\theta)$ and $\hat{\gamma}_\theta = \arg \max_{\gamma}(L(\theta, \gamma))$, and it is easy to see that $\hat{\theta}_P$ and $\hat{\gamma}_\theta$ are the maximum likelihood estimators $(\theta, \gamma) = \arg \max_{\theta, \gamma} L(\theta, \gamma)$. From here on, we will simply use $\hat{\theta}$ instead of $\hat{\theta}_P$ to denote the MLE of $\theta$.

We now assume that $X_1, X_2, \ldots, X_n$ are iid with probability density function $f(x; \theta, \gamma)$. Also define

$$\begin{bmatrix} I_{\theta\theta} & I_{\theta\gamma} \\ I_{\gamma\theta} & I_{\gamma\gamma} \end{bmatrix} = \begin{bmatrix} \mathbb{E}(-\frac{\partial^2 \ln f(x; \theta, \gamma)}{\partial \theta^2}) & \mathbb{E}(-\frac{\partial^2 \ln f(x; \theta, \gamma)}{\partial \theta \partial \gamma}) \\ \mathbb{E}(-\frac{\partial^2 \ln f(x; \theta, \gamma)}{\partial \gamma \partial \theta}) & \mathbb{E}(-\frac{\partial^2 \ln f(x; \theta, \gamma)}{\partial \gamma^2}) \end{bmatrix}$$

Probability of observing misleading evidence ($M_0$) when $\theta_0$ is assumed true:

$$M_0 = P_0\left[ \sup_{\theta'' \in \Theta_1} \frac{L_P(\theta''); x}{\sup_{\theta' \in \Theta_0} L_P(\theta'); x} \geq k \right]$$

$$= P_0\left[ \frac{L_P(\hat{\theta})}{L_P(\theta_0 + C)} \geq k \right] \cdot P_0(\hat{\theta} \geq \theta_0 + C) \cdot P_0(0 \leq \hat{\theta} \leq \theta_0 + C) \cdot P_0(\theta_0 \leq \hat{\theta} \leq \theta_0 + C) \cdot P_0(\theta_0 \geq \hat{\theta} \leq \theta_0)$$

$$+ P_0\left[ \frac{L_P(\hat{\theta})}{L_P(\theta_0 - C)} \geq k \right] \cdot P_0(\hat{\theta} \leq \theta_0 - C) \cdot P_0(-C \leq \hat{\theta} \leq \theta_0 + C) \cdot P_0(\theta_0 \leq \hat{\theta} \leq \theta_0 + C)$$

$$+ P_0\left[ \frac{L_P(\hat{\theta})}{L_P(\theta_0 - C)} \geq k \right] \cdot P_0(\hat{\theta} \leq \theta_0 - C) \cdot P_0(\theta_0 \leq \hat{\theta} \leq \theta_0 + C) \cdot P_0(\theta_0 \leq \hat{\theta} \leq \theta_0 + C)$$
$+ P_0 \left[ \frac{L_P(\hat{\theta})}{L_P(\theta_0 + C)} \right] \geq k |\hat{\theta} - \theta_0 - C| \cdot P_0(\hat{\theta} \leq \theta_0 - C) (4(a))$

$= P_0 \left[ 2 \ln \frac{L_P(\hat{\theta})}{L_P(\theta_0 + C)} + 2 \ln \frac{L_P(\theta_0)}{L_P(\theta_0 + C)} - 2 \ln k |\hat{\theta} - \theta_0 + C| \right] (4(b))$

$= P_0 \left[ 2 \ln \frac{L_P(\hat{\theta})}{L_P(\theta_0)} \right]_n \to \infty \xrightarrow{D} \left( \frac{\hat{\theta} - \theta_0}{1 - \rho_{\theta_0}^2 (1 - n \theta_0 \gamma_0)} \right)^2$

$= \lambda(n) = \chi^2,$

where $\rho_{\theta_0}^2 = \frac{\mathcal{I}_{\theta_0} \mathcal{I}_{\gamma_0} \theta_0}{\mathcal{I}_{\theta_0} \mathcal{I}_{\gamma_0} \theta_0}.$

Since $\hat{\theta} \geq \theta_0 + C$ implies

$\frac{\hat{\theta} - \theta_0}{1 - \rho_{\theta_0}^2 (1 - n \theta_0 \gamma_0)} \geq C \sqrt{n \mathcal{I}_{\theta_0} \mathcal{I}_{\gamma_0} \theta_0 (1 - \rho_{\theta_0}^2 (1 - n \theta_0 \gamma_0))}, \quad 2 \ln \frac{L_P(\hat{\theta})}{L_P(\theta_0)} |\hat{\theta} \geq \theta_0 + C \xrightarrow{n \to \infty} Z^2_\ast,$

where $Z_\ast$ is $TN(0, 1, C \sqrt{n \mathcal{I}_{\theta_0} \mathcal{I}_{\gamma_0} \theta_0 (1 - \rho_{\theta_0}^2 (1 - n \theta_0 \gamma_0))}, \infty).$

Now we will show that $2 \ln \frac{L_P(\theta_0)}{L_P(\theta_0 + C)} \mid \hat{\theta} \geq \theta_0 + C \xrightarrow{n \to \infty} Z_\ast.$

$2 \ln \frac{L_P(\theta_0)}{L_P(\theta_0 + C)}$

$= 2 \ell_p(\theta_0) - 2 \ell_p(\theta_0 + C)$

$= 2 \ell_p(\theta_0) - 2 \left[ \ell_p(\theta_0) + C \ell_p'(\theta_0) + \frac{1}{2} C^2 \ell_p''(\theta_0) + \frac{1}{6} C^3 \ell_p'''(\theta_0^*) \right]$

Taylor series expansion of $\ell_p(\theta_0 + C)$ at $\theta_0$

$= -2 C \ell_p'(\theta_0) - C^2 \ell_p''(\theta_0) - \frac{1}{3} C^3 \ell_p'''(\theta_0^*),$

where $\theta_0^*$ lies between $\theta_0$ and $\theta_0 + C.$
Taylor series expand $\ell'_P(\hat{\theta}) = \left. \frac{\partial \ell(\theta, \gamma)}{\partial \theta} \right|_{\hat{\theta}, \gamma(\hat{\theta})}$ at $\theta_0$, we get

$$\ell'_P(\hat{\theta}) = \ell'_P(\theta_0) + (\hat{\theta} - \theta_0)\ell''_P(\theta_0) + \frac{(\hat{\theta} - \theta_0)^2}{2}\ell'''_P(\theta_n^{**}),$$

where $\theta_n^{**}$ lies between $\theta$ and $\hat{\theta}$.

Since $\ell'_P(\hat{\theta}) = 0$ by the definition of MLE,

$$\ell'_P(\theta_0) = - (\hat{\theta} - \theta_0)\ell''_P(\theta_0) - \frac{(\hat{\theta} - \theta_0)^2}{2}\ell'''_P(\theta_n^{**})
\quad = \frac{-\ell''_P}{\sqrt{nI_{\theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)}} Z = N(0, 1),$$

Note that $\ell''_P(\theta_0) \rightarrow -I_{\theta_0 \theta_0} (1 - \rho_{\theta_0 \gamma_0}^2)$ (Royall, 2000, P767).

Then $2 \ln \frac{L_P(\theta_0)}{L_P(\theta_0 + C)} = -2C \sqrt{nI_{\theta_0 \theta_0} (1 - \rho_{\theta_0 \gamma_0}^2) Z + C^2 nI_{\theta_0 \theta_0} (1 - \rho_{\theta_0 \gamma_0}^2)}.$

Since $\hat{\theta} \geq \theta_0 + C$ implies

$$\sqrt{nI_{\theta_0 \theta_0} (1 - \rho_{\theta_0 \gamma_0}^2)} \geq C \sqrt{nI_{\theta_0 \theta_0} (1 - \rho_{\theta_0 \gamma_0}^2)},$$

$$2 \ln \frac{L_P(\theta_0)}{L_P(\theta_0 + C)} \hat{\theta} \geq \theta_0 + C = -2C \sqrt{nI(\theta_0) Z_\ast + C^2 nI(\theta_0)},$$

where $Z_\ast = TN(0, 1, C \sqrt{nI_{\theta_0 \theta_0} (1 - \rho_{\theta_0 \gamma_0}^2)}, \infty).$
(1)\(= P_0(\hat{\theta} \geq \theta_0 + C) = P_0\left(\frac{\hat{\theta} - \theta_0}{\sqrt{nI_{\theta_0}(1 - \rho_{\theta_0})}} \geq C \sqrt{nI_{\theta_0}(1 - \rho_{\theta_0})}\right)\)
\[= P_0\left(Z \geq C \sqrt{nI_{\theta_0}(1 - \rho_{\theta_0})}\right)\]
\[= P_0\left(Z \geq C \sqrt{nI_{\theta_0}(1 - \rho_{\theta_0})} + \sqrt{2 \ln k}\right)\]
\[= \Phi(-C \sqrt{nI_{\theta_0}(1 - \rho_{\theta_0})} - \sqrt{2 \ln k})\]

(2)\(= 0\) by the definition of MLE

(3)\(= 0\) by the definition of MLE

(4)\(= P_0\left[\frac{L_P(\hat{\theta})}{L_P(\theta - C)} \geq k | \hat{\theta} \leq \theta_0 - C\right]\)
\[= P_0\left[2 \ln \frac{L_P(\hat{\theta})}{L_P(\theta_0)} + 2 \ln \frac{L_P(\theta_0)}{L_P(\theta_0 - C)} - 2 \ln k \geq 0 | \hat{\theta} \leq \theta_0 - C\right]\]

Since \(\hat{\theta} \leq \theta_0 - C\) implies \(\frac{\hat{\theta} - \theta_0}{\sqrt{nI_{\theta_0}(1 - \rho_{\theta_0})}} \leq -C \sqrt{nI_{\theta_0}(1 - \rho_{\theta_0})}\),
\(2 \ln \frac{L_P(\hat{\theta})}{L_P(\theta_0)} \hat{\theta} \leq \theta_0 - C \xrightarrow{n \to \infty} Z^{**}\),
where \(Z^{**} = \text{TN}(0, 1, -\infty, -C \sqrt{nI_{\theta_0}(1 - \rho_{\theta_0})})\)

\(2 \ln \frac{L_P(\theta_0)}{L_P(\theta_0 - C)}\)
\[= 2 \ell_P(\theta_0) - 2 \ell_P(\theta_0 - C) = 2 \ell_P(\theta_0) - 2[\ell_P(\theta_0) - C \ell_P(\theta_0) + \frac{1}{2} C^2 \ell''_P(\theta_0)]\]
\[= 2C \ell_P'(\theta_0) - C^2 \ell''_P(\theta_0) + \frac{1}{3} \ell'''_P(\theta_0)\]
\[= 2C \sqrt{nI_{\theta_0}(1 - \rho_{\theta_0})} \frac{\ell_0'(\theta_0)}{\sqrt{nI_{\theta_0}(1 - \rho_{\theta_0})}} - C^2 \ell''_P(\theta_0) + \frac{1}{3} \ell'''_P(\theta_0)\]
\[\xrightarrow{n \to \infty} 2C \sqrt{nI_{\theta_0}(1 - \rho_{\theta_0})} Z^{**} + C^2 nI_{\theta_0}(1 - \rho_{\theta_0})\]
\[ 4(a) = P_0 \left[ \frac{L_P(\hat{\theta})}{L(\theta - C)} \geq k | \hat{\theta} \leq \theta_0 - C \right] \]
\[ = P_0 \left[ Z^2_{**} + 2C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} Z^2_{**} + C^2nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2) - 2 \ln k \geq 0 \right] \]
\[ = P_0 \left[ Z^{**} \geq -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} + \sqrt{2 \ln k} \right] + P_0 \left[ Z^{**} \leq -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} - \sqrt{2 \ln k} \right] \]
\[ = 0 + P_0 \left[ Z^{**} \leq -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} - \sqrt{2 \ln k} \right] \]
\[ = \frac{P_0 \left( Z \leq -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} \right)}{P_0 \left( Z \leq -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} \right)} \]

\[ 4(b) = P_0(\hat{\theta} \leq \theta_0 - C) \]
\[ = P_0 \left( \frac{\hat{\theta} - \theta_0}{\sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)}} \right) \leq -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} \]
\[ = P_0 \left( Z \leq -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} \right) \]

\[ 4 = 4(a) \cdot 4(b) \]
\[ = \frac{P_0 \left( Z \leq -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} - \sqrt{2 \ln k} \right)}{P_0 \left( Z \leq -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} \right)} \cdot P_0 \left( Z \leq -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} \right) \]
\[ = P_0 \left( Z \leq -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} - \sqrt{2 \ln k} \right) \]
\[ = \Phi \left( -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} - \sqrt{2 \ln k} \right) \]

\[ M_0 = (1) + (2) + (3) + (4) \]
\[ = \Phi \left( -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} - \sqrt{2 \ln k} \right) + 0 + 0 + \Phi \left( -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} - \sqrt{2 \ln k} \right) \]
\[ = 2 \Phi \left( -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0}^2)} - \sqrt{2 \ln k} \right) \]
\[ S_0 = P_0 \left[ \sup_{\theta' \in \Theta_0} \frac{L_P(\theta'; x)}{\sup_{\theta'' \in \Theta_1} L_P(\theta''; x)} \geq k \right] \]

\[ = P_0 \left[ \frac{L_P(\theta_0 + C)}{L_P(\hat{\theta})} \geq k | \hat{\theta} \geq \theta_0 + C \right] \cdot P_0(\hat{\theta} \geq \theta_0 + C) \quad (5) \]

\[ + \frac{P_0[\frac{L_P(\hat{\theta})}{L_P(\theta_0 + C)} \geq k | \theta_0 \leq \hat{\theta} \leq \theta_0 + C]}{6(a)} \cdot \frac{P_0[\theta_0 \leq \hat{\theta} \leq \theta_0 + C]}{6(b)} \]

\[ + \frac{P_0[\frac{L_P(\hat{\theta})}{L_P(\theta_0 - C)} \geq k | \theta_0 - C \leq \hat{\theta} \leq \theta_0]}{7(a)} \cdot \frac{P_0[\theta_0 - C \leq \hat{\theta} \leq \theta_0]}{7(b)} \]

\[ + \frac{P_0[\frac{L_P(\theta_0 - C)}{L_P(\theta)} \geq k | \theta \leq \theta_0 - C]}{8(a)} \cdot \frac{P_0[\theta \leq \theta_0 - C]}{8(b)} \]

\[(5) = 0 \text{ by the definition of the MLE.}\]

\[ 6(a) = \begin{cases} 
  \frac{P_0(0 \leq Z \leq C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0})} - \sqrt{2 \ln k})}{P_0(0 \leq Z \leq C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0})})}, & \text{if } C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0})} - \sqrt{2 \ln k} > 0 \\
  0, & \text{if } C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0})} - \sqrt{2 \ln k} \leq 0 
\end{cases} \]

\[ 6(b) = P_0(\theta_0 \leq \hat{\theta} \leq \theta_0 + C) = P_0(0 \leq Z \leq C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0})}) \]

\[ 6(\text{b}) = \begin{cases} 
  \Phi \left( C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0})} - \sqrt{2 \ln k} \right) - \Phi(0), & \text{if } n \geq \frac{2 \ln k}{C^2 nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0})} \\
  0, & \text{if } n < \frac{2 \ln k}{C^2 nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0})} 
\end{cases} \]

\[ 7(a) = \begin{cases} 
  P_0(-C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0})} + \sqrt{2 \ln k} \leq Z \leq 0), & \text{if } -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0})} + \sqrt{2 \ln k} > 0 \\
  0, & \text{if } -C \sqrt{nI_{\theta_0 \theta_0}(1 - \rho_{\theta_0 \gamma_0})} + \sqrt{2 \ln k} \leq 0 
\end{cases} \]
Appendix A.

\(7(b) = P_0(\theta_0 - C \leq \hat{\theta} \leq \theta_0) = P_0(-C \sqrt{nI_{\theta_0 \theta_0}} (1 - \rho_{\theta_0 \gamma_0}^2) \leq Z \leq 0)\)

\[
7 = \begin{cases} 
\Phi(0) - \Phi(-C \sqrt{nI_{\theta_0 \theta_0}} (1 - \rho_{\theta_0 \gamma_0}^2) + \sqrt{2 \ln k}), & \text{if } n \geq \frac{2 \ln k}{c^2 n I_{\theta_0 \theta_0} (1 - \rho_{\theta_0 \gamma_0}^2)} \\
0, & \text{if } n < \frac{2 \ln k}{c^2 n I_{\theta_0 \theta_0} (1 - \rho_{\theta_0 \gamma_0}^2)}
\end{cases}
\]

(8) = 0 by the definition of the MLE.

\[
S_0 = \begin{cases} 
\Phi(C \sqrt{nI_{\theta_0 \theta_0}} (1 - \rho_{\theta_0 \gamma_0}^2) - \sqrt{2 \ln k}) \\
-\Phi(-C \sqrt{nI_{\theta_0 \theta_0}} (1 - \rho_{\theta_0 \gamma_0}^2) + \sqrt{2 \ln k}), & \text{if } n > \frac{2 \ln k}{c^2 n I_{\theta_0 \theta_0} (1 - \rho_{\theta_0 \gamma_0}^2)} \\
0, & \text{if } n \leq \frac{2 \ln k}{c^2 n I_{\theta_0 \theta_0} (1 - \rho_{\theta_0 \gamma_0}^2)}
\end{cases}
\]

\[
W_0 = 1 - M_0 - S_0
\]

We will provide the results for \(M_1, W_1\) and \(S_1\) below without showing details of the derivations.

\[
M_1 = \begin{cases} 
\Phi\left[(\theta_0 + C - \theta_1) \sqrt{nI_{\theta_1 \theta_1}} (1 - \rho_{\theta_1 \gamma_1}^2) - \sqrt{2 \ln k}\right], & \text{if } n > \frac{2 \ln k}{c^2 n I_{\theta_1 \theta_1} (1 - \rho_{\theta_1 \gamma_1}^2)} \\
-\Phi\left[(\theta_0 - C - \theta_1) \sqrt{nI_{\theta_1 \theta_1}} (1 - \rho_{\theta_1 \gamma_1}^2) + \sqrt{2 \ln k}\right], & \text{if } n \leq \frac{2 \ln k}{c^2 n I_{\theta_1 \theta_1} (1 - \rho_{\theta_1 \gamma_1}^2)} \\
0, & \text{if } n \leq \frac{2 \ln k}{c^2 n I_{\theta_1 \theta_1} (1 - \rho_{\theta_1 \gamma_1}^2)}
\end{cases}
\]
Appendix A.

120

\[ S_1 = \Phi \left[ -(\theta_0 + C - \theta_1) \sqrt{nI_{\theta_\theta} (1 - \rho^2_{\theta_1 \gamma_1})} - \sqrt{2 \ln k} \right] + \Phi \left[ (\theta_0 - C - \theta_1) \sqrt{nI_{\theta_\theta} (1 - \rho^2_{\theta_1 \gamma_1})} - \sqrt{2 \ln k} \right] \]

\[
W_1 = \begin{cases} 
\Phi \left[ -(\theta_0 + C - \theta_1) \sqrt{nI_{\theta_\theta} (1 - \rho^2_{\theta_1 \gamma_1})} + \sqrt{2 \ln k} \right] \\
-\Phi \left[ -(\theta_0 + C - \theta_1) \sqrt{nI_{\theta_\theta} (1 - \rho^2_{\theta_1 \gamma_1})} - \sqrt{2 \ln k} \right] \\
\Phi \left[ (\theta_0 + C - \theta_1) \sqrt{nI_{\theta_\theta} (1 - \rho^2_{\theta_1 \gamma_1})} + \sqrt{2 \ln k} \right] \\
-\Phi \left[ (\theta_0 + C - \theta_1) \sqrt{nI_{\theta_\theta} (1 - \rho^2_{\theta_1 \gamma_1})} - \sqrt{2 \ln k} \right], & \text{if } n > \frac{2 \ln k}{C \ln n I_{\theta_\theta} (1 - \rho^2_{\theta_1 \gamma_1})} \\
\Phi \left[ (\theta_0 + C - \theta_1) \sqrt{nI_{\theta_\theta} (1 - \rho^2_{\theta_1 \gamma_1})} + \sqrt{2 \ln k} \right] \\
-\Phi \left[ (\theta_0 + C - \theta_1) \sqrt{nI_{\theta_\theta} (1 - \rho^2_{\theta_1 \gamma_1})} - \sqrt{2 \ln k} \right], & \text{if } n \leq \frac{2 \ln k}{C \ln n I_{\theta_\theta} (1 - \rho^2_{\theta_1 \gamma_1})}
\end{cases}
\]

Note that \( M_0 \) is bounded by \( 2\Phi(-\sqrt{2\ln k}) \) and \( M_1 \) is bounded by \( \Phi(-\sqrt{2\ln k}) \).

A.3.1 In the presence of multiple nuisance parameters

Let \( \theta \) denote the single parameter of interest and \( \gamma=(\gamma_1, \gamma_2, ..., \gamma_g) \) the \( g \) nuisance parameters, then Fisher's information matrix

\[
I = \begin{bmatrix}
I_{\theta \theta} & I_{\theta \gamma_1} & \cdots & I_{\theta \gamma_g} \\
I_{\gamma_1 \theta} & I_{\gamma_1 \gamma_1} & \cdots & I_{\gamma_1 \gamma_g} \\
\vdots & \vdots & \ddots & \vdots \\
I_{\gamma_g \theta} & I_{\gamma_g \gamma_1} & \cdots & I_{\gamma_g \gamma_g}
\end{bmatrix}
\]
where

\[ \begin{bmatrix} \mathcal{I}_{\theta\theta} & \mathcal{I}_{\theta\gamma} \\ \mathcal{I}_{\gamma\theta} & \mathcal{I}_{\gamma\gamma} \end{bmatrix} \]

The first element of the inverse of the Fisher's information matrix, \( \mathcal{I}^{-1} \), is

\[ b(\theta, \gamma) = \left[ \mathcal{I}_{\theta\theta} - \mathcal{I}_{\theta\gamma} \mathcal{I}_{\gamma\gamma}^{-1} \mathcal{I}_{\gamma\theta} \right]^{-1} \]

\[ = \left[ \mathcal{I}_{\theta\theta} (1 - \frac{\mathcal{I}_{\theta\gamma} \mathcal{I}_{\gamma\gamma}^{-1} \mathcal{I}_{\gamma\theta}}{\mathcal{I}_{\theta\theta}}) \right]^{-1} \]

\[ = \left[ \mathcal{I}_{\theta\theta} (1 - \rho_{\theta\gamma}^2) \right]^{-1}, \]

where \( \rho_{\theta\gamma}^2 = \frac{\mathcal{I}_{\theta\gamma} \mathcal{I}_{\gamma\gamma}^{-1} \mathcal{I}_{\gamma\theta}}{\mathcal{I}_{\theta\theta}}. \)

It is easy to see that M, W and S have the same formulation as in the one nuisance parameter case, except for \( \rho_{\theta\gamma}^2 = \frac{\mathcal{I}_{\theta\gamma} \mathcal{I}_{\gamma\gamma}^{-1} \mathcal{I}_{\gamma\theta}}{\mathcal{I}_{\theta\theta}} \) instead of \( \rho_{\theta\gamma}^2 = \frac{\mathcal{I}_{\theta\gamma} \mathcal{I}_{\gamma\gamma}}{\mathcal{I}_{\theta\theta} \mathcal{I}_{\gamma\gamma}}. \)