Empirical Likelihood and Bootstrap Inference with Constraints

by

Chunlin Wang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2017

© Chunlin Wang 2017

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Dongsheng Tu Professor (Queens University)
Supervisors	Dr. Pengfei Li Associate Professor
	Dr. Paul Marriott Professor
Internal Members	Dr. Yingli Qin Assistant Professor
	Dr. Changbao Wu Professor
Internal-external Member	Dr. Dinghai Xu Associate Professor (Department of Economics)

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

A version of Chapter 3 of this thesis has been prepared as a research paper submitted for publication in *Computational Statistics and Data Analysis*. A version of Chapter 4 of this thesis has been prepared as a research paper submitted for publication in *Journal of Multivariate Analysis*. Both papers are co-authored with my supervisors, where my contributions include deriving the main theorems, performing the simulations, analyzing a real data set, and writing the initial draft.

Abstract

Empirical likelihood and the bootstrap play influential roles in contemporary statistics. This thesis studies two distinct statistical inference problems, referred to as Part I and Part II, related to the empirical likelihood and bootstrap, respectively.

Part I of this thesis concerns making statistical inferences on multiple groups of samples that contain excess zero observations. A unique feature of the target populations is that the distribution of each group is characterized by a non-standard mixture of a singular distribution at zero and a skewed nonnegative component. In Part I of this thesis, we propose modelling the nonnegative components using a semiparametric, multiple-sample, density ratio model (DRM). Under this semiparametric setup, we can efficiently utilize information from the combined samples even with unspecified underlying distributions.

We first study the question of testing homogeneity of multiple nonnegative distributions when there is an excess of zeros in the data, under the proposed semiparametric setup. We develop a new empirical likelihood ratio (ELR) test for homogeneity and show that this ELR has a χ^2-type limiting distribution under the homogeneous null hypothesis. A nonparametric bootstrap procedure is proposed to calibrate the finite-sample distribution of the ELR. The consistency of this bootstrap procedure is established under both the null and alternative hypotheses. Simulation studies show that the bootstrap ELR test has an accurate nominal type I error, is robust to changes of underlying distributions, is competitive to, and sometimes more powerful than, several popular one- and two-part tests. A real data example is used to illustrate the advantages of the proposed test.

We next investigate the problem of comparing the means of multiple nonnegative distributions, with excess zero observations, under the proposed semiparametric setup. We develop a unified inference framework based on our new ELR statistic, and show that this ELR has a χ^2-type limiting distribution under a general null hypothesis. This allows us to construct a new test for mean equality. Simulation results show favourable performance of the proposed ELR test compared with other existing tests for mean equality, especially when the correctly specified basis function in the DRM is the logarithm function. A real data set is analyzed to illustrate the advantages of the proposed method.

In Part II of this thesis, we investigate the asymptotic behaviour of, the commonly used, bootstrap percentile confidence intervals when the parameters are subject to inequality constraints. We concentrate on the important one- and two-sample problems with data generated from distributions in the natural exponential family. Our attention is focused on quantifying asymptotic coverage probabilities of the percentile confidence intervals based on bootstrapping maximum likelihood estimators. We propose a novel local framework to study the subtle asymptotic behaviour of bootstrap percentile confidence intervals when the true parameter values are close to the boundary. Under this framework, we discover that when the true parameter is on, or close to, the restriction boundary, the local asymptotic coverage probabilities can always exceed the nominal level in the one-sample case; however, they can be, surprisingly, both under and over the nominal level in the two-sample case. The results provide theoretical justification and guidance on applying the bootstrap percentile method to constrained inference problems.

The two individual parts of this thesis are connected by being referred to as *constrained statistical inference*. Specifically, in Part I, the semiparametric density ratio model uses an exponential tilting constraint, which is a type of equality constraint, on the parameter space. In Part II, we deal with inequality constraints, such as a boundary or ordering constraints, on the parameter space. For both parts, an important regularity condition in traditional likelihood inference, that parameters should be interior points of the parameter space, is violated. Therefore, the respective inference procedures involve non-standard asymptotics that create new technical challenges.

Acknowledgements

First and foremost, I would like to reserve my heartfelt gratitude to my thesis supervisors, Dr. Pengfei Li and Dr. Paul Marriott, for your recognition, guidance, patience, and encouragement over past few years. I would have never explored so many elegant statistical thoughts and ideas without your inspiration and insightful discussions. I would have never realized my potential without your rigorous scientific attitude and high academic standard. Without you, I would have never become a better, wiser and stronger man in my life.

I am also very thankful to my thesis examining committee members for taking their precious time and effort to read my thesis carefully. All of your profound knowledge, generous sharing, and invaluable questions and comments have benefited me a lot.

Many thanks go to all the very kind staff members in the Department of Statistics and Actuarial Science. Especially, I wish to give my thanks to our department mother, Mary Lou Dufton, for her constant care during my study.

I am extremely lucky to have many great friends being around. All because of you, my memory in Waterloo is full of happiness.

There are still numerous individuals who I am indebted to. My any achievement was made possible with your efforts, no matter how small, that I can never forget.

Chunlin Wang
January 16th, 2017, in Waterloo

Dedication

To my beloved mom and dad for their unlimited love, support and understanding.

To the memory of my dearest grandpa.

Table of Contents

List of Tables

List of Figures

Chapter 1

Overview

1.1 Background and examples

This thesis exploits two important statistical inference problems, both of which take their motivation from real applications. We first introduce the research background by giving several motivating examples.

For Part I of this thesis, our research is motivated by the existence of a particular type of data which contains excess zero observations in addition to having skewed nonnegative outcomes. In many applications, multiple groups of samples with such important features are frequently encountered.

Example 1.1 *Muralidharan and Kale (2002) considered such a data set after monitoring rainfall distributions. Daily rainfall measurements were recorded over several years. There were often dry days which were recorded as having zero rainfall. All other data points were on a positive, continuous scale.*

Example 1.2 *Zhou and Tu (1999) provided an example involving the assessment of medical care expenditures. Here observations came from a control group and several intervention groups. In each group, a majority of inpatients had zero cost due to no hospitalizations during the study.*

There are many other examples in the literature, covering a broad range of applications, such as fishery surveys (Pennington, 1983), life sciences (Taylor and Pollard, 2009; Wagner et al., 2011; Gleiss et al., 2015), epidemiology (Bascoul-Mollevi et al., 2005; Bedrick and Hossain, 2013; Hu and Proschan, 2015), health economics (Tu and Zhou, 1999), reliability (Lambert, 1992), and tobacco consumption (Johnson et al., 2015). In addition, examples in more complex settings can be found in a recent special issue of the *Biometrical Journal* (Volume 58, Issue 2, March 2016) on "Models for Continuous Data with a Spike at Zero".

The scientific issues motivated by these examples often involve two questions.

Q1. How to test if several such populations are homogeneous? that is, testing if the data of each group are from the same underlying distribution?

Q2. How to make reliable inferences on the means of several such populations? for example, when testing mean equality?

Note that there are situations in which the excess zero observations are due to censoring or truncation (cf. Moulton and Halsey, 1995; Taylor et al., 2001). Modelling and analyzing the zero values in these situations may require more knowledge about the underlying data generating process, and thus are different from the "true" zeros that we consider in this thesis.

Part II of this thesis is devoted to situations where the parameters of interest are restricted by (linear) inequality constraints. These should be carefully differentiated from ones which have logical constraints (cf. Anaya-Izquierdo et al., 2014). The nature of these problems may be best introduced by using illustrative examples.

For a single scalar parameter, a boundary constraint can often be encountered in practice. Here we share two examples from physics and genetics.

Example 1.3 *Feldman and Cousins (1998) provided an example from high energy Physics when searching for neutrino oscillations. The observed signal can be formulated by the equation $X = S + B$, where the true signal S, and the background noise B, are two independent Poisson random variables, and B has a known mean b. Suppose the unknown* *mean of X is θ, and hence $\theta \geq b$. It is further claimed, in Woodroofe and Wang (2000), that the observed signal could be very weak which results in a challenging inference problem.*

Example 1.4 *Fu et al. (2006) investigated an example of genetic linkage analysis. The distribution of the number of recombinations between genetic traits and marker loci is usually described by a binomial distribution (Ott, 1999). The recombination fraction, denoted by θ, is a useful measure of genetic linkage. When the loci of two genes are completely unlinked, θ achieves the maximum possible value of $1/2$, and hence the recombination fraction θ is constrained by $\theta \in [0, 1/2]$.*

In multiple-sample problems, there are many real life examples where an ordering constraints can happen naturally. Here we mention two such examples.

Example 1.5 *Follmann (1996) presented an example from a case-control study involving blood pressure reduction. The treatment of sodium reduction is expected to have the effect of reducing blood pressure, relative to the control group. Inference on treatment effects can be made approximately by using normal distributions with ordered means.*

Example 1.6 *Li et al. (2010) considered an example from a pancreatic cancer biomarker study. For each patient, the serum samples were assayed for two antigens. Higher levels of the two antigens are known to be associated with higher risks of cancer. Therefore, for each antigen with categorized levels, the means of the binary outcome variable modelled by binomial probabilities should be ordered with regard to the levels.*

More motivating examples, with either boundary or ordering constraints, can be found in the book of Silvapulle and Sen (2004), and also in a special issue of the *Journal of Statistical Planning and Inference* (Volume 107, Issues 1-2, September 2002) on "Statistical Inference under Inequality Constraints".

Interval estimation of constrained parameters is a key measure of statistical accuracy, in additional to simply testing some pre-specified parameter point or set. In Part II of this thesis, we investigate confidence intervals constructed by, the frequently used, bootstrap percentile method. We aim to evaluate how reliable the bootstrap percentile confidence interval is by answering the following three specific questions.

Q3. Can it achieve nominal coverage level?

Q4. If not, does it over- or under-cover the true value of the parameter?

Q5. Can we more appropriately quantify the asymptotic coverage probability?

1.2 Main contributions and outline of the thesis

In this thesis, we study the empirical likelihood and bootstrap methods in problems that are motivated from a wide range of practical examples.

In Part I, we attempt to answer the aforementioned scientific questions Q1–Q2 arising in the context of multiple nonnegative distributions with excess zero observations. In many applications, multiple populations may naturally share some common characteristics. It is therefore desirable to borrow information across similar populations to improve inference procedures. At the same time, we also hope to avoid making too specific parametric model assumptions, which may not be realistic in applications. The density ratio model (DRM) is therefore attractive because of its semiparametric nature. We propose to link the distributions of the positive observations in multiple samples using the DRM, and build an empirical likelihood ratio (ELR) based inference framework. Under this semiparametric inference framework, we are able to exploit information from all the available data without having to specify the underlying distributions. This framework is valuable since it adds a semiparametric alternative to the existing class of fully parametric and nonparametric approaches for modelling and making inferences on multiple nonnegative distributions with excess zero observations. In Chapter 2, the DRM and its background are introduced.

In Chapter 3, we address the scientific question Q1 under the proposed semiparametric setup. We propose an ELR test for the homogeneity of multiple nonnegative distributions with excess zero observations. It is shown that the ELR has a χ^2-type limiting distribution under the homogeneous null hypothesis. In particular, the DRM assumption is always satisfied under this null hypothesis, and hence the asymptotic size of the ELR test can always be controlled at its nominal level even for misspecified basis functions. We further propose a nonparametric bootstrap procedure which improves the finite sample performance of the ELR test. The consistency of the proposed bootstrap procedure is established under both the null and alternative hypotheses. In addition, the proposed bootstrap ELR is computationally fast since it uses logistic regression routines available in standard statistical software, making it readily applicable in practice. The `R` code for implementing the bootstrap ELR test is also available in Appendix A.1. The theoretical and computational advantages of the proposed ELR test and the bootstrap procedure are illustrated by simulation studies and in a real data example.

In Chapter 4, we address the scientific question Q2 under the same semiparametric setup. We develop a unified inference framework, based on the ELR, for comparing the means of multiple nonnegative distributions with excess zero observations. Specifically, under the DRM assumption, the means of multiple samples can be estimated through unbiased estimating equations, and we show that the empirical likelihood method provides an effective inference platform. It is shown that this ELR has a χ^2-type limiting distribution under a general null hypothesis that includes mean equality as a special case. The numerical calculation of this proposed ELR statistic is also discussed. We have written `R` functions for implementing the ELR for testing mean equality, with the basis function in the DRM being the logarithm function, and they are available in Appendix A.2. We illustrate the good behaviour of the proposed ELR by finite sample simulation studies and also with real data example, where the emphasis is on testing mean equality. The simulation results show that the proposed ELR test can successfully control type I error in most scenarios, and that it is not sensitive to unequal sample sizes. Compared with other existing tests for mean equality, the ELR test shows a clear advantage in terms of power when the correctly specified basis function in the DRM is the logarithm function. These conclusions are further supported by a real data analysis.

In Part II of this thesis, we answer the scientific questions Q3–Q5 by showing an understanding of the behaviour of bootstrap percentile confidence intervals in the constrained inference problems. In Chapter 5, the bootstrap procedure and bootstrap percentile confidence interval are formally introduced.

In Chapter 6, we begin by investigating a simple normal example in which the mean parameter is constrained to be nonnegative. In this example, we construct a confidence interval for the mean by using the standard bootstrap percentile method. We quantify its exact coverage probability and observe that it can over-cover the true parameter when it is on, or close to, the boundary. We next quantify the exact coverage probabilities of the bootstrap percentile confidence intervals for two ordered normal means and their difference. The results show that these intervals can under- and over-cover the true mean parameters, but always over-cover the true difference of the two means, when the true mean difference is on, or close to, the boundary.

As we have seen from the motivating examples, the constrained inference problems may be associated with a wide range of data distributions. Therefore, we extend our exact finite sample results for the one- and two-sample normal distributions to the class of distributions in the natural exponential family.

We note that the non-standard coverage phenomena discussed above only occur when the true constrained parameters are on, or close to, the boundary. Here the magnitude of closeness crucially depends on sample size. This motivates us to propose using a local asymptotic framework to study the behaviour of bootstrap percentile confidence intervals for constrained mean parameters when the distributions are in the natural exponential family. Under this proposed framework, similar phenomena as seen in the normal examples can be observed asymptotically for one- and two-sample constrained problems. Our theoretical findings show that the bootstrap percentile confidence intervals may not always offer correct coverage level, and hence should be used with caution.

Finally, in Chapter 7, the achievements of this thesis are summarized and discussed, and possible directions for future research are highlighted.

Part I

Empirical likelihood and the density ratio model

Chapter 2

Introduction to Part I

2.1 The semiparametric density ratio model

In this chapter, we introduce the definition and show the properties of the semiparametric multiple-sample *density ratio model* (DRM) which will play an important role in Part I of this thesis. Suppose we have $m+1$ independent groups of samples as follows:

$$x_{i1}, \ldots, x_{in_i} \sim G_i(x), \quad i = 0, \ldots, m,$$

where n_i is the ith group's sample size which is a fixed number by design, and the $G_i(\cdot)$'s are cumulative distribution functions with common support which may be continuous or discrete. Let $dG_i(x)$ denote the probability density functions or probability mass functions of $G_i(x)$, for $i = 0, \ldots, m$. The definition of the multiple-sample density ratio model (Anderson, 1979; Qin and Zhang, 1997) is given as follows.

Definition 2.1 (Density ratio model) *The $G_i(x)$'s are said to satisfy the multiple-sample density ratio model if each $dG_i(x)$ satisfies*

$$dG_i(x) = \exp\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x)\} dG_0(x), \quad i = 0, \ldots, m, \tag{2.1}$$

for a pre-specified, non-trivial, basis function $\mathbf{q}(x)$ of dimension d, and unknown parameters α_i and $\boldsymbol{\beta}_i$. Clearly, $\alpha_0 = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$ for an arbitrarily selected baseline group.

This definition of the DRM does not require a specification on the baseline distribution $G_0(x)$. Hence, the DRM belongs to the class of semiparametric models, in the sense that it lies between the fully parametric and fully nonparametric approaches. It only specifies a parametric form for the log density ratios, where the parameters α_i's and $\boldsymbol{\beta}_i$'s characterize the discrepancy between dG_i and dG_0. Inference based on the DRM defined in (2.1) does not depend on the form of $G_0(x)$ and hence is robust to any assumptions on $G_0(x)$.

We give now a brief review of the connections between the DRM and several classical statistical models.

The DRM and multinomial logistic regression

Logistic regression models are widely used in analyzing data from case-control studies. The earliest roots of the DRM may date back to Anderson (1979) who was modelling multivariate logistic compounds, where a specific form of $\mathbf{q}(x) = x$ in the DRM was used.

The connection between the DRM and logistic regression can be explained by the equivalence between prospective and retrospective sampling schemes (Prentice and Pyke, 1979; Qin and Zhang, 1997). For a prospective sampling scheme, a study is usually designed by following a cohort of subjects over a period of time. Then we model the observed outcome of disease status at the end of the time period. Suppose $D = i$ for $i = 0, \ldots, m$ is the indicator variable of the ith disease group. Conditional on a covariate $X = x$, the classical multinomial logistic regression model is

$$\Pr(D = i|X = x) = \frac{\exp\{\alpha_i^* + \boldsymbol{\beta}_i^\top \mathbf{q}(x)\}}{1 + \sum_{k=1}^m \exp\{\alpha_k^* + \boldsymbol{\beta}_k^\top \mathbf{q}(x)\}}, \quad i = 0, 1, \ldots, m. \tag{2.2}$$

Due to the time and budget constraints, prospective sampling designs are sometime not feasible. Instead, a retrospective sampling scheme is used, where subjects are included in the study conditional on their disease status. Let $\Pr(X = x|D = i) = dG_i(x)$ with $G_i(x) = \Pr(X \le x|D = i)$. If G_i's satisfy the DRM (2.1), then applying Bayes' rule gives

$$\Pr(D = i|X = x) = \frac{\Pr(X = x|D = i)\Pr(D = i)}{\sum_{k=0}^m \Pr(X = x|D = k)\Pr(D = k)} = \frac{\exp\{\alpha_i^* + \boldsymbol{\beta}_i^\top \mathbf{q}(x)\}}{1 + \sum_{k=1}^m \exp\{\alpha_k^* + \boldsymbol{\beta}_k^\top \mathbf{q}(x)\}},$$

where $\alpha_i^* = \alpha_i - \log\{\Pr(D=0)/\Pr(D=i)\}$. Therefore, the DRM (2.1) is connected with the multinomial logistic regression model (2.2).

Note that logistic regression models and their related inference procedures are well established in the literature. The connection between the DRM and multinomial logistic regression provides a great computational advantage when implementing the DRM and related statistical inference procedures. Besides, this connection also provides a justification for borrowing the logistic regression model diagnostic and selection tools to related purposes for the DRM. The details will be covered in later sections.

The DRM and parametric exponential family

Suppose we have $m+1$ random variables, $X_0, X_1, \dots, X_m$, from the same exponential family with probability density function or probability mass function

$$g(x;\boldsymbol{\phi}_i) = h(x)\exp\left\{\mathbf{w}^\top(\boldsymbol{\phi}_i)\mathbf{T}(x) + c(\boldsymbol{\phi}_i)\right\}, \quad i=0,\dots,m,$$

where $\boldsymbol{\phi}_i$ is a parameter vector, and $h(\cdot)$, $\mathbf{w}(\cdot)$, $\mathbf{T}(\cdot)$, and $c(\cdot)$ are known functions, and the common support of X_i should not depend on $\boldsymbol{\phi}_i$'s.

Let $dG_i(x) = g(x;\boldsymbol{\phi}_i)$, then the ratios $dG_i(x)/dG_0(x)$, $i=0,\dots,m$, satisfy the DRM assumption (2.1) with

$$\alpha_i = c(\boldsymbol{\phi}_i) - c(\boldsymbol{\phi}_0), \quad \boldsymbol{\beta}_i = \mathbf{w}(\boldsymbol{\phi}_i) - \mathbf{w}(\boldsymbol{\phi}_0), \quad \text{and} \quad \mathbf{q}(x) = \mathbf{T}(x).$$

In the classical definition of the parametric exponential family, the form of $h(x)$ must be fully specified. A fundamental innovation in a semiparametric DRM is that the baseline density $dG_0(x)$, the counterpart of $h(x)$, is left unspecified as a nonparametric component since it plays no role in the estimation of α_i and $\boldsymbol{\beta}_i$. This adds to a good deal of flexibility to the specification of the DRM. In Table 2.1, we summarize some special and important members included in the rich DRM family.

As we can see from Table 2.1, the density ratio model assumptions are weaker than a fully specified parametric model, since $\mathbf{q}(x)$ is the only component that needs to be specified in the DRM (2.1). For example, the choice of basis function $\mathbf{q}(x) = \log(x)$

Table 2.1: Examples of members in the family of the semiparametric density ratio model.

Parametric distribution G_i	$\mathbf{q}(x)$	Description
Normal (μ_i, σ_i^2)	$\left(x, x^2\right)^\top$	μ_i: mean; σ_i^2: variance
Binomial (n, p_i)	x	p_i: success probability; n: known # of trails
Poisson (λ_i)	x	λ_i: rate
Negative Binomial (r, p_i)	x	p_i: success probability; r: known # of failures
Exponential (λ_i)	x	λ_i: rate
Log-normal (a_i, b_i)	$\left\{\log(x), \log^2(x)\right\}^\top$	a_i: mean in log scale; b_i: variance in log scale
Gamma (a_i, b_i)	$\{x, \log(x)\}^\top$	a_i: shape; b_i scale
Beta (a_i, b_i)	$\{\log(x), \log(1-x)\}^\top$	a_i: shape; b_i: scale
Weibull (k, λ_i)	x^k	λ_i: scale; k: known shape

embraces the log-normal distribution of same b_i's, and the gamma distribution of same b_i's. These commonly used examples in Table 2.1 also provide a guideline on choosing some candidate basis functions $\mathbf{q}(x)$ in practice.

More discussions on the connections between the DRM (2.1) and other classical statistical models can be found, for example, in Qin (1998) and Jiang and Tu (2012) for the Cox proportional hazards model (Cox, 1972), and in Qin and Zhang (1997) and Gilbert (2000) for the biased sampling problems.

A note on model space of the DRM

Although the DRM is flexible and has wide connections with classical statistical models, it is worth mentioning some exceptions under the Definition 2.1. First, this definition requires that all the G_i's should have common support. This, for example, prohibits the use of DRM to link the normal distribution with known variance (with support on $\mathbb{R}$) and the exponential distribution (with support on $\mathbb{R}^+$). Second, this definition also excludes the distributions with support depending on unknown parameters, for example, the generalized extreme value distribution and the generalized Pareto distribution. Lastly, the class of finite mixture distributions also does not satisfy the definition of DRM, although it is not unusual in applications.

2.2 Dual empirical likelihood for the DRM

In this section, we review an inference procedure for the DRM which uses the empirical likelihood (EL) method.

The EL method was proposed by Owen (1988, 1990, 1991). The EL is a nonparametric likelihood that has many attractive properties analogous to the parametric likelihood but requires no restrictive distributional assumption. In an important milestone paper, Qin and Lawless (1994) showed that auxiliary information, in the form of a set of general unbiased estimating equations, can be incorporated through the EL to improve the efficiency of estimation and the EL ratio based confidence intervals. The monograph of Owen (2001) serves as a comprehensive reference on empirical likelihood.

With the data structure as described in Section 2.1, if we treat all the distribution functions G_i's as unknown parameters, then the likelihood function is

$$\prod_{i=0}^{m}\prod_{j=1}^{n_i} dG_i(x_{ij}).$$

Under the density ratio model (2.1), the likelihood function becomes

$$\mathcal{L} \;=\; \prod_{i=0}^{m}\prod_{j=1}^{n_i} \exp\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij})\} dG_0(x_{ij}).$$

Following the generic recommendation in Owen (2001), we restrict the form of baseline distribution $G_0(x)$ to be a discretized distribution with support being the union of the combined observations of total sample size $n = \sum_{i=0}^{m} n_i$. Specifically,

$$G_0(x) = \sum_{i=0}^{m}\sum_{j=1}^{n_i} p_{ij} I(x_{ij} \le x).$$

Therefore, $p_{ij} = dG_0(x_{ij}) = G_0(x_{ij}) - G_0(x_{ij}-)$ is the probability mass assigned to x_{ij}.

For a compact presentation, let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$, and $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$. Then the empirical log-likelihood function of $(\boldsymbol{\theta}, G_0)$ can be written as

$$\tilde{\ell}(\boldsymbol{\theta}, G_0) \;=\; \log(\mathcal{L}) = \sum_{i=0}^{m}\sum_{j=1}^{n_i} \left\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij}) + \log(p_{ij})\right\},$$

where we have the following natural constraints:

$$p_{ij} > 0, \qquad \sum_{i=0}^{m} \sum_{j=1}^{n_i} p_{ij} \exp\{\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{q}(x_{ij})\} = 1, \quad k = 0, \ldots, m. \tag{2.3}$$

Hence, the maximum EL estimator $(\hat{\boldsymbol{\theta}}, \hat{G}_0)$ of $(\boldsymbol{\theta}, G_0)$ is defined as the solutions that maximize $\tilde{\ell}(\boldsymbol{\theta}, G_0)$ over the constrained space (2.3).

For inference about $\hat{\boldsymbol{\theta}}$, the profile empirical log-likelihood function would be much more convenient to work with. We follow similar procedures to those in Qin and Lawless (1994), using Lagrange multipliers to profile out the infinite dimensional parameter G_0.

We first set up the Lagrangian function. For any given $\boldsymbol{\theta}$,

$$\Lambda(G_0, \boldsymbol{\lambda}) = \tilde{\ell}(\boldsymbol{\theta}, G_0) + n \sum_{k=0}^{m} \lambda_k \left[1 - \sum_{i=0}^{m} \sum_{j=1}^{n_i} p_{ij} \exp\{\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{q}(x_{ij})\} \right],$$

where $\boldsymbol{\lambda} = (\lambda_0, ..., \lambda_m)^\top$ are Lagrangian multipliers. The point $\{p_{i1}, \ldots, p_{in_i} : i = 0, \ldots, m\}$ that maximize $\tilde{\ell}(\boldsymbol{\theta}, G_0)$ must also be a stationary point of Λ satisfying

$$\frac{\partial \Lambda(G_0, \boldsymbol{\lambda})}{\partial p_{ij}} = 0 \quad \text{and} \quad \frac{\partial \Lambda(G_0, \boldsymbol{\lambda})}{\partial \lambda_i} = 0. \tag{2.4}$$

We note that

$$\begin{aligned} 0 &= \sum_{i=0}^{m} \sum_{j=1}^{n_i} p_{ij} \frac{\partial \Lambda(G_0, \boldsymbol{\lambda})}{\partial p_{ij}} \\ &= \sum_{i=0}^{m} \sum_{j=1}^{n_i} p_{ij} \left\{ p_{ij}^{-1} - n \sum_{k=0}^{m} \lambda_k \exp\{\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{q}(x_{ij}) \right\} \\ &= n - n \sum_{k=0}^{m} \lambda_k \sum_{i=0}^{m} \sum_{j=1}^{n_i} p_{ij} \exp\left\{\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{q}(x_{ij})\right\} \\ &= n - n \sum_{k=0}^{m} \lambda_k \end{aligned} \tag{2.5}$$

where the last equation follows from the constraint (2.3).

By solving (2.4) together with (2.5), the function $\tilde{\ell}(\boldsymbol{\theta}, G_0)$ achieves its maximum when $\lambda_0 = 1 - \sum_{i=1}^m \lambda_i$ and

$$p_{ij} = \frac{1}{n}\frac{1}{1 + \sum_{k=1}^m \lambda_k[\exp\{\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{q}(x_{ij})\} - 1]}, \tag{2.6}$$

where $(\lambda_1, ..., \lambda_m)$ solves

$$\sum_{i=0}^m \sum_{j=1}^{n_i} \frac{\exp\{\alpha_r + \boldsymbol{\beta}_r^\top \mathbf{q}(x_{ij})\} - 1}{1 + \sum_{k=1}^m \lambda_k[\exp\{\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{q}(x_{ij})\} - 1]} = 0, \quad \text{for } r = 1, ..., m. \tag{2.7}$$

Using (2.6) to profile out the p_{ij}'s, the profile empirical log-likelihood function of $\boldsymbol{\theta}$, up to a constant, not depending on the unknown parameters, is

$$\tilde{\ell}^*(\boldsymbol{\theta}) = -\sum_{i=0}^m \sum_{j=1}^{n_i} \log\left\{1 + \sum_{t=1}^m \lambda_t[\exp\{\alpha_t + \boldsymbol{\beta}_t^\top \mathbf{q}(x_{ij})\} - 1]\right\} + \sum_{i=1}^m \sum_{j=1}^{n_i}\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij})\}.$$

Therefore, the maximum EL estimator of $\boldsymbol{\theta}$ can be obtained by $\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}\tilde{\ell}^*(\boldsymbol{\theta})$.

Unfortunately, there are at least two potential issues for using $\tilde{\ell}^*(\boldsymbol{\theta})$ to make inference on $\boldsymbol{\theta}$. First, in the computation of $\hat{\boldsymbol{\theta}}$, we need to maximize $\tilde{\ell}^*(\boldsymbol{\theta})$ numerically. This may not be an easy task since there are no analytical solutions for (2.7) in the definition of $\tilde{\ell}^*(\boldsymbol{\theta})$. Second, under the constraints (2.3), $\boldsymbol{\beta} = \mathbf{0}$ implies $\boldsymbol{\alpha} = \mathbf{0}$ and hence the function $\tilde{\ell}^*(\boldsymbol{\theta})$ may not be identified in a neighbourhood of $\boldsymbol{\theta} = \mathbf{0}$. Therefore, $\boldsymbol{\theta} = \mathbf{0}$ is no longer an interior point of the parameter space which violates the standard regularity conditions in traditional likelihood inference.

Keziou and Leoni-Aubin (2008) and Cai et al. (2016) argued that the above two issues of $\tilde{\ell}^*(\boldsymbol{\theta})$ can be resolved by optimising the following *dual* empirical log-likelihood function:

$$\ell(\boldsymbol{\theta}) = -\sum_{i=0}^m \sum_{j=1}^{n_i} \log\left[\rho_0 + \sum_{k=1}^m \rho_k \exp\{\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{q}(x_{ij})\}\right] + \sum_{i=1}^m \sum_{j=1}^{n_i}\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij})\}, \tag{2.8}$$

where $\rho_k = n_k/n$ for $k = 0, \ldots, m$.

We summarize the favourable properties of $\ell(\boldsymbol{\theta})$ by restating some key results of Keziou and Leoni-Aubin (2008) for the two-sample case, and Cai et al. (2016) for the multiple-sample case, in the following proposition.

Proposition 2.1 (Properties of dual EL) *Under some mild regularity conditions, for example, those in Cai et al. (2016), then for* $\ell(\boldsymbol{\theta})$ *defined in (2.8) and* $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \tilde{\ell}^*(\boldsymbol{\theta})$*, we have*

(a) $\ell(\boldsymbol{\theta})$ *is a concave function in* $\boldsymbol{\theta}$*, and it is strictly concave when* $\sum_{i=1}^m \sum_{j=1}^{n_i} \{1, \mathbf{q}^\top(x_{ij})\}^\top \{1, \mathbf{q}^\top(x_{ij})\}$ *is positive definite.*

(b) $\ell(\boldsymbol{\theta})$ *attains its maximum in a* $n^{1/3}$ *neighbourhood of the true parameter* $\boldsymbol{\theta}^*$ *as* $n \to \infty$.

(c) $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta})$ *and* $\ell(\hat{\boldsymbol{\theta}}) = \tilde{\ell}^*(\hat{\boldsymbol{\theta}})$.

Proof. We only prove Part (c). The proofs of Part (a) and Part (b) can be found by applying Theorem 3.1 in Keziou and Leoni-Aubin (2008) for the two-sample case, and Lemma 2.3 and Lemma 2.4 in Cai (2014) for the multiple-sample case.

Recall that $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\{\tilde{\ell}^*(\boldsymbol{\theta})\}$. Let $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_m)^\top$ be the solution of (2.7) with $\boldsymbol{\theta}$ being replaced by $\hat{\boldsymbol{\theta}}$. That is, $\hat{\boldsymbol{\lambda}}$ satisfies

$$\sum_{i=0}^m \sum_{j=1}^{n_i} \frac{\exp\{\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^\top \mathbf{q}(x_{ij})\} - 1}{1 + \sum_{t=1}^m \hat{\lambda}_t[\exp\{\hat{\alpha}_t + \hat{\boldsymbol{\beta}}_t^\top \mathbf{q}(x_{ij})\} - 1]} = 0, \quad r = 1, \ldots, m. \tag{2.9}$$

Plugging $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\lambda}}$ into (2.6), we denote for $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$

$$\hat{p}_{ij} = \frac{1}{n} \frac{1}{1 + \sum_{t=1}^m \hat{\lambda}_t[\exp\{\hat{\alpha}_t + \hat{\boldsymbol{\beta}}_t^\top \mathbf{q}(x_{ij})\} - 1]}. \tag{2.10}$$

In the first step, we find the explicit form of $\hat{\boldsymbol{\lambda}}$. For the convenience of our presentation, we let

$$h(\boldsymbol{\theta}, \boldsymbol{\lambda}) = -\sum_{i=0}^m \sum_{j=1}^{n_i} \log\left\{1 + \sum_{t=1}^m \lambda_t[\exp\{\alpha_t + \boldsymbol{\beta}_t^\top \mathbf{q}(x_{ij})\} - 1]\right\} + \sum_{i=1}^m \sum_{j=1}^{n_i} \{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij})\}.$$

Then $\tilde{\ell}^*(\boldsymbol{\theta}) = h(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with $\boldsymbol{\lambda}$ being the solution of (2.7) and $\tilde{\ell}^*(\hat{\boldsymbol{\theta}}) = h(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})$.

Note that when $\boldsymbol{\lambda}$ is the solution of (2.7), λ_i's are actually functions of $\boldsymbol{\theta}$. When $\tilde{\ell}^*(\boldsymbol{\theta})$ is maximized, the following equations should be satisfied, for $k = 1, \ldots, m$

$$0 = \frac{\partial \tilde{\ell}^*(\hat{\boldsymbol{\theta}})}{\partial \alpha_k}$$

$$
\begin{aligned}
&= \frac{\partial h(\hat{\boldsymbol\theta}, \hat{\boldsymbol\lambda})}{\partial \alpha_k} + \sum_{r=1}^{m} \frac{\partial h(\hat{\boldsymbol\theta}, \hat{\boldsymbol\lambda})}{\partial \lambda_r} \frac{\partial \lambda_r}{\alpha_k}\Big|_{\boldsymbol\theta=\hat{\boldsymbol\theta}} \\
&= -\hat\lambda_k \sum_{i=0}^{m}\sum_{j=1}^{n_i} \frac{\exp\{\hat\alpha_k + \hat{\boldsymbol\beta}_k^\top \mathbf{q}(x_{ij})\}}{1+\sum_{t=1}^{m}\hat\lambda_t[\exp\{\hat\alpha_t + \hat{\boldsymbol\beta}_t^\top \mathbf{q}(x_{ij})\}-1]} + n_k \\
&\quad -\sum_{r=1}^{m}\sum_{i=0}^{m}\sum_{j=1}^{n_i} \frac{\exp\{\hat\alpha_k + \hat{\boldsymbol\beta}_k^\top \mathbf{q}(x_{ij})\}-1}{1+\sum_{t=1}^{m}\hat\lambda_t[\exp\{\hat\alpha_t + \hat{\boldsymbol\beta}_t^\top \mathbf{q}(x_{ij})\}-1]} \cdot \frac{\partial \lambda_r}{\alpha_k}\Big|_{\boldsymbol\theta=\hat{\boldsymbol\theta}} \\
&= -n\hat\lambda_k \sum_{i=0}^{m}\sum_{j=1}^{n_i} \exp\{\hat\alpha_k + \hat{\boldsymbol\beta}_k^\top \mathbf{q}(x_{ij})\}\hat p_{ij} + n_k \\
&= n_k - n\hat\lambda_k,
\end{aligned}
$$

where the third equation is by (2.9) and (2.10), and the last equation follows from the constraint (2.3). In summary, when $\tilde\ell^*(\boldsymbol\theta)$ is maximized at $\boldsymbol\theta = \hat{\boldsymbol\theta}$, the Lagrange multipliers have the explicit solutions

$$\hat\lambda_k = \frac{n_k}{n} = \rho_k, \quad k = 1, \dots, m.$$

Hence $\ell(\boldsymbol\theta) = h(\boldsymbol\theta, \hat{\boldsymbol\lambda})$.

In the second step, we argue that

$$\frac{\partial \ell(\hat{\boldsymbol\theta})}{\partial \boldsymbol\theta} = \frac{\partial h(\hat{\boldsymbol\theta}, \hat{\boldsymbol\lambda})}{\partial \boldsymbol\theta} = \mathbf{0}. \tag{2.11}$$

It can be easily verified that, for $k = 1, \dots, m$

$$\frac{\partial h(\hat{\boldsymbol\theta}, \hat{\boldsymbol\lambda})}{\partial \alpha_k} = -n\hat\lambda_k \sum_{i=0}^{m}\sum_{j=1}^{n_i} \exp\{\hat\alpha_k + \hat{\boldsymbol\beta}_k^\top \mathbf{q}(x_{ij})\}\hat p_{ij} + n_k = n_k - n\hat\lambda_k = 0. \tag{2.12}$$

For $\partial\ell(\hat{\boldsymbol\theta})/\partial\boldsymbol\beta_k$, we notice that, for $k = 1, \dots, m$

$$
\begin{aligned}
\mathbf{0} &= \frac{\partial \tilde\ell^*(\hat{\boldsymbol\theta})}{\partial \boldsymbol\beta_k} \\
&= \frac{\partial h(\hat{\boldsymbol\theta}, \hat{\boldsymbol\lambda})}{\partial \boldsymbol\beta_k} + \sum_{r=1}^{m} \frac{\partial h(\hat{\boldsymbol\theta}, \hat\lambda)}{\partial \lambda_r} \frac{\partial \lambda_r}{\partial \boldsymbol\beta_k}\Big|_{\boldsymbol\theta=\hat{\boldsymbol\theta}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\partial h(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})}{\partial \boldsymbol{\beta}_k} - \sum_{r=1}^{m} \sum_{i=0}^{m} \sum_{j=1}^{n_i} \frac{\exp\{\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^\top \mathbf{q}(x_{ij})\} - 1}{1 + \sum_{t=1}^{m} \hat{\lambda}_t [\exp\{\hat{\alpha}_t + \hat{\boldsymbol{\beta}}_t^\top \mathbf{q}(x_{ij})\} - 1]} \cdot \frac{\partial \lambda_r}{\boldsymbol{\beta}_k} \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \\
&= \frac{\partial h(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})}{\partial \boldsymbol{\beta}_k}, \qquad (2.13)
\end{aligned}
$$

where in the last line, we have used (2.9). Hence, (2.12) and (2.13) together imply (2.11). Hence $\hat{\boldsymbol{\theta}}$ is a stationary point of $\ell(\boldsymbol{\theta})$. Due to the concavity of $\ell(\boldsymbol{\theta})$, we have

$$
\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}).
$$

Recall that $\tilde{\ell}^*(\hat{\boldsymbol{\theta}}) = h(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})$ and $\ell(\boldsymbol{\theta}) = h(\boldsymbol{\theta}, \hat{\boldsymbol{\lambda}})$. Then

$$
\tilde{\ell}^*(\hat{\boldsymbol{\theta}}) = h(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) = \ell(\hat{\boldsymbol{\theta}}).
$$

This completes the proof. □

We note that property (c) in Proposition 2.1 states that the maximum point and maximum value of $\tilde{\ell}^*(\boldsymbol{\theta})$ are identical to those of the dual empirical log-likelihood function $\ell(\boldsymbol{\theta})$. With property (c), the properties (a) and (b) guarantee that a unique $\hat{\boldsymbol{\theta}}$ exists and it is an interior point of the parameter space with probability one as $n \to \infty$. These properties are important for the following discussion on the numerical computation of $\hat{\boldsymbol{\theta}}$ and $\ell(\hat{\boldsymbol{\theta}})$.

Subsequently, plugging $\hat{\boldsymbol{\theta}}$ and $\hat{\lambda}_k = n_k/n$ into (2.6), we obtain

$$
\hat{p}_{ij} = \frac{1}{n} \frac{1}{1 + \sum_{k=1}^{m} \frac{n_k}{n} [\exp\{\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^\top \mathbf{q}(x_{ij})\} - 1]}.
$$

Finally, we can estimate the distribution function $G_k(x)$, for $k = 1, \ldots, m$, by

$$
\hat{G}_k(x) = \frac{1}{n} \sum_{i=0}^{m} \sum_{j=1}^{n_i} \exp\{\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^\top \mathbf{q}(x_{ij})\} \hat{p}_{ij} I(x_{ij} \le x). \qquad (2.14)
$$

Asymptotic properties

Under some mild regularity conditions, the $\sqrt{n}$-consistency of $\hat{\boldsymbol{\theta}}$ and the asymptotic normality of $\hat{\boldsymbol{\theta}}$ have been established in many situations; see for example, Qin (1998) and Keziou and Leoni-Aubin (2008) for the two-sample case, and Fokianos et al. (2001) and Chen and Liu (2013) for the multiple-sample case. Based on these results, Qin and Zhang (1997) and Zhang (2002) further showed that the semiparametric estimator $\{\hat{G}_0(x), \ldots, \hat{G}_m(x)\}$ in (2.14) converges weakly to a multivariate Gaussian process.

A note on computation

The solutions of $\hat{\boldsymbol{\theta}}$ and $\hat{p}_{ij}$ can not be found explicitly in general. In order to compute $\hat{\boldsymbol{\theta}}$, numerical optimization algorithms are required. To do this we can use the equivalence between the DRM (2.1) and the multinomial logistic regression model (2.2). For a given $\mathbf{q}(x)$, the log-likelihood function of the multinomial logistic regression model in (2.2) is

$$\check{\ell}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}) = -\sum_{i=0}^{m}\sum_{j=1}^{n_i} \log\left[1 + \sum_{k=1}^{m} \exp\{\alpha_k^* + \boldsymbol{\beta}_k^\top \mathbf{q}(x_{ij})\}\right] + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\{\alpha_i^* + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij})\},$$

where $\boldsymbol{\alpha}_i^* = (\alpha_1, \ldots, \alpha_m^*)^\top$ is the vector of intercepts and $\boldsymbol{\beta}_i$'s are vectors of regression parameters interpreted as the log odds ratios associated with the ith group. Compared with the dual likelihood $\ell(\boldsymbol{\theta})$ in (2.8), it is easy to find that

$$\ell(\boldsymbol{\theta}) = \check{\ell}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}) - \sum_{i=0}^{m} n_i \log(\rho_i),$$

and $\alpha_i = \alpha_i^* + \log(\rho_0/\rho_i)$ for $i = 1\ldots, m$. This fact makes the computation of $\hat{\boldsymbol{\theta}}$ and $\ell(\hat{\boldsymbol{\theta}})$ very straightforward using logistic regression routines, for example, via the `R` function `multinom` available in the package "`nnet`" (R Development Core Team, 2014). Alternatively, one can use the `R` package "`drmdel`" (Cai, 2015) for the dual EL inference under the DRM; this includes solving more general computational problems in Chen and Liu (2013) and Cai et al. (2016).

2.3 The DRM diagnostic and selection

Despite the flexibility of the DRM, (2.1), misspecifying the DRM can still have an impact on the inference results (Fokianos and Kaimi, 2006). To assess the DRM assumption

(2.1), a Kolmogorov-Smirnov-type goodness-of-fit test of Zhang (2002), for multinomial logistic models based on case-control data, may be used. However, this may only be helpful in identifying several competing models with suitable basis functions $\mathbf{q}(x)$. In practice, Fokianos (2007) suggested selecting a suitable basis function $\mathbf{q}(x)$ in a DRM among several competing models using Akaike's information criterion (AIC) (Akaike, 1973) which is defined as

$$\text{AIC} = -2\ell(\hat{\boldsymbol{\theta}}) + 2(md + m).$$

Here $\ell(\hat{\boldsymbol{\theta}})$ is the dual EL and $md + m$ are the number of parameters in $\boldsymbol{\theta}$. The performance of AIC was evaluated in Fokianos (2007), and it was shown to be robust for moderate sample sizes.

2.4 Recent developments on the DRM

The idea of the DRM as a powerful semiparametric tool was not being fully recognized until the establishment of the theoretical foundation of empirical likelihood by Owen (1988). Later, the idea of the DRM was popularized by Qin and Zhang (1997) and Qin (1998). In particular, Qin and Zhang (1997) demonstrated the equivalence between the DRM and logistic regression models, and further developed procedures to assess the goodness-of-fit of the logistic regression models based on case-control data. Qin (1998) formally established the large sample properties of the maximum EL estimators of the DRM parameters in the two-sample case.

We now selectively review some fundamental work for EL inference under the DRM. For the multiple-sample estimation problems, Fokianos (2004) and Aubin and Leoni-Aubin (2008) considered density estimation under the DRM. Diao et al. (2012) studies the estimation efficiency of parameters in the DRM. Chen and Liu (2013) discussed the quantile and quantile function estimation problem using the DRM. For the multiple-sample hypothesis testing problems, Fokianos et al. (2001) constructed a Wald-type test of the linear hypothesis about the DRM parameters in one-way layout. Zhang (2002) considered testing the validity of the generalized logistic model with multiple categories based on case-control data using the connection with the DRM. Cai (2014) constructed an EL ratio statistic for testing a general composite hypothesis about the DRM parameters. Wang (2014) proposed a Wald-type statistic for testing the means of several populations under the DRM.

There has been a large literature that investigate the merit of the DRM in a wide range of important statistical problems. For examples, Qin (1999), Zou et al. (2002) and Li et al. (2016) applied the DRM to semiparametric mixture models. Guan (2004) and Hu et al. (2008) extended the use of the DRM for the change point problems. Fokianos and Savvides (2008) and Kedem et al. (2008) explored the use of DRM in the context of time series data. Shen et al. (2012) and Chan (2013) implemented the DRM in the analysis of survival data. Jiang and Tu (2012) and Jiang et al. (2016) studied making inference under the DRM with censored data. Qin and Zhang (2003), Wan and Zhang (2007) and Wan and Zhang (2013) examined using the DRM for optimally estimating receiver operating characteristic curves. Davidov et al. (2010) and Davidov et al. (2014) generalized the DRM to inference problems under order restriction.

Other significant recent advances include the work of Luo and Tsai (2012) and Huang and Rathouz (2012) who introduced proportional likelihood ratio models to regress the DRM parameter $\boldsymbol{\beta}$ on a set of covariates. Liu et al. (2014) built up a local empirical likelihood framework to allow for varying coefficients in a DRM. The paper of de Carvalho and Davison (2014) investigated modelling several multivariate extremal distributions using the DRM. Chen et al. (2014) and Ning and Chen (2015) suggested a new pseudo-likelihood framework for inference problems under the DRM. Qin et al. (2015) used the DRM idea to integrate valuable auxiliary information from external large data sources for case-control studies. Chen et al. (2016) relaxed the log-linear form of the DRM and proposed a monotonic DRM to further improve the robustness of the inference procedures.

To the best of our knowledge, the semiparametric DRM has not been used in modelling multiple samples with excess zero observations. In Chapters 3 and 4, we will model and develop inference procedures for multiple samples with excess zero observations under such a semiparametric framework.

Chapter 3

Testing homogeneity for multiple nonnegative distributions with excess zero observations

3.1 Introduction

This chapter is devoted to answering the scientific question Q1 outlined in Section 1.1 of Chapter 1. That is, given multiple groups with an excess of zero observations, we aim to test the homogeneity of their distributions. This hypothesis testing problem has been considered as one of the fundamental problems in the literature; see for example, Lachenbruch (1976, 2001, 2002); Tse et al. (2009); Bedrick and Hossain (2013); Johnson et al. (2015). Specifically, suppose we have $m+1$ independent groups of samples as follows:

$$x_{i1}, \ldots, x_{in_i} \sim F_i(x) = \nu_i I(x=0) + (1-\nu_i) I(x>0) G_i(x), \quad i = 0, \ldots, m, \tag{3.1}$$

where n_i is the ith group's sample size, $I(\cdot)$ is an indicator function and the $G_i(\cdot)$'s are cumulative distribution functions with common support which may be continuous or discrete. In this chapter, we concentrate on continuous distributions $G_i(\cdot)$'s whose support consists of nonnegative real numbers; but we proposed, in Section 7.1, ways that the method can be applied to discrete distributions. For random samples with excess zero observations, the zero outcomes, in fact, contain valuable information and thus should not be simply discarded. The above formulation, (3.1), which is a non-standard mixture model of a point mass distribution at zero and a continuous nonnegative component, is an intuitive way to account for the unique features of such data. Our interest is to test whether the $m+1$ mixture distributions F_i's are homogenous. In addition, one may also be interested in testing the equality of their means. This topic will be further covered in Chapter 4.

In the literature, two-part tests have been widely used to compare groups of samples from the non-standard mixture structure in (3.1). Lachenbruch (2001, 2002) comprehensively studied two-part tests for two such populations. A two-part test is a two degrees of freedom test based on the sum of a test statistic for the equality of the proportions of zero counts and a conditional χ^2-test statistic for the positive part. The test for the latter part may be a nonparametric Wilcoxon-Mann-Whitney rank sum test or a two-sample t-test. If more than two populations are under consideration, we can replace these tests with a Kruskal-Wallis test or an ANOVA F-test, respectively. On the other hand, the parametric likelihood ratio test can also be used for the second part after assuming a parametric form on the nonzero data, such as a log-normal distribution or a gamma distribution. The two-part tests and their extensions have been successfully implemented in various applications; see for example, Bascoul-Mollevi et al. (2005), Taylor and Pollard (2009), and Wagner et al. (2011). Further ideas and comparisons of some existing one- and two-part procedures may be found in Delucchi and Bostrom (2004), Follmann et al. (2009), Hallstrom (2010), and Hu and Proschan (2015).

In numerical studies (see Sections 3.3 and 3.6), we show that the existing two-part tests are either inefficient when no parametric assumptions are made for the nonnegative components or are not robust when the parametric models are assumed. In many applications, multiple populations may naturally share some common characteristics. It is therefore desirable to borrow efficiency across similar populations to improve testing power. At the same time, we also hope that a test is robust to deviations from the model assumptions. The semiparametric density ratio model defined in Chapter 2 is a natural tool to use here. We propose to compare the distributions of the continuous nonnegative components in (3.1), by the DRM to exploit information from all available samples. Let $dG_i(x)$ denote the density of $G_i(x)$ for $i = 0, \dots, m$. We link $G_i(x)$ through the DRM such that

$$dG_i(x) = \exp\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x)\} dG_0(x), \quad i = 0, \dots, m. \tag{3.2}$$

for a non-trivial pre-specified basis function $\mathbf{q}(x)$ of dimension d, and corresponding unknown parameters α_i and $\boldsymbol{\beta}_i$ ($\alpha_0 = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$). Without specifying the baseline density $dG_0(x)$, we propose a test based on the DRM that does not depend on the form of $G_0(x)$ and hence is robust to the assumption on $G_0(x)$.

Under this semiparametric setup, we propose an empirical likelihood ratio (ELR) test for homogeneity under (3.1) and (3.2). We show that the proposed ELR test is also a two-part test: the first part tests the equality of zero proportions ν_i's and the second part tests the homogeneity in the continuous components of the model. We show that the asymptotic null distribution of the ELR is χ^2-type as the total sample size goes to infinity. We also explore using a nonparametric bootstrap procedure to calibrate the distributions of the proposed test statistic in finite-sample situations. This bootstrap procedure is shown to approximate the null distribution of the ELR test statistic under both the null and alternative hypotheses. Software implementing the bootstrap ELR test has been developed in the `R` language (R Development Core Team, 2014) and is supplemented in the Appendix A.1 at the end of this thesis.

We note that developing the limiting distribution of the ELR is technically challenging. First, in the second part of the ELR, the number of observations, i.e., the number of positive observations in each group, is a random number, thus differs from the work of Fokianos et al. (2001), Zhang (2002), Cai et al. (2016) and this creates new challenges. In particular we have to deal with random sums of independent and identically distributed random variables. Second, the two parts of the ELR both have χ^2-type null limiting distributions. Hence we need to show asymptotic independence so that their summation still has a χ^2-type null limiting distribution. We comment that investigating the asymptotic properties of the bootstrap procedure under both the null and alternative hypotheses is also technically challenging. Existing results may not be directly applied. We refer to Janssen and Pauls (2003) for more discussion.

We further note that under the null hypothesis of homogeneity, the DRM in (3.2) is automatically satisfied regardless of the choice of basis function $\mathbf{q}(x)$. This is because the null hypothesis corresponds to a reduced form of the DRM with all $\alpha_i = 0$ and $\boldsymbol{\beta}_i = \mathbf{0}$. This property is attractive because it ensures that the asymptotic size of the test can always be controlled at its nominal level even if the form of basis function $\mathbf{q}(x)$ is misspecified. Hence the size of the proposed ELR test is robust to both the choice of $\mathbf{q}(x)$ and the assumption on $G_0(x)$.

The rest of this chapter is organized as follows. In Section 3.2, we present the ELR test for homogeneity under the DRM, investigate its asymptotic properties, and propose a nonparametric bootstrap procedure to calibrate its finite-sample distribution. We further study the theoretical properties of the bootstrap procedure under both the null and alternative hypotheses. In Section 3.3, we report some simulation results. A real example is given in Section 3.4. For the convenience of presentation, all proofs and complete simulation results are in Sections 3.5 and 3.6, respectively.

3.2 Main results

In this section, we present an empirical likelihood method for testing homogeneity of distributions under (3.1) and (3.2). For a compact presentation, let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$, and $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$. Under the condition (3.2), testing the null hypothesis, that all the $m+1$ groups of samples come from the same population, i.e. $F_0 = \cdots = F_m$, is equivalent to testing

$$H_0 : \nu_0 = \cdots = \nu_m \quad \text{and} \quad \boldsymbol{\theta} = \mathbf{0}. \tag{3.3}$$

3.2.1 Empirical likelihood method

Let n_{i0} and n_{i1} be the (random) numbers of zeros and positive observations for the ith group, respectively, for $i = 0, \ldots, m$. Without loss of generality, we use $x_{i1}, \ldots, x_{in_{i1}}$ to denote the positive observations for the ith group.

With the given multiple groups of observations from (3.1), the likelihood function of the parameters ν_i's and G_i's is

$$\prod_{i=0}^{m}\left\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\prod_{j=1}^{n_{i1}}dG_i(x_{ij})\right\}.$$

Under the density ratio model (3.2), the likelihood function becomes

$$\mathcal{L}=\prod_{i=0}^{m}\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\prod_{i=0}^{m}\prod_{j=1}^{n_{i1}}\exp\{\alpha_i+\boldsymbol{\beta}_i^\top\mathbf{q}(x_{ij})\}dG_0(x_{ij}).$$

Along the lines of empirical likelihood (see Section 2.2), we restrict the form of baseline distribution G_0 to be

$$G_0(x)=\sum_{i=0}^{m}\sum_{j=1}^{n_{i1}}p_{ij}I(x_{ij}\le x).$$

Let $\boldsymbol{\nu}=(\nu_0,\dots,\nu_m)^\top$, then the empirical log-likelihood function of $(\boldsymbol{\nu},\boldsymbol{\theta},G_0)$ can be written as

$$\log(\mathcal{L})=\ell_0(\boldsymbol{\nu})+\tilde{\ell}_1(\boldsymbol{\theta},G_0),$$

where

$$\ell_0(\boldsymbol{\nu})=\sum_{i=0}^{m}\log\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\},\quad \tilde{\ell}_1(\boldsymbol{\theta},G_0)=\sum_{i=0}^{m}\sum_{j=1}^{n_{i1}}\left\{\alpha_i+\boldsymbol{\beta}_i^\top\mathbf{q}(x_{ij})+\log(p_{ij})\right\}.$$

Here $\ell_0(\boldsymbol{\nu})$ is the binomial log-likelihood corresponding to the number of zero observations and $\tilde{\ell}_1(\boldsymbol{\theta},G_0)$ is the empirical log-likelihood associated with the positive observations.

We have the following natural constraints:

$$p_{ij}>0,\quad \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}}p_{ij}\exp\{\alpha_r+\boldsymbol{\beta}_r^\top\mathbf{q}(x_{ij})\}=1,\quad r=0,\dots,m. \tag{3.4}$$

Recall the derivation of dual empirical log-likelihood function in Section 2.2. By replacing n_i in (2.8) with n_{i1}, we have the following dual empirical log-likelihood function of $\boldsymbol{\theta}$:

$$\ell_1(\boldsymbol{\theta})=-\sum_{i=0}^{m}\sum_{j=1}^{n_{i1}}\log\left[\rho_0+\sum_{r=1}^{m}\rho_r\exp\{\alpha_r+\boldsymbol{\beta}_r^\top\mathbf{q}(x_{ij})\}\right]+\sum_{i=1}^{m}\sum_{j=1}^{n_{i1}}\{\alpha_i+\boldsymbol{\beta}_i^\top\mathbf{q}(x_{ij})\}, \tag{3.5}$$

where $\rho_r = n_{r1}/n_{\cdot 1}$ for $r = 0, \ldots, m$ with $n_{\cdot 1} = \sum_{i=0}^{m} n_{i1}$. Note that this dual empirical log-likelihood $\ell_1(\boldsymbol{\theta})$ retains the nice properties given in Proposition 2.1.

Let $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}) = \text{argmax}_{\boldsymbol{\nu},\boldsymbol{\theta}}\{\ell_0(\boldsymbol{\nu}) + \ell_1(\boldsymbol{\theta})\}$ be the maximum EL estimate of $(\boldsymbol{\nu}, \boldsymbol{\theta})$. Clearly, $\hat{\boldsymbol{\nu}}$ and $\hat{\boldsymbol{\theta}}$ maximize $\ell_0(\boldsymbol{\nu})$ and $\ell_1(\boldsymbol{\theta})$, respectively. Therefore, the ELR test statistic for testing H_0 is defined as

$$R_n = 2\left\{\ell_0(\hat{\boldsymbol{\nu}}) - \max_{H_0} \ell_0(\boldsymbol{\nu})\right\} + 2\ell_1(\hat{\boldsymbol{\theta}}),$$

where, in the last step, we use the fact that $\ell_1(\boldsymbol{\theta}) = \ell_1(\mathbf{0}) = 0$ under H_0. In R_n, the subscript n denotes the total sample size, i.e., $n = \sum_{i=0}^{m} n_i$.

3.2.2 Large sample property

In this subsection, we study the asymptotic distribution of the ELR test statistic R_n under H_0.

Suppose that the true value of $\boldsymbol{\nu}$ is $\boldsymbol{\nu}^* = (\nu_0^*, \ldots, \nu_0^*)^\top$ under H_0. The true value of $\boldsymbol{\theta}$ is $\mathbf{0}$ under H_0. Then the asymptotic result in this section relies on the following regularity conditions.

C1. $\nu_0^* \in (0, 1)$.

C2. $\lim_{\min\{n_0,\ldots,n_m\}\to\infty} n_i/n \to \rho_i^*$, where $\rho_i^* \in (0, 1)$ for $i = 0, \ldots, m$.

C3. $\int \left(1, \mathbf{q}^\top(x)I(x > 0)\right)^\top \left(1, \mathbf{q}^\top(x)I(x > 0)\right) dF_0(x)$ exists and is positive definite.

C4. $\int \exp\{\boldsymbol{\beta}_i^\top \mathbf{q}(x)\}I(x > 0)dF_0(x) < \infty$ in a neighbourhood of $\boldsymbol{\beta}_i = \mathbf{0}$ for $i = 0, \ldots, m$.

Condition C1 states that the parameter $\boldsymbol{\nu}^*$ is an interior point of the parameter space of $\boldsymbol{\nu}$ such that $\ell_0(\boldsymbol{\nu})$ has regular properties. Condition C2 assumes that the ratio of each group sample size to n converges to a constant as $\min\{n_0, \ldots, n_m\} \to \infty$. For simplicity and convenience of presentation, we write $\rho_i^* = n_i/n$ and assume that it is a constant. This does not affect our technical development. Under C1 and C2, there is no need to distinguish the stochastic orders with respect to n or n_i. Condition C3 is an identifiability condition, and it ensures that the components of $\{1, \mathbf{q}^\top(x)I(x>0)\}$ are linearly independent under $F_0(x)$, and hence $\mathbf{q}(x)$ can not be a constant function. Conditions C3 and C4 guarantee that a second-order approximation of $\ell_1(\boldsymbol{\theta})$ is applicable.

The limiting distribution of R_n is given in the following theorem. Let $\xrightarrow{d}$ denote "convergence in distribution".

Theorem 3.1 *Suppose we have $m+1$ groups of samples from (3.1) and condition (3.2) is satisfied. Assume, also, that the regularity conditions C1–C4 hold. Under the null hypothesis H_0 given in (3.3), we have*

$$R_n \xrightarrow{d} \chi^2_{m(d+1)}$$

as $n \to \infty$, where $\chi^2_{m(d+1)}$ denotes a chi-squared random variable with $m(d+1)$ degrees of freedom.

For the convenience of presentation, we defer the proof of Theorem 3.1 to Section 3.5.1.

We provide a remark on the result in Theorem 3.1 by considering the degrees of freedom in the χ^2-limiting distribution. Let $R_{n,1} = 2\{\ell_0(\hat{\boldsymbol{\nu}}) - \max_{H_0} \ell_0(\boldsymbol{\nu})\}$ and $R_{n,2} = 2\ell_1(\hat{\boldsymbol{\theta}})$. Then the ELR test statistic $R_n = R_{n,1} + R_{n,2}$. Here the first part $R_{n,1}$ tests the equality of zero proportions $\boldsymbol{\nu}$ and is shown to have a χ^2_m null limiting distribution. The second part $R_{n,2}$ tests homogeneity and is shown to have a χ^2_{md} null limiting distribution. For second part we actually have total $dim(\boldsymbol{\theta}) = m+md$ number of parameters, but the number of free parameters is md since $\boldsymbol{\beta} = \mathbf{0}$ automatically implies that $\boldsymbol{\alpha} = \mathbf{0}$. Adding the asymptotically independent $R_{n,1}$ and $R_{n,2}$ together, we get that the null limiting distribution of R_n is $\chi^2_{m(d+1)}$. In addition, some technical difficulties in our proof have been highlighted in Section 3.1.

3.2.3 A bootstrap procedure

As we show through simulations in Section 3.3, the approximation of the limiting distribution to the finite-sample distribution of the ELR is not satisfactory in many situations.

Intuitively we can see two possible causes: (i) $R_{n,1}$ and $R_{n,2}$ are not independent in finite-sample situations; and (ii) even when the total sample size is moderately large, the asymptotic results can still be unreliable if the zero proportions, ν_i's, are close to boundaries at 0 or 1.

A natural method for improving the finite-sample approximations is to use the bootstrap (Efron and Tibshirani, 1993). Let $\mathbf{X}^\omega = \{x_{i1}^\omega, \dots, x_{in_i}^\omega : i = 0, \dots, m\}$ be the non-parametric bootstrap sample, by resampling with replacement from the combined observed data $\mathbf{X} = \{x_{i1}, \dots, x_{in_i} : i = 0, \dots, m\}$. Let R_n^ω be the ELR test statistic based on $\mathbf{X}^\omega$.

Next we study the asymptotic properties of R_n^ω under both the null and alternative hypotheses, which depend on a new set of regularity conditions. Let $\bar{F}(x) = \sum_{i=0}^m \rho_i^* F_i(x)$ and $\bar{\nu}_0^* = \sum_{i=0}^m \rho_i^* \nu_i$.

D1. $\bar{\nu}_0^* \in (0, 1)$.

D2. $\lim_{\min\{n_0,\dots,n_m\}\to\infty} n_i/n \to \rho_i^*$, where $\rho_i^* \in (0,1)$ for $i = 0, \dots, m$.

D3. $\int \left(1, \mathbf{q}^\top(x) I(x > 0)\right)^\top \left(1, \mathbf{q}^\top(x) I(x > 0)\right) d\bar{F}(x)$ exists and is positive definite.

D4. $\int \exp\{\boldsymbol{\beta}_i^\top \mathbf{q}(x)\} I(x > 0) d\bar{F}(x) < \infty$ in a neighbourhood of $\boldsymbol{\beta}_i = \mathbf{0}$ for $i = 0, \dots, m$.

The following theorem establishes the asymptotic properties of the proposed bootstrap procedure.

Theorem 3.2 *Assume that the regularity conditions D1–D4 hold. Then conditional on the observed data* $\mathbf{X}$*, we have*

$$R_n^\omega \xrightarrow{d} \chi^2_{m(d+1)}$$

in probability as $n \to \infty$*. Hence*

$$\sup_x \left| \Pr(R_n^\omega \le x | \mathbf{X}) - \Pr(\chi^2_{m(d+1)} \le x) \right| \to 0$$

in probability as $n \to \infty$*.*

The proof of Theorem 3.2 is given in Section 3.5.2. We give some remarks about Conditions D1–D4 and the results in Theorem 3.2.

(a) The proof of Theorem 3.2 does not rely on any information about $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$ specified under the null hypothesis in (3.3). Hence, the result is valid for the whole model space under (3.1) and (3.2). That is, the true underlying model that generates the observed data $\mathbf{X}$ can be from the null and also alternative hypotheses.

(b) Since Theorem 3.2 is proved to cover the cases that the observed data $\mathbf{X}$ is from the null and also alternative hypotheses, we require the regularity conditions to cover both cases. Conditions D1–D4 serve this purpose. They become Conditions C1–C4 when the observed data is generated under the null hypothesis. The comments for Conditions C1–C4 in Section 3.2.2 can be similarly extended to Conditions D1–D4.

(c) The conditional asymptotic distribution of R_n^ω always approximates $\chi^2_{m(d+1)}$, under both the null and alternative hypotheses that generate the observed data $\mathbf{X}$. This is a desirable property for a resampling method in hypothesis testing context (Pauly et al., 2015). It ensures that the quantiles of the bootstrap distribution of R_n^ω always converges in probability to the quantiles of $\chi^2_{m(d+1)}$, which is used to find a p-value under the null hypothesis even if the null hypothesis is not true. Hence, together with Theorem 3.1, the resulting bootstrap ELR test can asymptotically maintain the nominal type I error under H_0. This further implies that the ELR test can also maintain its consistent power behaviour for any fixed alternative based on bootstrap critical values.

Based on Theorem 3.2, the following nonparametric procedure is then suggested to approximate the p-value of the ELR test.

Step 1. For each b from 1 to B, we repeat the following:

Step 1.1. Generate a bootstrap sample $\left\{x_{i1}^{(b)}, \ldots, x_{in_i}^{(b)} : i = 0, \ldots, m\right\}$ by random sampling with replacement from the combined observed data $\{x_{i1}, \ldots, x_{in_i} : i = 0, \ldots, m\}$.

Step 1.2. Based on this bootstrap sample $\left\{x_{i1}^{(b)}, \ldots, x_{in_i}^{(b)} : i = 0, \ldots, m\right\}$, we can calculate the bootstrap ELR test statistic $R_n^{(b)}$;

Step 2. From Step 1, we obtained a set $\left\{R_n^{(b)} : b = 1, \ldots, B\right\}$. We approximate the p-value of R_n by $\sum_{b=1}^{B} I\left\{R_n^{(b)} \geq R_n\right\}/B$, where R_n is the observed ELR test statistic.

Note that in Step 1.1 we separate the obtained bootstrap sample into $m+1$ segments with each segment length n_i and call the ith segment the bootstrap sample for the ith group, for $i = 0, \ldots, m$.

The choice of B depends on the desired precision level. In our simulation, we set $B = 999$. We show in the next section that the ELR test coupled with the above nonparametric bootstrap procedure produces accurate type I errors in almost all the simulation settings.

3.3 Simulation studies

In this section, we assess the finite-sample performance of the proposed ELR test and the nonparametric bootstrap procedure through Monte Carlo simulation. We further compare the proposed method with some existing methods:

- two-part parametric likelihood ratio tests (LRT) based on the assumption of a log-normal distribution for the nonnegative part (LRT-LN) or the assumption of a gamma distribution for the nonnegative part (LRT-GAM);
- the one-part Wald-type statistic (WTS1) proposed by Pauly et al. (2015) for testing a linear hypothesis about the means without using any distributional assumptions under very general heteroscedastic factorial designs, and also its two-part version (WTS2);
- the one-part generalized ANOVA-type statistic (ATS) proposed by Brunner et al. (1997), which was shown in Vallejo et al. (2010) and Pauly et al. (2015) to accurately maintain the preassigned level in most cases and it is conservative for skewed distributions;
- the one-part Kruskal-Wallis test (KW1), and its two-part version (KW2);

- the one-part ANOVA F-test (ANOVA1), and its two-part version (ANOVA2).

The KW test, the ANOVA test, and the fully parametric LRTs are classical methods to test the equality of multiple distributions. Their two-part versions are also classical methods and are studied in Lachenbruch (2001). In the context of zero-excess data, it is well known that skewness is an important characteristic. Vallejo et al. (2010) concluded that the ATS method should be recommended when the distributions are moderately skewed. Lastly, the WTS is a distribution-free and valid method in general heteroscedastic factorial designs and may be adapted to the case with excess zeros (Pauly et al., 2015).

We fix the number of populations under consideration to be $m + 1 = 3$ and generate random observations, conditional on all $\hat{\nu}_i$'s $\neq 0$ or 1, from (3.1) with all the G_i's being log-normal, or all the G_i's being gamma. Note that if any $\hat{\nu}_i$'s $= 0$ or 1, then some test statistics may not be well defined. This is not a problem in practice. However, note that, when any true zero proportion ν_i is too close to the boundary 0 or 1, Anaya-Izquierdo et al. (2014) found that boundary effects can dominate the sampling distribution of $\hat{\nu}_i$. Hence, the true ν_i's are considered to be between 0.2 and 0.7 in our simulation settings. A diagnostic tool was proposed in Anaya-Izquierdo et al. (2014) which defines how far ν_i is required to be from the boundary so that first order asymptotics remain adequate. Further, note that, when the combined sample contains only a few nonzero observations, resampling methods may not be very helpful to remove such boundary effects (Chen et al., 2003).

In the following, we use $\text{LN}(a_i, b_i)$ to denote a log-normal distribution with mean a_i and variance b_i both with respect to the log scale (i.e., the mean and variance of the associated normal random variable) and $\text{GAM}(a_i, b_i)$ to denote a gamma distribution with shape parameter a_i and scale parameter b_i. The parameter settings under the null (LN_1–LN_3 and GAM_1–GAM_3) and alternative (LN_4–LN_{15} and GAM_4–GAM_{15}) models are given in Table 3.1.

We consider the case with equal sample sizes by setting n_i to be 20, 50 or 100 for $i = 0, 1, 2$; and also consider the case with unequal sample sizes that $(n_0, n_1, n_2) = (50, 100, 150)$.

For all tests, the type I error rates and power at the 5% significance level are calculated based on 10,000 repetitions. The type I error rates are calculated using the limiting distribution and the bootstrap procedure proposed in Section 3.2.3 with $B = 999$. For fair comparison, the bootstrap procedure with $B = 999$ is used to calculate the powers of all methods. For the one-part and two-part WTS test statistics, the computation is completed by using `R` package "`GFD`" (Friedrich et al., 2016). Following Pauly et al. (2015), we also use the permutation method with 10,000 permutation samples (the default number in "`GFD`") to calculate the type I error rates and powers of one-part and two-part WTS tests.

For the clarity of presentation, we only present the results of ELR test under the correctly specified basis function $\mathbf{q}(x)$. That is, ELR test under $\mathbf{q}(x) = \{x, \log(x)\}^\top$ for the GAM models; and ELR test under $\mathbf{q}(x) = \{\log(x), \log^2(x)\}^\top$ for the LN models. For more choices of $\mathbf{q}(x)$, interested readers may refer to Section 3.6. Moreover, we only present the comparisons of the ELR test with the correctly specified LRT (which serves as the benchmark test), the WTS2, and the KW2. The ANOVA2 has similar or even less power than WTS2. The one-part tests are in general comparable to, or less powerful than, the corresponding two-part versions. Hence, the power comparisons with ANOVA2 and one-part tests are not presented in this current section. Further, for WTS2 method, bootstrap and permutation procedures give the consistent results; see Section 3.6. For fair comparison, we only present the results using the bootstrap procedure for WTS2. In Section 3.6, complete simulation results for all tests are available.

3.3.1 Type I error

The simulated type I error rates for the four selected representative tests are summarized in Tables 3.2–3.3. We have observed, based on our simulation results, that the type I error rates based on asymptotic distributions for all tests tend to be inflated, particularly, when sample sizes are small for ELR, LRT and WTS2, and when sample sizes are small and zero proportions are low for KW2. On the other hand, the bootstrap procedure provides satisfactory adjustment to the type I error rates, except for LN_3 and GAM_3 with sample sizes $(20, 20, 20)$. In these settings, we may encounter a number of bootstrap samples contain at most one nonzero observations. For such a bootstrap sample, some test statistics

Table 3.1: Parameter settings for simulation studies. In the first column, each LN_1–LN_{15} and each GAM_1–GAM_{15} denote mixture models whose continuous parts follow the distributions $LN(a_i, b_i)$ and $GAM(a_i, b_i)$, respectively, for $i = 0, 1, 2$. The last two columns are the means and variances corresponding to each model.

Model	(ν_0, ν_1, ν_2)	(a_0, a_1, a_2)	(b_0, b_1, b_2)	Means	Variances
LN_1	(0.2, 0.2, 0.2)	(0.0, 0.0, 0.0)	(1.0, 1.0, 1.0)	(1.65, 1.65, 1.65)	(4.67, 4.67, 4.67)
LN_2	(0.4, 0.4, 0.4)	(0.0, 0.0, 0.0)	(1.0, 1.0, 1.0)	(1.65, 1.65, 1.65)	(4.67, 4.67, 4.67)
LN_3	(0.7, 0.7, 0.7)	(0.0, 0.0, 0.0)	(1.0, 1.0, 1.0)	(1.65, 1.65, 1.65)	(4.67, 4.67, 4.67)
LN_4	(0.2, 0.3, 0.4)	(0.0, 0.0, 0.0)	(1.0, 1.0, 1.0)	(1.32, 1.15, 0.99)	(4.17, 3.84, 3.45)
LN_5	(0.4, 0.4, 0.4)	(0.0, 0.5, 1.0)	(2.0, 2.0, 2.0)	(1.63, 2.69, 4.43)	(30.10, 81.82, 222.40)
LN_6	(0.6, 0.6, 0.6)	(0.0, 0.0, 0.0)	(1.0, 2.0, 3.0)	(0.66, 1.09, 1.79)	(2.52, 20.66, 158.16)
LN_7	(0.5, 0.6, 0.7)	(0.0, 0.5, 1.0)	(3.0, 2.0, 1.0)	(2.24, 1.79, 1.34)	(196.69, 56.15, 14.57)
LN_8	(0.6, 0.6, 0.6)	(0.0, 0.5, 1.0)	(3.0, 2.0, 1.0)	(1.79, 1.79, 1.79)	(158.16, 56.15, 18.63)
LN_9	(0.3, 0.4, 0.5)	(0.0, 0.15, 0.34)	(2.0, 2.0, 2.0)	(1.90, 1.90, 1.90)	(34.60, 40.97, 49.89)
LN_{10}	(0.4, 0.5, 0.6)	(0.0, 0.0, 0.0)	(2.0, 2.36, 2.81)	(1.63, 1.63, 1.63)	(30.10, 53.95, 107.90)
LN_{11}	(0.4, 0.5, 0.6)	(0.0, 0.5, 1.0)	(2.69, 2.05, 1.5)	(2.30, 2.30, 2.30)	(124.67, 77.32, 54.07)
LN_{12}	(0.5, 0.5, 0.5)	(0.0, 0.5, 1.0)	(2.46, 1.98, 1.5)	(1.71, 2.21, 2.88)	(65.93, 65.93, 65.93)
LN_{13}	(0.3, 0.4, 0.5)	(0.0, 0.07, 0.15)	(2.0, 2.0, 2.0)	(1.90, 1.75, 1.58)	(34.60, 34.60, 34.60)
LN_{14}	(0.3, 0.4, 0.5)	(0.0, 0.0, 0.0)	(2.0, 2.07, 2.15)	(1.90, 1.69, 1.46)	(34.60, 34.60, 34.60)
LN_{15}	(0.4, 0.5, 0.6)	(0.0, 0.5, 1.0)	(2.28, 1.88, 1.5)	(1.88, 2.11, 2.30)	(54.07, 54.07, 54.07)
GAM_1	(0.2, 0.2, 0.2)	(1.0, 1.0, 1.0)	(1.0, 1.0, 1.0)	(0.80, 0.80, 0.80)	(0.96, 0.96, 0.96)
GAM_2	(0.4, 0.4, 0.4)	(1.0, 1.0, 1.0)	(1.0, 1.0, 1.0)	(0.60, 0.60, 0.60)	(0.84, 0.84, 0.84)
GAM_3	(0.7, 0.7, 0.7)	(1.0, 1.0, 1.0)	(1.0, 1.0, 1.0)	(0.30, 0.30, 0.30)	(0.51, 0.51, 0.51)
GAM_4	(0.2, 0.3, 0.4)	(1.0, 1.0, 1.0)	(2.0, 2.0, 2.0)	(1.6, 1.4, 1.2)	(3.84, 3.64, 3.36)
GAM_5	(0.6, 0.6, 0.6)	(1.0, 1.5, 2.0)	(2.0, 2.0, 2.0)	(0.8, 1.2, 1.6)	(2.56, 4.56, 7.04)
GAM_6	(0.6, 0.6, 0.6)	(1.0, 1.0, 1.0)	(1.0, 2.0, 3.0)	(0.4, 0.8, 1.2)	(0.64, 2.56, 5.76)
GAM_7	(0.4, 0.5, 0.6)	(1.0, 1.5, 3.0)	(3.0, 2.0, 1.0)	(1.8, 1.5, 1.2)	(7.56, 5.25, 3.36)
GAM_8	(0.5, 0.5, 0.5)	(1.0, 1.5, 3.0)	(3.0, 2.0, 1.0)	(1.5, 1.5, 1.5)	(6.75, 5.25, 3.75)
GAM_9	(0.4, 0.5, 0.6)	(1.5, 1.8, 2.25)	(2.0, 2.0, 2.0)	(1.8, 1.8, 1.8)	(5.76, 6.84, 8.46)
GAM_{10}	(0.4, 0.5, 0.6)	(1.0, 1.0, 1.0)	(2.0, 2.4, 3.0)	(1.2, 1.2, 1.2)	(3.36, 4.32, 5.76)
GAM_{11}	(0.4, 0.5, 0.6)	(2.0, 3.0, 4.0)	(2.0, 1.6, 1.5)	(2.4, 2.4, 2.4)	(8.64, 9.60, 12.24)
GAM_{12}	(0.4, 0.4, 0.4)	(1.0, 1.5, 3.0)	(2.0, 1.53, 0.92)	(1.20, 1.37, 1.66)	(3.36, 3.36, 3.36)
GAM_{13}	(0.3, 0.4, 0.5)	(1.5, 1.56, 1.66)	(2.0, 2.0, 2.0)	(2.1, 1.87, 1.66)	(6.09, 6.09, 6.09)
GAM_{14}	(0.3, 0.4, 0.5)	(1.0, 1.0, 1.0)	(2.0, 2.08, 2.20)	(1.4, 1.25, 1.1)	(3.64, 3.64, 3.64)
GAM_{15}	(0.4, 0.5, 0.6)	(2.0, 3.0, 4.0)	(2.0, 1.52, 1.26)	(2.4, 2.28, 2.02)	(8.64, 8.64, 8.64)

may not be well defined. Our treatment is to simply delete such bootstrap sample. By doing so, the type I error may be slightly inflated.

In Section 3.6, we demonstrate, by additional simulations, that similar conclusions on the type I error rates can be drawn for the bootstrap ELR test under five particular basis functions which may be incorrectly specified. This confirms that the asymptotic size of the

Table 3.2: Type I error rates (%) for testing H_0 at significance level 0.05 when data are generated from a log-normal mixture model with parameter settings given in Table 3.1. The ELR test is defined under $\mathbf{q}(x) = \{\log(x), \log^2(x)\}^\top$.

Model	(n_0, n_1, n_2)	ELR		LRT-LN		WTS2		KW2	
		asymptotic	bootstrap	asymptotic	bootstrap	asymptotic	bootstrap	asymptotic	bootstrap
LN_1	$(20, 20, 20)$	8.12	4.87	7.03	4.12	7.33	4.66	5.73	4.81
	$(50, 50, 50)$	6.12	4.89	5.79	4.81	5.54	4.64	5.10	4.88
	$(50, 100, 150)$	5.89	5.20	5.76	5.22	6.72	4.84	5.37	5.31
	$(100, 100, 100)$	5.70	5.06	5.32	4.89	5.43	4.91	4.80	4.69
LN_2	$(20, 20, 20)$	8.58	4.78	7.35	3.99	7.00	4.64	4.98	4.87
	$(50, 50, 50)$	6.80	5.03	5.91	4.75	6.40	5.14	5.11	4.99
	$(50, 100, 150)$	6.10	5.17	5.64	5.00	7.14	4.92	5.35	5.40
	$(100, 100, 100)$	6.03	5.17	5.50	5.17	5.77	5.11	5.52	5.46
LN_3	$(20, 20, 20)$	12.54	5.66	10.02	3.47	8.80	5.79	3.26	5.79
	$(50, 50, 50)$	7.94	4.81	6.61	4.36	7.00	4.65	4.96	4.81
	$(50, 100, 150)$	6.91	4.99	6.30	4.95	8.65	5.33	5.55	5.59
	$(100, 100, 100)$	6.42	4.98	5.73	4.58	6.10	5.00	4.76	4.80

NOTE: the Monte Carlo error is 0.218 (%) under the null models LN_1-LN_3.

Table 3.3: Type I error rates (%) for testing H_0 at significance level 0.05 when data are generated from a gamma mixture model with parameter settings given in Table 3.1. The ELR test is defined under $\mathbf{q}(x) = \{x, \log(x)\}^\top$.

Model	(n_0, n_1, n_2)	ELR		LRT-GAM		WTS2		KW2	
		asymptotic	bootstrap	asymptotic	bootstrap	asymptotic	bootstrap	asymptotic	bootstrap
GAM_1	$(20, 20, 20)$	7.97	4.63	7.06	4.39	8.04	5.18	6.13	5.16
	$(50, 50, 50)$	5.98	4.53	5.38	4.38	6.18	4.98	5.09	4.97
	$(50, 100, 150)$	5.58	4.81	5.33	4.67	5.82	4.74	5.04	4.93
	$(100, 100, 100)$	5.67	5.09	5.42	5.09	5.54	5.06	5.12	4.94
GAM_2	$(20, 20, 20)$	8.91	4.99	7.37	4.32	7.99	5.03	5.29	5.14
	$(50, 50, 50)$	6.22	4.59	5.33	4.32	5.82	4.72	4.90	4.81
	$(50, 100, 150)$	5.84	4.92	5.32	4.74	6.26	4.88	5.13	5.11
	$(100, 100, 100)$	5.52	5.04	5.32	4.88	5.50	5.07	5.13	5.15
GAM_3	$(20, 20, 20)$	11.95	5.75	10.01	3.66	9.68	5.10	3.10	5.60
	$(50, 50, 50)$	7.87	4.99	6.94	4.64	7.34	4.77	5.13	5.02
	$(50, 100, 150)$	6.70	4.95	6.16	4.84	7.42	5.01	5.04	5.07
	$(100, 100, 100)$	6.60	4.99	5.90	4.77	6.36	5.24	5.33	5.39

NOTE: the Monte Carlo error is 0.218 (%) under the null models GAM_1-GAM_3.

ELR test is robust to the choice of $\mathbf{q}(x)$ and the assumption on $G_0(x)$.

3.3.2 Testing power

For clarity and to aid discussion, we categorize the alternative model specifications into a 3×4 table given in Table 3.4. For the rows, these are considered to include the scenarios that the non-homogeneity is: due to both means and variances (Scenario R.I), mainly due to the variances (Scenario R.II), or mainly due to the means (Scenario R.III). For the columns, the specifications can be divided into scenarios in which: either the zero proportions or the nonnegative components are held constant (Scenario C.I), the parameters b_i's of nonnegative components are held constant (Scenario C.II), the parameters a_i's of non-negative components are held constant (Scenario C.III), or all the parameters are different (Scenario C.IV).

Table 3.4: Scenarios categorized according to the alternative model settings in Table 3.1.

Scenario	C.I	C.II	C.III	C.IV
R.I	LN_4 & GAM_4	LN_5 & GAM_5	LN_6 & GAM_6	LN_7 & GAM_7
R.II	LN_8 & GAM_8	LN_9 & GAM_9	LN_{10} & GAM_{10}	LN_{11} & GAM_{11}
R.III	LN_{12} & GAM_{12}	LN_{13} & GAM_{13}	LN_{14} & GAM_{14}	LN_{15} & GAM_{15}

The simulated powers for four selected representative tests, adjusted using bootstrap procedure, are plotted in Figures 3.1–3.2. Together with the additional simulation results in Section 3.6, we make the following comments.

(a) There seems to be no single dominant method in terms of the power of the tests in all the scenarios considered.

(b) The fully parametric LRT is known as the asymptotically most powerful test if the parametric model is correctly specified. It can be observed that the proposed bootstrap ELR test performs almost as well as the fully parametric LRT under correct model assumptions in all the settings investigated. But not surprisingly, the bootstrap ELR test performs better than the parametric LRT under incorrect parametric assumptions, from the results in Section 3.6. These findings confirm that the bootstrap ELR test is robust against the departure of parametric assumptions.

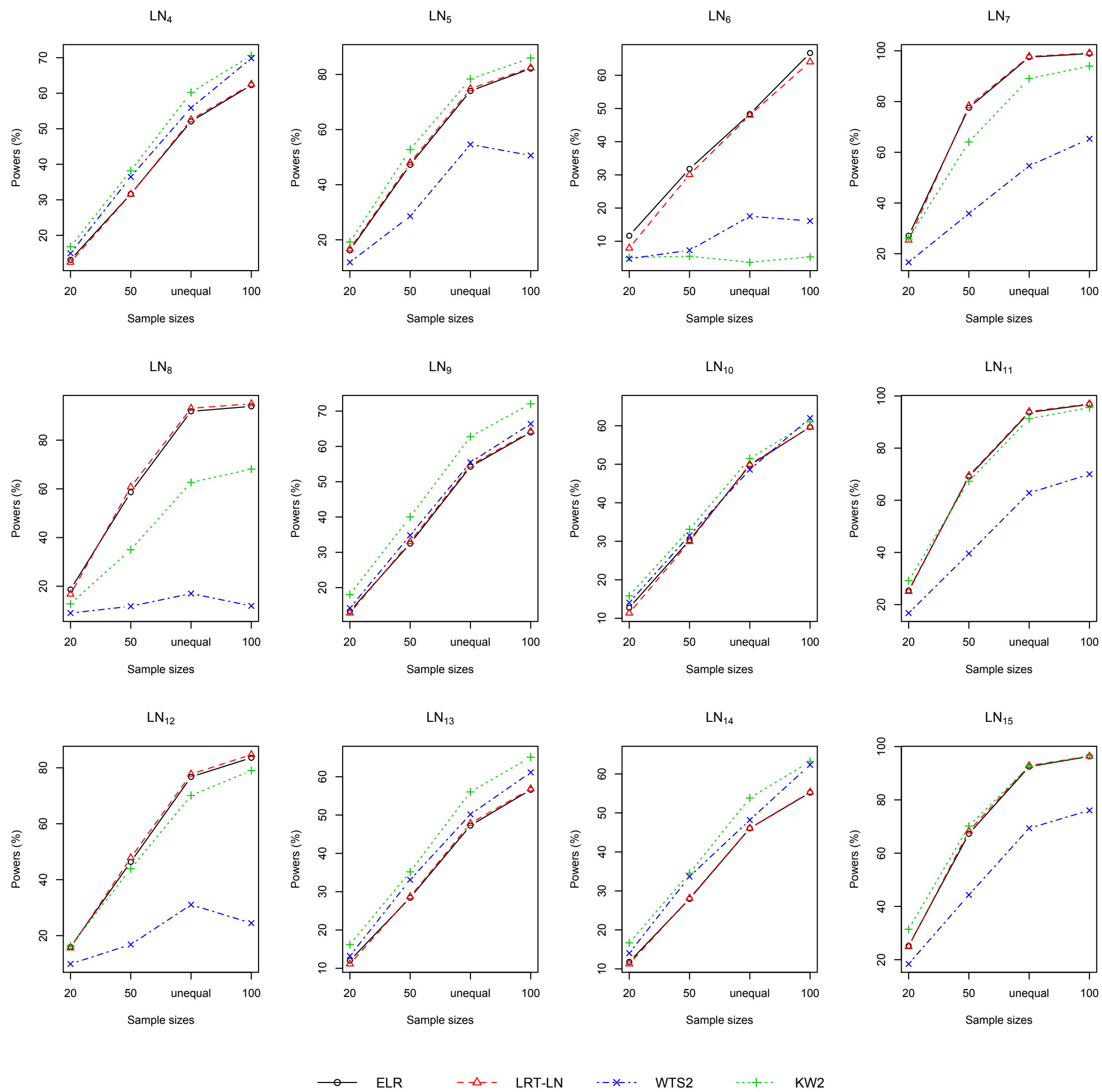

Figure 3.1: Simulated powers (%) of rejecting H_0 at significance level 0.05 when data are generated from a log-normal mixture model with parameter settings given in Table 3.1. The horizontal axis denotes four combinations of sample sizes (n_0, n_1, n_2) equal to $(20, 20, 20)$, $(50, 50, 50)$, $(50, 100, 150)$ and $(100, 100, 100)$, from left to right.

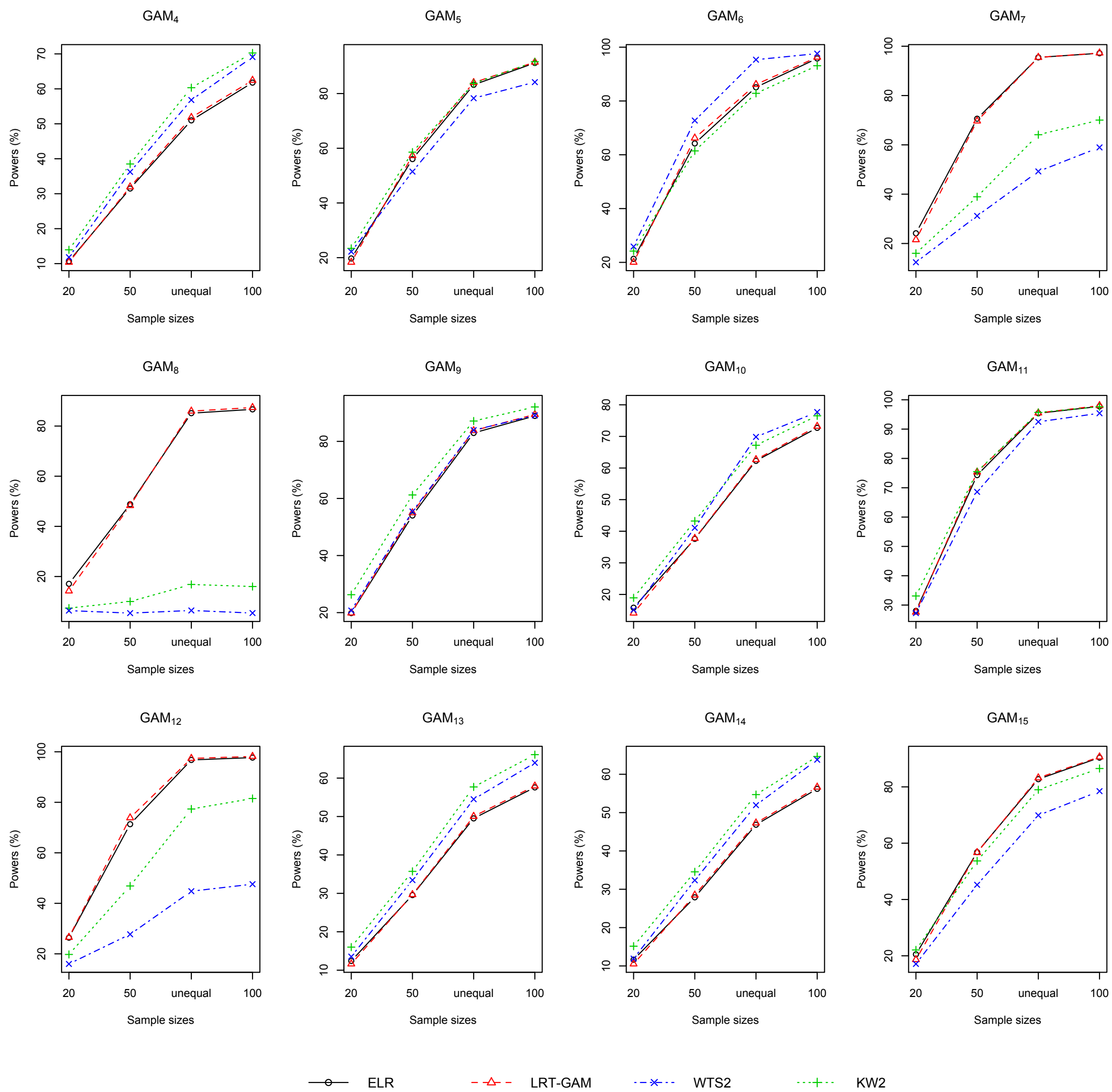

Figure 3.2: Simulated powers (%) of rejecting H_0 at significance level 0.05 when data are generated from a gamma mixture model with parameter settings given in Table 3.1. The horizontal axis denotes four combinations of sample sizes (n_0, n_1, n_2) equal to $(20, 20, 20)$, $(50, 50, 50)$, $(50, 100, 150)$ and $(100, 100, 100)$, from left to right.

(c) The comparison with one-part tests is provided in Section 3.6. In some settings, the one-part tests may preform slightly better than the corresponding two-part versions, but can fail dramatically such as in Scenario C.IV for KW1, and in Scenario R.II for ANOVA1, WTS1, and ATS. This phenomenon may be due to the fact that the one-part tests ignore the mixture structure in (3.1). In general, the ELR test is more powerful than the one-part tests except for KW1 under LN_4, LN_{10}, LN_{14}, GAM_4, and GAM_{14}. The performance of ELR test and KW1 are very comparable under LN_4, GAM_4, and GAM_{14}, while the KW1 is slightly more powerful for small sample sizes. Under LN_{10} and LN_{14}, the KW1 is more powerful than the ELR test for all sample sizes. These two models belong to the case that the mixture proportions and variances on the log scale (i.e. b_i's) are different in a log-normal mixture model.

(d) For all the two-part tests, the first parts are likelihood ratio tests for testing homogeneity in the zero proportions. Therefore, the power differences of two-part tests are due to the second parts for testing homogeneity in $G_i(x)$'s. If both a_i's and b_i's are different (i.e., Scenarios C.I and C.IV except for LN_4 and GAM_4), the ELR test is the most powerful test, or one of the most powerful tests. For example, the ELR test has the most advantage for LN_7, LN_8, GAM_7, and GAM_8. If a_i's or b_i's or both are the same (i.e., Scenarios C.II and C.III, LN_4, and GAM_4), then KW2 and/or WTS2 could be sometimes more powerful than the ELR test except for LN6. This limitation of the ELR test is mainly because the $\mathbf{q}(x)$ in the DRM (3.2) is over-fitted. If a more parsimonious basis function $\mathbf{q}(x)$ is used, then the ELR test could become comparable to KW2 and/or WTS2. More precisely, in LN_{13} and $(n_0, n_1, n_1) = (50, 100, 150)$, the powers of KW2, WTS2, and ELR with basis function $\mathbf{q}(x) = \{\log(x), \log^2(x)\}^\top$ are 56.04%, 50.19%, and 47.25%, respectively. The most parsimonious basis function under LN_{13} is $\mathbf{q}(x) = \log(x)$. With this basis function, the power of the ELR test increases to 55.33%, which becomes comparable to other two-part tests. Lastly, we explain the trend exception of LN_6. Note that in LN_6, the means of G_i's are only slightly different compared with the magnitude of variances. Recall that WTS2 is mainly designed for testing the mean differences in G_i's. This explains why the WTS2 is less powerful than the ELR test for LN_6. Also note that the non-homogeneity in F_i's under LN_6 is only due to the differences in b_i's. Since the KW test is invariant to the monotone transformation (Kruskal, 1952), the KW2 is not able to detect the differences in b_i's (the variances in log scale) of a log-normal mixture model. This explains why the KW2 is much less powerful than the ELR test for LN_6.

(e) To examine how sensitive the power of the ELR test is to the choice of $\mathbf{q}(x)$, more comprehensive simulation results, that cover under-fitting, over-fitting, and misspecification of a DRM, are given in Section 3.6. In general, we have observed that the bootstrap ELR tests based on the correctly specified basis functions have higher power than those based on the misspecified, under-fitted or over-fitted basis functions. Hence, in practice, to achieve appreciable testing power, we suggest selecting a suitable basis function $\mathbf{q}(x)$ in a DRM among several competing models using some information criteria, for example, Akaike's information criterion (AIC) as described in Section 2.3. We note that our current hypothesis testing problem is slightly different from that considered in Fokianos (2007). Hence, the use of AIC for selecting $\mathbf{q}(x)$ and its impact on the power performance may deserve further investigation, which is beyond the scope of this chapter.

3.4 Testing with a real data set

In this section, we employ a real data example to illustrate the proposed ELR test and the nonparametric bootstrap procedure.

The aim of this real example was to investigate precipitation distribution changes which plays an important role in meteorology. The data here is available from the website of the University of Waterloo weather station data archive (`http://weather.uwaterloo.ca/data.html`). The data set records daily precipitation measurements (in millimetres, and in winter the number is referred to as the snow-water equivalent) for several years since 2003 based on the GeoNor T-200B Precipitation Gauge (installed in December 2002) located in the North Campus of the University of Waterloo, Canada.

For illustration purposes, we consider the data for years 2003–2006. We are interested in detecting if there are any precipitation distribution changes over these years. To reduce the time dependence among the observations, we take every 4th measurement into our analysis, i.e. the observations on days $1, 5, 9, \ldots, 361$, which gives a total sample size $n = 364$. Similar conclusions were drawn when we considered other subsets of the data. Some summary statistics are:

- the estimate of $\boldsymbol{\nu}^\top$ is $(0.3187, 0.3956, 0.4176, 0.4176)$ with sample sizes $(91, 91, 91, 91)$;
- the sample means are $(2.0923, 3.5396, 3.3978, 3.4978)$;
- the sample variances are $(16.6392, 41.0662, 76.1044, 59.4987)$.

We further fit this data by log-normal mixture model and gamma mixture model under the null and alternative hypotheses by the parametric maximum likelihood method. The details are provided in Table 3.5. These models are used in our confirmative simulation later on. It seems that the means and variances fitted by the gamma mixture model match the real data quite closely.

Table 3.5: Fitted parameters for log-normal mixture model and gamma mixture model under the null and alternative hypotheses for Waterloo precipitation data. The models LN_{16} and GAM_{16} are fitted under the null hypothesis; and the models LN_{17} and GAM_{17} are fitted under the alternative hypothesis. The last two columns are the means and variances corresponding to each model, and $\mathbf{1}^\top = (1, 1, 1, 1)$.

Model	$(\nu_0, \nu_1, \nu_2, \nu_3)$	(a_0, a_1, a_2, a_3)	(b_0, b_1, b_2, b_3)	Means	Variances
LN_{16}	$0.39 \times \mathbf{1}^\top$	$0.52 \times \mathbf{1}^\top$	$2.58 \times \mathbf{1}^\top$	$3.74 \times \mathbf{1}^\top$	$285.90 \times \mathbf{1}^\top$
LN_{17}	$(0.32, 0.40, 0.42, 0.42)$	$(0.18, 1.05, 0.36, 0.52)$	$(2.18, 1.70, 3.08, 3.00)$	$(2.43, 4.04, 3.91, 4.39)$	$(70.62, 131.52, 559.29, 645.92)$
GAM_{16}	$0.39 \times \mathbf{1}^\top$	$0.56 \times \mathbf{1}^\top$	$9.11 \times \mathbf{1}^\top$	$3.13 \times \mathbf{1}^\top$	$34.74 \times \mathbf{1}^\top$
GAM_{17}	$(0.32, 0.40, 0.42, 0.42)$	$(0.65, 0.82, 0.46, 0.50)$	$(4.72, 7.12, 12.69, 12.04)$	$(2.09, 3.54, 3.40, 3.50)$	$(11.92, 33.41, 51.40, 50.89)$

We applied the proposed bootstrap ELR tests and other testing methods to test if the precipitation distribution changed for the four years. The various test statistics and their p-values based on 10,000 times bootstrap resampling are summarized in Table 3.6.

It is seen that the proposed ELR tests based on the first three basis functions all give highly significant p-values at the 5% significance level. The bootstrap ELR test based

Table 3.6: Test statistics, corresponding p-values, and related simulation results based on the bootstrap procedures with $B = 10,000$ for Waterloo precipitation data.

Method	Precipitation data			Confirmative simulation			
	AIC	Test statistic	p-value	LN_{16}	LN_{17}	GAM_{16}	GAM_{17}
ELR							
$\mathbf{q}(x) = \{x, \log(x)\}^\top$	614.54	23.36	0.0070	5.40	82.81	4.96	90.74
$\mathbf{q}(x) = \{\log(x), \log^2(x)\}^\top$	616.78	21.12	0.0155	5.26	87.48	4.90	81.93
$\mathbf{q}(x) = \{x, \log(x), \log^2(x)\}^\top$	618.94	24.96	0.0210	5.16	81.63	5.18	86.08
$\mathbf{q}(x) = x$	622.62	9.28	0.1746	5.41	42.52	5.02	75.76
$\mathbf{q}(x) = \log(x)$	619.97	11.93	0.0672	5.07	75.34	5.15	60.00
KW2	-	11.70	0.0696	5.14	77.22	4.98	65.62
KW1	-	2.33	0.5005	5.10	22.08	4.89	12.27
ANOVA2	-	7.84	0.2538	4.79	33.86	5.09	68.28
ANOVA1	-	0.91	0.4383	4.37	11.64	5.07	31.50
LRT-LN	-	18.30	0.0155	5.12	87.58	4.45	66.74
LRT-GAM	-	20.20	0.0361	4.48	44.06	4.89	90.74
ATS	-	0.91	0.4383	4.37	11.64	5.07	31.50
WTS1	-	5.10	0.1870	4.96	19.76	4.87	47.94
WTS2	-	12.18	0.0747	4.87	42.32	4.96	82.37

on $\mathbf{q}(x) = \{x, \log(x)\}^\top$ has the smallest AIC. Hence, the DRM with $\mathbf{q}(x) = \{x, \log(x)\}^\top$ provides the best fit to the data among the five commonly used basis functions as guided by Table 2.1. Therefore, under a reasonably fitted DRM, we gain some evidence that the precipitation distribution was changing over 2003–2006 based on the observed data. Even though we are not able to exhaustively search all possible basis functions $\mathbf{q}(x)$, our ELR test based on this fitted $\mathbf{q}(x) = \{x, \log(x)\}^\top$ already has enough power of rejecting the homogeneous null hypothesis at 5% significance level, and our conclusion is reliable. This is because of the fact that misspecifing, over- or under-fitting the basis function $\mathbf{q}(x)$ in the DRM may only result in some loss of power (Section 3.6), while the type I error can still be controlled regardless the choice of $\mathbf{q}(x)$. The total computational time for obtaining the p-value of bootstrap ELR test based on $\mathbf{q}(x) = \{x, \log(x)\}^\top$ is around 2 minutes in an IMAC with a 3.4-GHz Intel Core i7 processor.

The two parametric likelihood ratio tests also produce significant p-values, however, at the risk of misspecifying the underlying parametric models, which may significantly decrease the power of the parametric likelihood ratio test as we have discussed in Section 3.2.3. On the other hand, the proposed bootstrap ELR test guarantees the control of type

I error with data-discovered underlying distributions and thus the conclusion is much more reliable.

The above results may be further verified by the confirmative simulation with model settings in Table 3.5. The simulation results are based on 10,000 repetitions. For each repetition, $B = 10,000$ bootstrap samples are used to calculate the p-value for each test. For WTS1 and WTS2, we also use $10,000$ permutation samples to calculate the p-values. The results are very similar to those from the bootstrap method and thus are omitted here. The results are summarized in Table 3.6, which show that the resampling adjusted type I error rates for all tests are well controlled for the null models LN_{16} and GAM_{16}. Since a_i's, and b_i's are all different, the ELR test with basis function $\mathbf{q}(x) = \{x, \log(x)\}^\top$ and LRT-GAM are the most powerful tests under GAM_{17}. Recall that the alternative model GAM_{17} well captures some characteristics of this data. The simulation results under GAM_{16} and GAM_{17} further confirm that our findings for the ELR test with basis function $\mathbf{q}(x) = \{x, \log(x)\}^\top$ may not be obtained by chance.

3.5 Proofs

This section contains proofs of Theorem 3.1 and Theorem 3.2.

3.5.1 Proof of Theorem 3.1

Recall that the true value of $\boldsymbol{\nu}$ is $\boldsymbol{\nu}^* = (\nu_0^*, \ldots, \nu_0^*)^\top$ under H_0. The true value of $\boldsymbol{\theta}$ is $\mathbf{0}$ under H_0. The regularity conditions provided in Section 3.2.2 will be needed throughout this proof.

Recall that $R_n = R_{n,1} + R_{n,2}$ with

$$R_{n,1} = 2\left\{\ell_0(\hat{\boldsymbol{\nu}}) - \max_{H_0} \ell_0(\boldsymbol{\nu})\right\}, \qquad R_{n,2} = 2\ell_1(\hat{\boldsymbol{\theta}}).$$

The roadmap for proving Theorem 3.1 is as follows. In Step 1, we argue that the null limiting distribution of $R_{n,1}$ is χ^2_m. In Step 2, we argue that the null limiting distribution of $R_{n,2}$ is χ^2_{md}. In Step 3, we argue that $R_{n,1}$ and $R_{n,2}$ are asymptotically independent. Theorem 3.1 then follows.

Note that $\ell_0(\boldsymbol{\nu})$ is the summation of binomial log-likelihood based on n_{i0}'s and $R_{n,1}$ is the parametric likelihood ratio test statistic for testing the null hypothesis that $\nu_0 = \cdots = \nu_m$. According to the classical likelihood theory (Serfling, 1980, Chapter 4), we have under H_0

$$R_{n,1} = \frac{1}{n}\frac{\partial \ell_0(\boldsymbol{\nu}^*)}{\partial \boldsymbol{\nu}^\top}\left\{\boldsymbol{\Sigma}^{-1} - \nu_0^*(1-\nu_0^*)\mathbf{1}_{m+1}\mathbf{1}_{m+1}^\top\right\}\frac{\partial \ell_0(\boldsymbol{\nu}^*)}{\partial \boldsymbol{\nu}} + o_p(1), \tag{3.6}$$

where $\boldsymbol{\Sigma} = \frac{1}{\nu_0^*(1-\nu_0^*)}\text{diag}(\rho_0^*, \ldots, \rho_m^*)$ and $\mathbf{1}_{m+1}$ is a $(m+1)$-dimensional vector with all ones. It can be verified that

$$\frac{\partial \ell_0(\boldsymbol{\nu}^*)}{\partial \boldsymbol{\nu}^\top} = \left\{\frac{n_{00} - n_0\nu_0^*}{\nu_0^*(1-\nu_0^*)}, \ldots, \frac{n_{m0} - n_m\nu_0^*}{\nu_0^*(1-\nu_0^*)}\right\}^\top.$$

By the central limit theorem, we have that

$$n^{-1/2}\frac{\partial \ell_0(\boldsymbol{\nu}^*)}{\partial \boldsymbol{\nu}} \xrightarrow{d} N(0, \boldsymbol{\Sigma}).$$

After some algebra, we can further check

$$\left(\boldsymbol{\Sigma}^{-1} - \nu_0^*(1-\nu_0^*)\mathbf{1}_{m+1}\mathbf{1}_{m+1}^\top\right)\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^{-1} - \nu_0^*(1-\nu_0^*)\mathbf{1}_{m+1}\mathbf{1}_{m+1}^\top\right) = \boldsymbol{\Sigma}^{-1} - \nu_0^*(1-\nu_0^*)\mathbf{1}_{m+1}\mathbf{1}_{m+1}^\top$$

and $\text{rank}\left(\boldsymbol{\Sigma}^{-1} - \nu_0^*(1-\nu_0^*)\mathbf{1}_{m+1}\mathbf{1}_{m+1}^\top\right) = m$. Hence, it follows that

$$R_{n,1} \xrightarrow{d} \chi^2_m.$$

This finishes Step 1.

In Step 2, we derive the limiting distribution of $R_{n,2}$ under the null hypothesis, which requires the asymptotic properties of $\partial \ell_1(\mathbf{0})/\partial \boldsymbol{\theta}$, $\partial^2 \ell_1(\mathbf{0})/\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top$ and $\hat{\boldsymbol{\theta}}$. To present these results, we need some notation. In the following, the expectation, variance and covariance are taken under $\nu_0^* I(x=0) + (1-\nu_0^*)G_0(x)I(x>0)$, the true null distribution. Without loss of generality, we assume that $E\{\mathbf{q}(X)I(X>0)\} = \mathbf{0}$. Otherwise a transformation can be applied. Let $\mathbf{Q} = \left\{\sum_{j=1}^{n_0}\mathbf{q}^\top(x_{0j})I(x_{0j}>0), \ldots, \sum_{j=1}^{n_m}\mathbf{q}^\top(x_{mj})I(x_{mj}>0)\right\}^\top$, $\mathbf{I}_m$ denotes a $m \times m$ unit diagonal matrix, and $\mathbf{1}_m$ denotes a m-dimensional vector with all entries being one. Further, let $\mathbf{H} = \text{diag}(\boldsymbol{\rho}) - \boldsymbol{\rho}\boldsymbol{\rho}^\top$ with $\boldsymbol{\rho} = (\rho_1^*, \ldots, \rho_m^*)^\top$.

Lemma 3.1 *Assume the same conditions as Theorem 3.1. We have under H_0, as $n \to \infty$*

(a) $\frac{\partial \ell_1(\mathbf{0})}{\partial \boldsymbol{\alpha}} = \mathbf{0}$, $\frac{\partial \ell_1(\mathbf{0})}{\partial \boldsymbol{\beta}} = \{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathbf{Q} + o_p(n^{1/2})$, *and*

$$n^{-1/2}\frac{\partial \ell_1(\mathbf{0})}{\partial \boldsymbol{\beta}} \xrightarrow{d} N(\mathbf{0}, \mathbf{V}),$$

where $\mathbf{V} = \mathbf{H} \otimes Var\{\mathbf{q}(X)I(X > 0)\}$ *and* $\otimes$ *denotes the Kroneker product;*

(b) $-\frac{1}{n}\frac{\partial^2 \ell_1(\mathbf{0})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} \to \mathbf{U}$ *in probability, where*

$$\mathbf{U} = \begin{pmatrix} (1-\nu_0^*)\mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{pmatrix};$$

(c) $n^{1/2}\hat{\boldsymbol{\alpha}} = o_p(1)$ *and* $n^{1/2}\hat{\boldsymbol{\beta}} = n^{-1/2}\mathbf{V}^{-1}\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathbf{Q} + o_p(1)$.

Proof. First, we consider Part (a). With $\ell_1(\boldsymbol{\theta})$ defined in (3.5), it can be verified that for $r = 1, \ldots, m$,

$$\frac{\partial \ell_1(\mathbf{0})}{\partial \alpha_r} = 0 \quad \text{and} \quad \frac{\partial \ell_1(\mathbf{0})}{\partial \boldsymbol{\beta}_r} = \sum_{j=1}^{n_{r1}} \mathbf{q}(x_{rj}) - \rho_r \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} \mathbf{q}(x_{ij}).$$

We emphasize that in the above segment of the score function, the ρ_r and n_{i1}'s are random variables which prevent us from applying standard large sample theories directly. We use indicator variables to circumvent this technical difficulty. Note that Conditions C1 and C2 imply that $\rho_r \to \rho_r^*$ almost surely under H_0. Then, for $r = 1, \ldots, m$,

$$\frac{\partial \ell_1(\mathbf{0})}{\partial \boldsymbol{\beta}_r} = \sum_{j=1}^{n_r} \mathbf{q}(x_{rj})I(x_{rj} > 0) - \rho_r^* \sum_{i=0}^{m}\sum_{j=1}^{n_i} \mathbf{q}(x_{ij})I(x_{ij} > 0) + o_p(n^{1/2}).$$

Therefore, we have an expression of the score function

$$\frac{\partial \ell_1(\mathbf{0})}{\partial \boldsymbol{\alpha}} = \mathbf{0} \quad \text{and} \quad \frac{\partial \ell_1(\mathbf{0})}{\partial \boldsymbol{\beta}} = \{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathbf{Q} + o_p(n^{1/2}).$$

After some algebra, we can verify that

$$E\Big[\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathbf{Q}\Big] = \mathbf{0}$$

and

$$\mathrm{Var}\Big[\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathbf{Q}\Big] = n\mathbf{H} \otimes \mathrm{Var}\left\{\mathbf{q}(X)I(X > 0)\right\} = n\mathbf{V}.$$

Note that by using indicator variables, elements in $\mathbf{Q}$ are written as the summations of independent random vectors. Therefore, applying the standard central limit theorem, as the total sample size $n \to \infty$, we have

$$n^{-1/2}\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathbf{Q} \xrightarrow{d} N(\mathbf{0}, \mathbf{V}).$$

Using Slutsky's theorem, we further have

$$n^{-1/2}\frac{\partial \ell_1(\mathbf{0})}{\partial \boldsymbol{\beta}} \xrightarrow{d} N(\mathbf{0}, \mathbf{V}).$$

This finishes Part (a).

Next, we consider Part (b). It can be verified that, for $1 \le r, s \le m$,

$$\begin{aligned}
\frac{\partial^2 \ell_1(\mathbf{0})}{\partial \alpha_r \partial \alpha_s} &= -(\delta_{rs}\rho_r - \rho_r\rho_s)n_{\cdot 1}, \\
\frac{\partial^2 \ell_1(\mathbf{0})}{\partial \alpha_r \partial \boldsymbol{\beta}_s^\top} &= -(\delta_{rs}\rho_r - \rho_r\rho_s)\sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} \mathbf{q}^\top(x_{ij}), \\
\frac{\partial^2 \ell_1(\mathbf{0})}{\partial \boldsymbol{\beta}_r \partial \boldsymbol{\beta}_s^\top} &= -(\delta_{rs}\rho_r - \rho_r\rho_s)\sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} \mathbf{q}(x_{ij})\mathbf{q}^\top(x_{ij}),
\end{aligned}$$

where δ_{rs} is 1 when $r = s$ and 0 otherwise.

By noting that $\rho_r = \rho_r^* + o_p(1)$ and using indicator variables, we can write the above second derivatives as sums over a constant range of all n observations. As an illustration, we consider

$$\begin{aligned}
\frac{\partial^2 \ell_1(\mathbf{0})}{\partial \boldsymbol{\beta}_r \partial \boldsymbol{\beta}_s^\top} &= -(\delta_{rs}\rho_r - \rho_r\rho_s)\sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} \mathbf{q}(x_{ij})\mathbf{q}^\top(x_{ij}) \\
&= -(\delta_{rs}\rho_r^* - \rho_r^*\rho_s^*)\sum_{i=0}^{m}\sum_{j=1}^{n_{i}} \mathbf{q}(x_{ij})\mathbf{q}^\top(x_{ij})I(x_{ij} > 0) + o_p(n).
\end{aligned}$$

Using the Kronecker product $\otimes$, $\partial^2 \ell_1(\mathbf{0})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ becomes

$$\frac{\partial^2 \ell_1(\mathbf{0})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = -\mathbf{H} \otimes \begin{pmatrix} \sum_{i=0}^{m}\sum_{j=1}^{n_i} I(x_{ij} > 0) & \sum_{i=0}^{m}\sum_{j=1}^{n_i} \mathbf{q}^\top(x_{ij})I(x_{ij} > 0) \\ \sum_{i=0}^{m}\sum_{j=1}^{n_i} \mathbf{q}(x_{ij})I(x_{ij} > 0) & \sum_{i=0}^{m}\sum_{j=1}^{n_i} \mathbf{q}(x_{ij})\mathbf{q}^\top(x_{ij})I(x_{ij} > 0) \end{pmatrix} + o_p(n).$$

By applying the standard weak law of large numbers, as the total sample size $n \to \infty$,

$$-\frac{1}{n}\frac{\partial^2 \ell_1(\mathbf{0})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \to \mathbf{U} = \begin{pmatrix} (1-\nu_0^*)\mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{pmatrix}$$

in probability, where we have used the fact that $E\{\mathbf{q}(X)I(X>0)\} = \mathbf{0}$. This finishes Part (b).

Last, we consider Part (c). Note that $\hat{\boldsymbol{\theta}}$ satisfies that

$$\partial \ell_1(\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} = \mathbf{0}.$$

Conditions C1–C4 ensures that the matrix $\mathbf{U}$ is positive definite and hence we can apply first order Taylor expansion to $\partial \ell_1(\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}$ to obtain the approximation of $\hat{\boldsymbol{\theta}}$ as follows:

$$\mathbf{0} = \frac{\partial \ell_1(\mathbf{0})}{\partial \boldsymbol{\theta}} + \left(\frac{\partial^2 \ell_1(\mathbf{0})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right) \hat{\boldsymbol{\theta}} + o_p(n^{1/2}).$$

With Parts (a) and (b), we further have

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \mathbf{U}^{-1} \begin{pmatrix} \mathbf{0} \\ \{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathbf{Q} \end{pmatrix} + o_p(n^{-1/2}) \\ &= \begin{pmatrix} (1-\nu_0^*)\mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{0} \\ \{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathbf{Q} \end{pmatrix} + o_p(n^{-1/2}), \end{aligned}$$

which implies Part (c). □

We now move back to show the limiting distribution of $R_{n,2} = 2\ell_1(\hat{\boldsymbol{\theta}})$ under H_0. Since $\ell_1(\mathbf{0}) = 0$ and $\mathbf{U}$ is positive definite, the second-order Taylor expansion can be used to approximate $R_{n,2}$ as

$$R_{n,2} = 2\frac{\partial \ell_1(\mathbf{0})}{\partial \boldsymbol{\theta}^\top}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}}^\top \left(\frac{\partial^2 \ell_1(\mathbf{0})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right) \hat{\boldsymbol{\theta}} + o_p(1).$$

With Lemma 4.1, the approximation of $R_{n,2}$ is simplified to

$$R_{n,2} = \frac{1}{n}\big[\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathbf{Q}\big]^\top \mathbf{V}^{-1} \big[\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathbf{Q}\big] + o_p(1), \qquad (3.7)$$

which converges in distribution to χ^2_{md}. Note here $\text{rank}(\mathbf{V}^{-1}) = \dim(\boldsymbol{\beta}) = md$. This finishes Step 2.

The remaining step is to show $R_{n,1}$ and $R_{n,2}$ are asymptotically independent. In light of Lemma 4.1 and asymptotic expansions (3.6) and (3.7), it suffices to show that

$$\mathrm{Cov}\left[\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathbf{Q}, \frac{\partial \ell_0(\boldsymbol{\nu}^*)}{\partial \boldsymbol{\nu}}\right] = \{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^\top) \otimes \mathbf{I}_d\}\mathrm{Cov}\left\{\mathbf{Q}, \frac{\partial \ell_0(\boldsymbol{\nu}^*)}{\partial \boldsymbol{\nu}}\right\} = \mathbf{0}.$$

Recall that $\mathbf{Q}$ can be partitioned into $m+1$ blocks with each block being a d-dimensional vector. For any $0 \le r, s \le m$, the covariance between the $(r+1)$th block of $\mathbf{Q}$ and $(s+1)$th element of $\partial \ell_0(\boldsymbol{\nu}^*)/\partial \boldsymbol{\nu}$ is $\mathrm{Cov}\left\{\sum_{j=1}^{n_r} \mathbf{q}(x_{rj}) I(x_{rj} > 0), \partial \ell_0(\boldsymbol{\nu}^*)/\partial \nu_s\right\}$ which is $\mathbf{0}$ when $r \neq s$ since the observations for the rth and sth groups are independent. When $r = s$,

$$\begin{aligned}
& \mathrm{Cov}\left\{\sum_{j=1}^{n_r} \mathbf{q}(x_{rj}) I(x_{rj} > 0), \frac{\partial \ell_0(\boldsymbol{\nu}^*)}{\partial \nu_s}\right\} \\
= & \frac{1}{\nu_0^*(1-\nu_0^*)}\mathrm{Cov}\left\{\sum_{j=1}^{n_r} \mathbf{q}(x_{rj}) I(x_{rj} > 0), n_{r0} - n_r \nu_0^*\right\} \\
= & \frac{1}{\nu_0^*(1-\nu_0^*)}\mathrm{Cov}\left\{\sum_{j=1}^{n_r} \mathbf{q}(x_{rj}) I(x_{rj} > 0), \sum_{j=1}^{n_r} I(x_{rj} = 0)\right\} \\
= & \frac{1}{\nu_0^*(1-\nu_0^*)}\sum_{j=1}^{n_r} \mathrm{Cov}\left\{\mathbf{q}(x_{rj}) I(x_{rj} > 0), I(x_{rj} = 0)\right\} \\
= & \frac{1}{\nu_0^*(1-\nu_0^*)}\sum_{j=1}^{n_r} E\left\{\mathbf{q}(x_{rj}) I(x_{rj} > 0) I(x_{rj} = 0)\right\} = \mathbf{0}.
\end{aligned}$$

Hence $R_{n,1}$ and $R_{n,2}$ are asymptotically independent, and the null limiting distribution of $R_n = R_{n,1} + R_{n,2}$ is $\chi^2_{m(d+1)}$. This completes the proof of Theorem 3.2. □

3.5.2 Proof of Theorem 3.2

The regularity conditions provided in Section 3.2.3 will be needed throughout this proof.

We first define some notation. Recall that $\mathbf{X}^\omega = \{x_{i1}^\omega, \ldots, x_{in_i}^\omega : i = 0, \ldots, m\}$ is the nonparametric bootstrap sample from $\mathbf{X} = \{x_{i1}, \ldots, x_{in_i} : i = 0, \ldots, m\}$. That is,

$$\mathbf{X}^\omega | \mathbf{X} \sim \hat{F}_n(x) = \frac{1}{n}\sum_{i=0}^{m}\sum_{j=1}^{n_i} I(x_{ij} \le x).$$

For a given bootstrap sample $\mathbf{X}^\omega$, let n_{i0}^ω and n_{i1}^ω be the numbers of zeros and positive observations for the ith group, respectively, for $i = 0, \ldots, m$. Without loss of generality, we use $\left\{x_{i1}^\omega, \ldots, x_{in_{i1}^\omega}^\omega : i = 0, \ldots, m\right\}$ to denote the positive observations in $\mathbf{X}^\omega$.

Recall the forms of $\hat{\boldsymbol{\nu}}$ and $\mathbf{Q}$ defined in the proof of Theorem 3.1. The bootstrap version of $\hat{\boldsymbol{\nu}}$ is defined as $\hat{\boldsymbol{\nu}}^\omega = (\hat{\nu}_0^\omega, \ldots, \hat{\nu}_m^\omega)^\top$ with $\hat{\nu}_i^\omega = n_{i0}^\omega / n_i$ for $i = 0, \ldots, m$. The bootstrap version of $\mathbf{Q}$ is defined as

$$\mathbf{Q}^\omega = \left\{ \sum_{j=1}^{n_0} \mathbf{q}^\top(x_{0j}^\omega) I(x_{0j}^\omega > 0), \ldots, \sum_{j=1}^{n_m} \mathbf{q}^\top(x_{mj}^\omega) I(x_{mj}^\omega > 0) \right\}^\top.$$

Since $\mathbf{X}^\omega | \mathbf{X} \sim \hat{F}_n(x)$, it is easy to verify that

$$E\left(\hat{\boldsymbol{\nu}}^\omega | \mathbf{X}\right) = (\bar{v}_0, \ldots, \bar{v}_0)^\top \quad \text{and} \quad E(\mathbf{Q}^\omega | \mathbf{X}) = (n_0 \bar{\mathbf{q}}^\top, \ldots, n_m \bar{\mathbf{q}}^\top)^\top := \bar{\mathbf{Q}},$$

where $\bar{v}_0 = \sum_{i=0}^m n_{i0}/n$ and $\bar{\mathbf{q}} = \frac{1}{n} \sum_{i=0}^m \sum_{j=1}^{n_i} \mathbf{q}(x_{ij}) I(x_{ij} > 0)$.

The proof of Theorem 3.2 depends on the asymptotic properties of $\hat{\boldsymbol{\nu}}^\omega$ and $\mathbf{Q}^\omega$ conditional on the observed data $\mathbf{X}$. As a preparation, we study them first. Let

$$\mathbf{Z}_n^\omega = \begin{pmatrix} n_{00}^\omega - n_0 \bar{\nu}_0 \\ \vdots \\ n_{m0}^\omega - n_m \bar{\nu}_0 \\ \mathbf{Q}^\omega - \bar{\mathbf{Q}} \end{pmatrix}.$$

Let $E_{\bar{F}}$ and $\mathrm{Var}_{\bar{F}}$ represent the expectation and variance calculated under $\bar{F}$. Also let $X \sim \bar{F}(x)$. Without loss of generality, we assume that $E_{\bar{F}}\{\mathbf{q}(X) I(X > 0)\} = \mathbf{0}$.

Lemma 3.2 *Assume the same conditions as Theorem 3.2. We have, as $n \to \infty$*

(a) conditional on the observed data $\mathbf{X}$,

$$\frac{1}{\sqrt{n}} \mathbf{Z}_n^\omega \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega})$$

in probability, where $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{22} \end{pmatrix}$ *and*

$$\boldsymbol{\Omega}_{11} = diag\Big\{\rho_0^* \bar{\nu}_0^* (1 - \bar{\nu}_0^*), \ldots, \rho_m^* \bar{\nu}_0^* (1 - \bar{\nu}_0^*)\Big\},$$
$$\boldsymbol{\Omega}_{22} = diag\Big[\rho_0^* \, Var_{\bar{F}}\left\{\mathbf{q}(X) I(X > 0)\right\}, \ldots, \rho_m^* \, Var_{\bar{F}}\left\{\mathbf{q}(X) I(X > 0)\right\}\Big];$$

(b) $\hat{\boldsymbol{\nu}}^\omega - \bar{\boldsymbol{\nu}}^* = O_p(n^{-1/2})$ *and* $\mathbf{Q}^\omega = O_p(n^{1/2})$*, where* $\bar{\boldsymbol{\nu}}^* = (\bar{v}_0^*, \ldots, \bar{v}_0^*)^\top$.

Proof. First, we consider Part (a). Note that $n_{i0}^\omega = \sum_{j=0}^{n_i} I(x_{ij}^\omega = 0)$. Hence conditional on the observed data $\mathbf{X}$, each row of $\mathbf{Z}_n^\omega$ is sum of independently and identically distributed observations. Then by the fact that $\mathbf{X}^\omega | \mathbf{X} \sim \hat{F}_n(x)$, we have

$$E\left(\frac{1}{\sqrt{n}}\mathbf{Z}_n^\omega | \mathbf{X}\right) = \mathbf{0}, \quad \mathrm{Var}\left(\frac{1}{\sqrt{n}}\mathbf{Z}_n^\omega | \mathbf{X}\right) = \mathbf{\Omega}_n = \begin{pmatrix} \mathbf{\Omega}_{n,11} & \mathbf{\Omega}_{n,12} \\ \mathbf{\Omega}_{n,21} & \mathbf{\Omega}_{n,22} \end{pmatrix},$$

where

$$\begin{aligned}
\mathbf{\Omega}_{n,11} &= \mathrm{diag}\left\{\rho_0^* \bar{\nu}_0(1-\bar{\nu}_0), \ldots, \rho_m^* \bar{\nu}_0(1-\bar{\nu}_0)\right\}, \\
\mathbf{\Omega}_{n,22} &= \mathrm{diag}\left[\rho_0^*\left\{\frac{1}{n}\sum_{i,j}\mathbf{q}(x_{ij})\mathbf{q}^\top(x_{ij})I(x_{ij}>0) - \bar{\mathbf{q}}\bar{\mathbf{q}}^\top\right\},\right. \\
&\qquad \left. \ldots, \rho_m^*\left\{\frac{1}{n}\sum_{i,j}\mathbf{q}(x_{ij})\mathbf{q}^\top(x_{ij})I(x_{ij}>0) - \bar{\mathbf{q}}\bar{\mathbf{q}}^\top\right\}\right], \\
\mathbf{\Omega}_{n,21} &= \mathbf{\Omega}_{n,12}^\top = \mathrm{diag}(-\rho_0^*\bar{\nu}_0\bar{\mathbf{q}}, \ldots, -\rho_m^*\bar{\nu}_0\bar{\mathbf{q}}).
\end{aligned}$$

After some calculus, it can be verified that $E(\bar{\mathbf{q}}) = E_{\bar{F}}\{\mathbf{q}(X)I(X>0)\} = \mathbf{0}$. Hence by the weak law of large numbers, we have $\mathbf{\Omega}_n \to \mathbf{\Omega}$ in probability. By Conditions D1–D4, $\mathbf{\Omega}$ is positive definite.

By Berry-Esseén inequality (Shao and Tu, 1995, Section 3.1, p. 74) together with Cramér-Wold theorem (Serfling, 1980, p. 17), or the results in Janssen and Pauls (2003), we get that, conditional on $\mathbf{X}$,

$$\mathbf{\Omega}_n^{-1/2}\frac{1}{\sqrt{n}}\mathbf{Z}_n^\omega \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$$

in probability. Note that $\mathbf{\Omega}_n \to \mathbf{\Omega}$ in probability implies that $\mathbf{\Omega}_n^{-1/2} \to \mathbf{\Omega}^{-1/2}$ in probability. By conditional Slutsky's theorem (Cheng, 2015), conditional on the observed data $\mathbf{X}$,

$$\frac{1}{\sqrt{n}}\mathbf{Z}_n^\omega \xrightarrow{d} N(\mathbf{0}, \mathbf{\Omega})$$

in probability. This finishes the proof of Part (a).

Next, we consider Part (b). Since $\mathbf{\Omega}$ is positive definite, we have $n_{i0}^{\omega} - n_i\bar{\nu}_0 = O_p(\sqrt{n})$ and $\mathbf{Q}^{\omega} - \bar{\mathbf{Q}} = O_p(\sqrt{n})$.

It is easy to verify that $E(\bar{\nu}_0) = \bar{\nu}_0^*$ and $\bar{\mathbf{Q}} = O_p(n^{1/2})$ since $E_{\bar{F}}\{\mathbf{q}(X)I(X > 0)\} = \mathbf{0}$. Hence by the classic central limit theorem, we obtain that $\bar{\nu}_0 - \bar{\nu}_0^* = O_p(n^{-1/2})$ and $\bar{\mathbf{Q}} = O_p(n^{1/2})$.

Combining the above results, we get that

$$\hat{\nu}_i^{\omega} - \nu_0^* = O_p(n^{-1/2}), \quad \text{and} \quad \mathbf{Q}^{\omega} = O_p(n^{1/2}).$$

This completes the proof of Part (b). □

We now move back to the proof of Theorem 3.2. For a given bootstrap sample $\mathbf{X}^{\omega}$, we define the bootstrap versions of $\ell_0(\boldsymbol{\nu})$ and $\ell_1(\boldsymbol{\theta})$ as

$$\ell_0^{\omega}(\boldsymbol{\nu}) = \sum_{i=0}^{m} \log\{\nu_i^{n_{i0}^{\omega}}(1-\nu_i)^{n_{i1}^{\omega}}\}$$

and

$$\ell_1^{\omega}(\boldsymbol{\theta}) = -\sum_{i=0}^{m}\sum_{j=1}^{n_{i1}^{\omega}} \log\left[\rho_0^{\omega} + \sum_{r=1}^{m}\rho_r^{\omega}\exp\{\alpha_r + \boldsymbol{\beta}_r^{\top}\mathbf{q}(x_{ij}^{\omega})\}\right] + \sum_{i=1}^{m}\sum_{j=1}^{n_{i1}^{\omega}}\{\alpha_i + \boldsymbol{\beta}_i^{\top}\mathbf{q}(x_{ij}^{\omega})\},$$

where $\rho_r^{\omega} = n_{r1}^{\omega}/n_{\cdot 1}^{\omega}$ with $n_{\cdot 1}^{\omega} = \sum_{i=0}^{m} n_{i1}^{\omega}$. Then the bootstrap ELR statistic can be written as $R_n^{\omega} = R_{n,1}^{\omega} + R_{n,2}^{\omega}$ with

$$R_{n,1}^{\omega} = 2\left\{\sup_{\boldsymbol{\nu}}\ell_0^{\omega}(\boldsymbol{\nu}) - \sup_{H_0}\ell_0^{\omega}(\boldsymbol{\nu})\right\}, \quad R_{n,2}^{\omega} = 2\left\{\sup_{\boldsymbol{\theta}}\ell_1^{\omega}(\boldsymbol{\theta}) - \sup_{H_0}\ell_1^{\omega}(\boldsymbol{\theta})\right\}.$$

With the results in Lemma 3.2, we are able to find quadratic approximations of $R_{n,1}^{\omega}$ and $R_{n,2}^{\omega}$.

First, we consider $R_{n,1}^{\omega}$. Note that $\hat{\boldsymbol{\nu}}^{\omega}$ maximizes $\ell_0^{\omega}(\boldsymbol{\nu})$. Then $R_{n,1}^{\omega} = 2\big\{\ell_0^{\omega}(\hat{\boldsymbol{\nu}}^{\omega}) - \sup_{H_0}\ell_1^{\omega}(\boldsymbol{\theta})\big\}$. By the second order Taylor expansion and Lemma 3.2, we have

$$\ell_0^{\omega}(\hat{\boldsymbol{\nu}}^{\omega}) - \ell_0^{\omega}(\bar{\boldsymbol{\nu}}^*) = \frac{n}{2}\sum_{i=0}^{m}\rho_i^*\frac{1}{\bar{\nu}_0^*(1-\bar{\nu}_0^*)}(\hat{\nu}_i^{\omega} - \bar{\nu}_0^*)^2 + o_p(1). \tag{3.8}$$

Similarly, we have

$$\sup_{H_0} \ell_0^\omega(\boldsymbol{\nu}) - \ell_0^\omega(\bar{\boldsymbol{\nu}}^*) = \frac{n}{2}\sum_{i=0}^m \rho_i^* \frac{1}{\bar{\nu}_0^*(1-\bar{\nu}_0^*)}(\bar{\nu}_0^\omega - \bar{\nu}_0^*)^2 + o_p(1), \tag{3.9}$$

where $\bar{\nu}_0^\omega = \sum_{i=0}^m n_{i0}^\omega / n$. Combining (3.8) and (3.9) gives

$$\begin{aligned} R_{n,1}^\omega &= n\sum_{i=0}^m \rho_i^* \frac{1}{\bar{\nu}_0^*(1-\bar{\nu}_0^*)}(\hat{\nu}_i^\omega - \bar{\nu}_0^\omega)^2 + o_p(1) \\ &= \frac{n}{\bar{\nu}_0^*(1-\bar{\nu}_0^*)}\sum_{i=0}^m \rho_i^* \{\hat{\nu}_i^\omega - \bar{\nu}_0 - (\bar{\nu}_0^\omega - \bar{\nu}_0)\}^2 + o_p(1). \end{aligned}$$

Let $\mathbf{S}_n^\omega = (S_{n0}^\omega, \ldots, S_{nm}^\omega)^\top$ with $S_{ni}^\omega = \hat{\nu}_i^\omega - \bar{\nu}_0$ for $i = 0, \ldots, m$. Then we can have a compact form for $R_{n,1}^\omega$ as

$$R_{n,1}^\omega = (\sqrt{n}\mathbf{S}_n^\omega)^\top \boldsymbol{\Sigma}^* (\sqrt{n}\mathbf{S}_n^\omega) - \bar{\nu}_0^*(1-\bar{\nu}_0^*)(\sqrt{n}\mathbf{S}_n^\omega)^\top \boldsymbol{\Sigma}^* \mathbf{1}_{m+1}\mathbf{1}_{m+1}^\top \boldsymbol{\Sigma}^* (\sqrt{n}\mathbf{S}_n^\omega) + o_p(1),$$

where $\boldsymbol{\Sigma}^* = \frac{1}{\bar{\nu}_0^*(1-\bar{\nu}_0^*)}\text{diag}(\rho_0^*, \ldots, \rho_m^*)$. Using Part (a) of Lemma 3.2, we have, conditional on $\mathbf{X}$,

$$\sqrt{n}\mathbf{S}_n^\omega \xrightarrow{d} N\left(\mathbf{0}, (\boldsymbol{\Sigma}^*)^{-1}\right)$$

in probability. Finally, we can further check that

$$\begin{aligned} &\left(\boldsymbol{\Sigma}^* - \bar{\nu}_0^*(1-\bar{\nu}_0^*)\boldsymbol{\Sigma}^*\mathbf{1}_{m+1}\mathbf{1}_{m+1}^\top\boldsymbol{\Sigma}^*\right)(\boldsymbol{\Sigma}^*)^{-1}\left(\boldsymbol{\Sigma}^* - \bar{\nu}_0^*(1-\bar{\nu}_0^*)\boldsymbol{\Sigma}^*\mathbf{1}_{m+1}\mathbf{1}_{m+1}^\top\boldsymbol{\Sigma}^*\right) \\ = \;& \boldsymbol{\Sigma}^* - \bar{\nu}_0^*(1-\bar{\nu}_0^*)\boldsymbol{\Sigma}^*\mathbf{1}_{m+1}\mathbf{1}_{m+1}^\top\boldsymbol{\Sigma}^* \end{aligned}$$

and $\text{rank}\left(\boldsymbol{\Sigma}^* - \bar{\nu}_0^*(1-\bar{\nu}_0^*)\boldsymbol{\Sigma}^*\mathbf{1}_{m+1}\mathbf{1}_{m+1}^\top\boldsymbol{\Sigma}^*\right) = m$. Hence, by the conditional continuous mapping theorem (Kosorok, 2008, Theorem 10.8) and conditional Slutsky's theorem (Cheng, 2015), conditional on $\mathbf{X}$, we get that

$$R_{n,1}^\omega \xrightarrow{d} \chi_m^2$$

in probability.

Next, we consider $R_{n,2}^\omega$. Let $\hat{\boldsymbol{\theta}}^\omega = \text{argmax}_{\boldsymbol{\theta}} \ell_1^\omega(\boldsymbol{\theta})$. It can be checked that $\sup_{H_0} \ell_1^\omega(\boldsymbol{\theta}) = 0$. Then $R_{n,2}^\omega = 2\ell_1^\omega(\hat{\boldsymbol{\theta}}^\omega)$. Note that, no matter the original observed data $\mathbf{X}$ is from the null or alternative hypothesis, the bootstrapped observations for all groups are homogeneous. This is because that the bootstrapped observations for all groups come from the same empirical distribution function $\hat{F}_n(x)$. Following Cai et al. (2016), we still have $\hat{\boldsymbol{\theta}}^{\omega} \to \mathbf{0}$ in probability and $\hat{\boldsymbol{\theta}}^{\omega} = O_p(n^{-1/2})$.

In the following lemma, we summarize some useful properties of $\partial \ell_1^{\omega}(\mathbf{0})/\partial \boldsymbol{\theta}$, $\partial^2 \ell_1^{\omega}(\mathbf{0})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}$, and $\hat{\boldsymbol{\theta}}^{\omega}$, which are helpful to find the conditional distribution of $R_{n,2}^{\omega}$. The proof is similar to that of Lemma 4.1 and hence is omitted.

Lemma 3.3 *Assume the same conditions as Theorem 3.2. We have, as $n \to \infty$*

(a) $\frac{\partial \ell_1^{\omega}(\mathbf{0})}{\partial \boldsymbol{\alpha}} = \mathbf{0}$, $\frac{\partial \ell_1^{\omega}(\mathbf{0})}{\partial \boldsymbol{\beta}} = \{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho} \mathbf{1}_m^{\top}) \otimes \mathbf{I}_d\} \mathbf{Q}^{\omega} + o_p(n^{1/2})$;

(b) $-\frac{1}{n} \frac{\partial^2 \ell_1^{\omega}(\mathbf{0})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}} \to \bar{\mathbf{U}}$ *in probability, where*

$$\bar{\mathbf{U}} = \begin{pmatrix} (1 - \bar{\nu}_0^*)\mathbf{H} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{V}} \end{pmatrix}$$

and $\bar{\mathbf{V}} = \mathbf{H} \otimes Var_{\bar{F}} \{\mathbf{q}(X) I(X > 0)\}$;

(c) $\sqrt{n}\hat{\boldsymbol{\alpha}}^{\omega} = o_p(1)$ *and* $\sqrt{n}\hat{\boldsymbol{\beta}}^{\omega} = \frac{1}{\sqrt{n}} \bar{\mathbf{V}}^{-1} \{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho} \mathbf{1}_m^{\top}) \otimes \mathbf{I}_d\} \mathbf{Q}^{\omega} + o_p(1)$.

We now move back to $R_{n,2}^{\omega} = 2\ell_1^{\omega}(\hat{\boldsymbol{\theta}}^{\omega})$. By the second order Taylor expansion and Lemma 3.3, we have the following quadratic expansion for $R_{n,2}^{\omega}$:

$$R_{n,2}^{\omega} = \frac{1}{n} \big[\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho} \mathbf{1}_m^{\top}) \otimes \mathbf{I}_d\} \mathbf{Q}^{\omega}\big]^{\top} \bar{\mathbf{V}}^{-1} \big[\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho} \mathbf{1}_m^{\top}) \otimes \mathbf{I}_d\} \mathbf{Q}^{\omega}\big] + o_p(1).$$

After some algebra work, it can be verified that

$$\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho} \mathbf{1}_m^{\top}) \otimes \mathbf{I}_d\} \bar{\mathbf{Q}} = \mathbf{0}.$$

Hence

$$R_{n,2}^{\omega} = \frac{1}{n} \big[\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho} \mathbf{1}_m^{\top}) \otimes \mathbf{I}_d\} (\mathbf{Q}^{\omega} - \bar{\mathbf{Q}})\big]^{\top} \bar{\mathbf{V}}^{-1} \big[\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho} \mathbf{1}_m^{\top}) \otimes \mathbf{I}_d\} (\mathbf{Q}^{\omega} - \bar{\mathbf{Q}})\big] + o_p(1).$$

Recall that Part (a) of Lemma 3.2 implies that conditional on $\mathbf{X}$,

$$\frac{1}{\sqrt{n}}(\mathbf{Q}^{\omega} - \bar{\mathbf{Q}}) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Omega}_{22})$$

in probability and

$$\frac{1}{\sqrt{n}}\{(-\boldsymbol{\rho}, \mathbf{I}_m - \boldsymbol{\rho}\mathbf{1}_m^{\top}) \otimes \mathbf{I}_d\}(\mathbf{Q}^{\omega} - \bar{\mathbf{Q}}) \xrightarrow{d} N(\mathbf{0}, \bar{\mathbf{V}})$$

in probability. Note $\mathrm{rank}(\bar{\mathbf{V}}^{-1}) = \dim(\boldsymbol{\beta}) = md$. By conditional Slutsky's theorem (Cheng, 2015), we conclude that conditional on $\mathbf{X}$,

$$R_{n,2}^{\omega} \xrightarrow{d} \chi^2_{md}$$

in probability.

Lastly, Part (a) of Lemma 3.2 implies that conditional on $\mathbf{X}$, $R_{n,1}^{\omega}$ and $R_{n,2}^{\omega}$ are asymptotically independent. Hence, conditional on the data $\mathbf{X}$,

$$R_n^{\omega} \xrightarrow{d} \chi^2_{m(d+1)}$$

in probability. As a result, under both the null and alternative hypotheses,

$$\sup_x \left|\Pr(R_n^{\omega} \leq x|\mathbf{X}) - \Pr(\chi^2_{m(d+1)} \leq x)\right| \to 0$$

in probability. This completes the proof of Theorem 3.2. □

3.6 Additional simulation results

In this section, we add more simulations to assess the finite-sample performance of the proposed ELR test and the nonparametric bootstrap procedure. The simulation settings stay unchanged as in Section 3.3. We consider more competitors and more choices of basis functions.

To examine how sensitive the power of the ELR test is to the choice of user-specified basis function $\mathbf{q}(x)$ in a DRM, we consider the ELR defined under five particular basis functions:

- ELR1: $\mathbf{q}(x) = \{x, \log(x)\}^\top$, which is, in general, correctly specified under GAM models but is misspecified under LN models;
- ELR2: $\mathbf{q}(x) = \{\log(x), \log^2(x)\}^\top$, which is, in general, correctly specified under LN models but is misspecified under GAM models;
- ELR3: $\mathbf{q}(x) = \{x, \log(x), \log^2(x)\}^\top$, which covers both distribution families, but perhaps at the price of over-fitting;
- ELR4: $\mathbf{q}(x) = x$, which is correctly specified when the parameters a_i's in GAM models are set to be equal, but otherwise, is under-fitting under other GAM models and is misspecified under LN models;
- ELR5: $\mathbf{q}(x) = \log(x)$, which is correctly specified when the parameters b_i's in both LN and GAM models are set to be equal, but otherwise, is under-fitting under other LN and GAM models.

Same as before, the type I error rates and powers at the 5% significance level are calculated based on 10,000 repetitions. The type I error rates are calculated based on the limiting distribution as well as adjusted by the nonparametric bootstrap procedure with $B = 999$. The results are summarized in Tables 3.7–3.10.

As suggested in Pauly et al. (2015), we also use the permutation method with 10,000 permutation samples (the default number in "`GFD`") to calculate the type I error rates and powers of WTS1 and WTS2. The results along with those from the bootstrap method are presented in Table 3.11. As we can see, the results from two resampling methods are quite consistent.

Table 3.7: Simulated probabilities (%) of rejecting H_0 at significance level 0.05 when data are generated from a mixture model with parameter settings given in Table 3.1 and $(n_0, n_1, n_2) = (20, 20, 20)$. The abbreviations "asym." and "boot." denote the asymptotic distribution and bootstrap procedure, respectively, that are used in calculation.

Model		$(n_0, n_1, n_2) = (20, 20, 20)$													
		ELR1	ELR2	ELR3	ELR4	ELR5	KW2	KW1	ANOVA2	ANOVA1	LRT-LN	LRT-GAM	ATS	WTS1	WTS2
LN_1	asym.	8.67	8.12	11.43	7.41	6.56	5.73	4.78	5.98	4.24	7.03	16.12	2.29	6.28	7.33
	boot.	4.86	4.87	4.94	4.93	4.74	4.81	4.96	4.84	4.40	4.12	4.11	4.40	4.39	4.66
LN_2	asym.	8.85	8.58	12.76	7.11	6.31	4.98	4.67	5.27	3.57	7.35	15.46	2.26	6.04	7.00
	boot.	4.84	4.78	5.11	4.94	4.85	4.87	4.82	4.51	4.32	3.99	3.97	4.32	4.98	4.64
LN_3	asym.	12.47	12.54	19.75	7.16	6.70	3.26	2.95	5.07	2.72	10.02	15.22	1.41	4.92	8.80
	boot.	5.75	5.66	6.95	5.67	5.70	5.79	5.83	5.13	5.19	3.47	3.82	5.19	5.78	5.79
LN_4	boot.	13.37	13.11	11.71	15.60	15.49	16.78	15.04	16.24	8.09	12.42	10.56	8.09	9.33	14.93
LN_5	boot.	16.84	16.18	15.31	17.76	21.72	19.20	9.73	11.79	13.56	16.70	14.22	13.56	13.83	11.86
LN_6	boot.	11.45	11.67	8.94	5.84	5.26	5.11	4.44	5.35	5.35	7.94	7.47	5.35	5.87	4.73
LN_7	boot.	27.04	27.10	19.45	17.25	24.26	25.94	11.92	17.86	4.57	25.36	17.03	4.57	4.30	16.64
LN_8	boot.	18.30	18.63	15.31	10.21	15.39	12.77	3.96	6.33	6.32	16.72	10.92	6.32	9.87	8.96
LN_9	boot.	13.63	13.32	11.92	16.05	17.02	18.03	11.32	17.30	4.28	12.78	10.30	4.28	4.63	14.21
LN_{10}	boot.	12.99	12.75	11.12	14.68	14.28	15.84	16.86	15.53	4.89	11.35	9.11	4.89	7.48	14.08
LN_{11}	boot.	25.53	25.38	20.67	20.01	28.89	29.19	8.62	19.97	4.35	25.03	16.65	4.35	4.62	16.74
LN_{12}	boot.	15.74	15.93	15.18	13.02	18.92	16.26	7.58	8.86	9.01	15.59	10.53	9.01	10.43	9.87
LN_{13}	boot.	12.23	12.03	10.27	14.17	14.90	16.18	13.32	15.60	5.21	11.14	8.63	5.21	5.73	13.24
LN_{14}	boot.	12.21	11.80	10.28	15.02	15.21	16.67	16.26	15.76	5.82	11.27	8.69	5.82	6.83	14.03
LN_{15}	boot.	25.44	25.19	20.99	23.45	31.87	31.43	9.06	22.70	4.78	24.92	17.94	4.78	4.82	18.41
GAM_1	asym.	7.97	8.53	11.00	7.54	7.02	6.13	4.98	6.59	5.31	16.48	7.06	3.78	7.17	8.04
	boot.	4.63	4.59	5.05	5.18	4.93	5.16	5.20	5.09	5.13	4.12	4.39	5.13	5.14	5.18
GAM_2	asym.	8.91	9.46	12.66	7.03	6.67	5.29	5.15	6.19	4.46	15.54	7.37	3.29	7.10	7.99
	boot.	4.99	5.28	5.68	5.17	5.00	5.14	5.37	5.08	4.99	4.31	4.32	4.99	5.36	5.03
GAM_3	asym.	11.95	12.03	19.53	6.72	6.34	3.10	2.86	5.24	3.04	16.25	10.01	1.71	5.94	9.68
	boot.	5.75	5.72	6.55	5.32	5.28	5.60	5.69	5.02	5.16	3.27	3.66	5.16	6.06	5.10
GAM_4	boot.	10.69	10.88	9.83	13.06	12.23	13.95	13.91	13.70	8.18	8.75	10.39	8.18	9.70	11.84
GAM_5	boot.	19.72	18.72	18.64	24.40	27.88	23.40	8.04	22.60	16.08	19.08	18.32	16.08	15.46	22.20
GAM_6	boot.	21.32	20.20	19.04	30.36	24.68	24.20	8.28	24.72	20.52	16.36	20.04	20.52	19.36	25.88
GAM_7	boot.	24.15	24.26	15.71	12.93	15.67	16.02	12.51	12.97	8.93	21.77	21.52	8.93	8.03	12.39
GAM_8	boot.	17.08	16.60	11.96	5.84	8.48	7.40	6.52	5.52	6.20	15.72	14.24	6.20	6.24	6.40
GAM_9	boot.	19.83	19.58	17.49	24.09	27.23	26.28	9.48	25.55	5.03	18.75	19.85	5.03	5.13	20.73
GAM_{10}	boot.	15.81	15.61	14.51	19.35	18.45	18.93	12.78	20.04	5.76	11.76	14.10	5.76	5.65	15.14
GAM_{11}	boot.	28.01	27.90	23.57	30.23	34.99	33.10	7.68	30.98	5.06	28.84	27.57	5.06	5.36	27.26
GAM_{12}	boot.	26.40	26.60	20.72	15.44	24.80	19.76	10.88	14.36	12.32	30.88	26.52	12.32	12.16	16.00
GAM_{13}	boot.	12.41	12.16	10.46	14.95	15.05	16.01	13.33	15.13	8.59	10.69	11.63	8.59	8.19	13.54
GAM_{14}	boot.	11.64	11.51	10.55	14.06	14.10	15.15	13.87	14.51	7.54	8.95	10.52	7.54	8.27	11.91
GAM_{15}	boot.	20.52	20.64	16.22	18.97	21.84	22.08	9.57	19.08	6.09	19.34	18.65	6.09	6.67	17.11

* NOTE: the Monte Carlo error is 0.218 (%) under the null models LN_1-LN_3 and GAM_1-GAM_3.

Table 3.8: Simulated probabilities (%) of rejecting H_0 at significance level 0.05 when data are generated from a mixture model with parameter settings given in Table 3.1 and $(n_0, n_1, n_2) = (50, 50, 50)$. The abbreviations "asym." and "boot." denote the asymptotic distribution and bootstrap procedure, respectively, that are used in calculation.

Model		$(n_0, n_1, n_2) = (50, 50, 50)$													
		ELR1	ELR2	ELR3	ELR4	ELR5	KW2	KW1	ANOVA2	ANOVA1	LRT-LN	LRT-GAM	ATS	WTS1	WTS2
LN_1	asym.	6.54	6.12	7.76	5.68	5.30	5.10	4.70	4.79	3.71	5.79	17.17	2.87	5.33	5.54
	boot.	4.91	4.89	5.12	4.84	4.82	4.88	4.66	4.61	4.27	4.81	4.43	4.27	4.58	4.64
LN_2	asym.	7.30	6.80	8.61	6.38	5.61	5.11	5.21	5.03	4.00	5.91	17.71	2.88	6.03	6.40
	boot.	5.20	5.03	5.35	5.23	5.11	4.99	5.30	5.04	4.64	4.75	4.81	4.64	4.97	5.14
LN_3	asym.	8.22	7.94	11.17	6.78	5.85	4.96	4.81	5.24	3.48	6.61	15.87	2.22	5.64	7.00
	boot.	4.95	4.81	5.21	4.95	4.98	4.81	5.10	4.83	4.46	4.36	4.30	4.46	4.91	4.65
LN_4	boot.	30.92	31.64	26.51	37.01	37.44	38.17	32.38	38.35	12.49	31.56	22.84	12.49	12.95	36.45
LN_5	boot.	46.68	47.21	42.03	38.38	55.90	52.77	18.52	26.44	32.33	47.95	33.48	32.33	33.17	28.58
LN_6	boot.	29.57	31.80	23.79	10.93	5.59	5.48	4.98	8.93	9.59	30.09	25.47	9.59	7.96	7.29
LN_7	boot.	75.05	77.55	68.25	34.90	63.56	64.08	21.11	36.43	4.72	78.32	44.61	4.72	4.92	35.84
LN_8	boot.	55.57	58.62	49.54	11.45	37.49	34.98	8.05	8.18	7.47	60.74	26.03	7.47	10.84	11.77
LN_9	boot.	32.30	32.46	28.42	36.08	39.79	40.01	21.20	38.29	4.46	33.01	20.97	4.46	5.31	34.82
LN_{10}	boot.	29.90	30.28	26.21	33.05	32.25	33.11	36.13	36.23	5.55	29.92	19.96	5.55	7.50	31.54
LN_{11}	boot.	66.19	69.06	60.39	40.45	68.02	67.29	13.65	41.81	4.54	69.47	36.35	4.54	5.23	39.56
LN_{12}	boot.	43.82	46.36	40.58	19.68	46.95	43.90	11.79	14.03	15.88	47.83	20.62	15.88	17.66	16.81
LN_{13}	boot.	28.39	28.41	24.66	33.19	34.70	35.18	27.42	35.70	5.95	28.69	17.69	5.95	6.99	33.10
LN_{14}	boot.	28.02	27.94	24.18	33.39	33.81	34.53	34.44	35.51	7.43	28.13	17.04	7.43	9.19	33.67
LN_{15}	boot.	65.33	67.27	59.70	47.24	71.53	70.16	12.82	47.50	5.84	68.30	40.11	5.84	6.23	44.32
GAM_1	asym.	5.98	6.44	7.37	5.61	5.54	5.09	4.85	5.26	4.72	16.99	5.38	4.19	5.75	6.18
	boot.	4.53	4.57	4.68	4.92	4.84	4.97	4.89	4.78	4.85	4.56	4.38	4.85	4.72	4.98
GAM_2	asym.	6.22	6.65	8.53	5.48	5.37	4.90	5.11	5.04	4.87	16.63	5.33	4.13	6.11	5.82
	boot.	4.59	5.03	5.23	4.59	4.75	4.81	5.10	4.65	5.05	4.64	4.32	5.05	5.02	4.72
GAM_3	asym.	7.87	8.41	10.96	6.18	6.21	5.13	4.86	5.61	4.45	16.25	6.94	3.26	6.38	7.34
	boot.	4.99	5.09	5.04	4.88	5.10	5.02	4.86	4.90	5.09	4.62	4.64	5.09	4.71	4.77
GAM_4	boot.	31.49	31.37	27.29	37.41	37.55	38.51	33.90	37.90	14.98	22.96	31.95	14.98	14.96	36.24
GAM_5	boot.	56.05	55.06	48.36	53.19	65.57	58.58	10.31	51.20	35.20	57.93	57.31	35.20	34.87	51.47
GAM_6	boot.	64.21	59.67	57.43	74.41	56.39	61.44	10.65	66.09	55.67	42.29	66.21	55.67	57.77	72.73
GAM_7	boot.	70.60	68.78	58.71	30.97	43.83	38.95	25.38	31.32	17.60	68.74	69.68	17.60	15.86	31.19
GAM_8	boot.	48.83	46.86	38.13	5.44	15.21	10.04	6.10	5.42	5.22	52.46	48.29	5.22	5.33	5.41
GAM_9	boot.	54.02	54.17	47.81	57.70	63.59	61.23	15.02	58.28	5.02	51.81	54.98	5.02	5.03	55.44
GAM_{10}	boot.	37.59	36.19	32.70	45.02	41.25	43.23	22.30	46.37	5.85	25.98	37.73	5.85	5.44	41.08
GAM_{11}	boot.	74.31	74.03	66.64	69.26	79.62	75.49	11.87	69.94	5.58	75.22	75.35	5.58	4.97	68.60
GAM_{12}	boot.	71.34	70.19	61.37	25.50	60.15	46.88	17.17	25.08	20.86	80.17	73.86	20.86	20.66	27.74
GAM_{13}	boot.	29.55	29.09	25.16	34.70	35.11	35.72	25.87	35.09	11.58	23.87	29.67	11.58	11.45	33.45
GAM_{14}	boot.	27.90	28.18	24.24	33.73	33.70	34.55	29.30	34.33	11.17	20.74	28.48	11.17	10.94	32.34
GAM_{15}	boot.	56.81	56.32	47.68	45.97	57.64	53.70	18.20	45.76	8.69	59.09	56.66	8.69	8.47	45.22

* NOTE: the Monte Carlo error is 0.218 (%) under the null models LN_1-LN_3 and GAM_1-GAM_3.

Table 3.9: Simulated probabilities (%) of rejecting H_0 at significance level 0.05 when data are generated from a mixture model with parameter settings given in Table 3.1 and $(n_0, n_1, n_2) = (100, 100, 100)$. The abbreviations "asym." and "boot." denote the asymptotic distribution and bootstrap procedure, respectively, that are used in calculation.

Model		$(n_0, n_1, n_2) = (100, 100, 100)$													
		ELR1	ELR2	ELR3	ELR4	ELR5	KW2	KW1	ANOVA2	ANOVA1	LRT-LN	LRT-GAM	ATS	WTS1	WTS2
LN_1	asym.	5.83	5.70	6.95	5.50	5.06	4.80	4.74	4.81	4.53	5.32	18.66	3.75	5.50	5.43
	boot.	4.95	5.06	5.11	5.05	4.91	4.69	4.83	5.01	4.90	4.89	4.60	4.90	4.83	4.91
LN_2	asym.	6.33	6.03	6.92	5.81	5.83	5.52	5.07	4.94	4.57	5.50	18.42	3.70	5.79	5.77
	boot.	5.09	5.17	4.90	5.09	5.44	5.46	4.94	4.91	5.05	5.17	4.65	5.05	4.87	5.11
LN_3	asym.	6.96	6.42	8.68	6.39	5.22	4.76	4.99	5.14	4.28	5.73	17.00	3.31	5.96	6.10
	boot.	5.13	4.98	5.20	5.18	4.69	4.80	4.96	5.02	4.94	4.58	4.74	4.94	4.84	5.00
LN_4	boot.	62.33	62.29	56.07	69.97	70.38	70.62	59.80	71.33	19.35	62.55	45.14	19.35	19.81	69.83
LN_5	boot.	81.64	82.14	76.65	62.37	87.79	85.96	34.00	47.78	56.62	82.45	57.22	56.62	56.36	50.60
LN_6	boot.	57.99	66.73	55.52	22.47	5.02	5.28	5.36	15.48	19.78	64.04	53.74	19.78	18.87	16.11
LN_7	boot.	98.27	98.90	97.71	62.64	93.93	93.97	39.69	66.04	6.91	99.15	73.90	6.91	6.43	65.30
LN_8	boot.	90.49	93.87	89.82	10.80	70.51	68.11	11.43	8.14	8.52	94.92	48.19	8.52	11.45	11.91
LN_9	boot.	63.49	63.98	57.47	66.81	72.04	72.07	39.20	69.03	4.73	64.25	37.29	4.73	4.75	66.40
LN_{10}	boot.	57.43	59.66	53.41	62.38	60.33	60.99	64.34	65.69	5.77	59.63	35.01	5.77	7.37	61.99
LN_{11}	boot.	95.87	96.79	94.58	69.11	96.04	95.54	22.20	70.91	4.76	96.98	61.21	4.76	5.37	69.99
LN_{12}	boot.	80.70	83.58	78.21	27.36	81.65	79.00	18.64	20.74	24.27	84.70	34.02	24.27	26.85	24.46
LN_{13}	boot.	56.44	56.59	50.03	62.90	64.64	65.10	50.83	65.76	7.11	56.84	30.95	7.11	7.98	61.13
LN_{14}	boot.	54.55	55.20	48.78	61.78	62.63	63.18	61.26	64.94	9.54	55.33	29.04	9.54	11.20	62.34
LN_{15}	boot.	95.80	96.33	93.95	76.80	96.97	96.57	20.74	77.62	6.94	96.44	66.76	6.94	7.84	76.04
GAM_1	asym.	5.67	6.24	6.63	5.52	5.38	5.12	4.98	5.34	4.78	17.75	5.42	4.51	5.36	5.54
	boot.	5.09	5.26	5.18	5.16	4.93	4.94	5.08	5.06	4.92	4.92	5.09	4.92	5.17	5.06
GAM_2	asym.	5.52	6.02	6.93	5.19	5.29	5.13	5.16	4.86	5.02	16.63	5.32	4.72	5.57	5.50
	boot.	5.04	5.11	5.08	4.85	5.13	5.15	5.10	4.76	5.02	4.58	4.88	5.02	4.76	5.07
GAM_3	asym.	6.60	6.93	8.58	5.92	5.61	5.33	5.24	5.64	4.98	17.13	5.90	4.26	5.95	6.36
	boot.	4.99	5.02	5.42	5.23	5.12	5.39	5.24	5.26	5.19	4.64	4.77	5.19	4.79	5.24
GAM_4	boot.	61.78	61.76	55.72	69.67	69.98	70.26	59.61	70.07	25.32	45.06	62.44	25.32	25.22	69.10
GAM_5	boot.	91.19	90.72	86.60	85.62	94.83	91.62	16.22	84.82	64.91	89.86	91.49	64.91	64.77	84.17
GAM_6	boot.	95.81	93.10	93.05	97.84	87.47	93.10	15.49	96.46	89.72	72.38	96.28	89.72	92.24	97.61
GAM_7	boot.	97.15	96.58	94.58	58.61	78.81	70.04	45.99	59.00	34.11	96.61	97.26	34.11	34.20	58.97
GAM_8	boot.	86.66	83.45	79.73	5.81	29.87	16.01	6.96	5.70	5.37	88.87	87.39	5.37	4.94	5.43
GAM_9	boot.	88.92	88.73	84.74	89.58	93.16	92.06	25.22	89.81	5.05	85.27	89.44	5.05	4.85	89.11
GAM_{10}	boot.	72.71	70.01	66.54	79.46	73.68	76.52	40.26	80.31	5.08	50.21	73.21	5.08	5.23	77.71
GAM_{11}	boot.	97.75	97.73	96.48	95.48	98.62	97.63	19.10	95.65	4.73	97.82	98.00	4.73	5.02	95.41
GAM_{12}	boot.	97.71	97.16	95.33	43.87	92.73	81.51	30.03	43.62	34.63	98.87	98.17	34.63	35.50	47.59
GAM_{13}	boot.	57.58	57.50	51.11	64.77	66.00	66.10	46.48	64.95	18.77	47.77	57.94	18.77	19.20	63.98
GAM_{14}	boot.	56.20	56.35	49.93	64.48	64.03	64.64	53.65	65.12	16.31	40.32	56.64	16.31	16.16	63.80
GAM_{15}	boot.	90.39	90.27	85.75	78.83	90.02	86.52	32.15	78.88	11.85	91.44	90.63	11.85	11.99	78.48

* NOTE: the Monte Carlo error is 0.218 (%) under the null models LN_1-LN_3 and GAM_1-GAM_3.

Table 3.10: Simulated probabilities (%) of rejecting H_0 at significance level 0.05 when data are generated from a mixture model with parameter settings given in Table 3.1 and $(n_0, n_1, n_2) = (50, 100, 150)$. The abbreviations "asym." and "boot." denote the asymptotic distribution and bootstrap procedure, respectively, that are used in calculation.

Model		$(n_0, n_1, n_2) = (50, 100, 150)$													
		ELR1	ELR2	ELR3	ELR4	ELR5	KW2	KW1	ANOVA2	ANOVA1	LRT-LN	LRT-GAM	ATS	WTS1	WTS2
LN_1	asym.	6.36	5.89	7.05	5.86	5.66	5.37	5.17	5.22	4.57	5.76	18.44	4.67	6.65	6.72
	boot.	5.23	5.20	5.17	5.11	5.20	5.31	5.20	5.18	4.91	5.22	4.92	4.71	4.84	4.84
LN_2	asym.	6.27	6.10	7.36	5.86	5.62	5.35	5.49	4.97	4.71	5.64	17.56	5.07	7.12	7.14
	boot.	5.11	5.17	5.03	5.02	5.30	5.40	5.53	4.92	5.11	5.00	4.60	5.00	5.06	4.92
LN_3	asym.	7.51	6.91	8.89	6.82	6.01	5.55	5.18	5.73	4.47	6.30	17.18	4.92	7.90	8.65
	boot.	5.23	4.99	4.97	5.53	5.63	5.59	5.36	5.45	5.11	4.95	4.94	4.98	4.83	5.33
LN_4	boot.	52.19	52.06	46.28	59.48	59.94	60.23	51.88	60.47	18.51	52.54	37.53	9.42	10.12	55.89
LN_5	boot.	72.73	74.01	67.33	57.12	80.83	78.34	26.24	27.40	34.00	74.82	47.67	60.89	58.95	54.60
LN_6	boot.	41.76	48.34	36.55	12.75	3.17	3.67	4.64	5.09	4.67	47.98	34.54	23.52	18.74	17.53
LN_7	boot.	96.20	97.50	95.97	60.03	90.59	89.04	37.73	59.23	13.75	97.78	76.20	3.62	3.57	54.65
LN_8	boot.	88.02	91.84	88.14	16.35	69.40	62.63	9.36	8.38	8.83	93.09	52.37	12.26	17.30	16.99
LN_9	boot.	54.16	54.25	48.31	58.97	62.78	62.77	33.81	57.56	4.58	54.69	31.67	3.92	3.75	55.49
LN_{10}	boot.	48.79	49.73	42.84	52.33	50.71	51.48	57.40	52.76	5.59	50.00	28.01	2.67	2.33	48.60
LN_{11}	boot.	91.70	93.67	91.06	66.09	92.85	91.28	19.01	62.00	5.08	94.00	59.85	6.38	7.58	62.82
LN_{12}	boot.	73.02	76.75	71.53	29.48	75.40	70.09	14.78	13.80	15.34	77.72	31.57	28.60	32.33	31.05
LN_{13}	boot.	47.17	47.25	41.93	54.19	55.33	56.04	44.91	54.73	8.10	47.84	26.26	3.40	3.30	50.19
LN_{14}	boot.	46.01	46.05	40.50	52.69	52.99	53.82	54.19	54.04	10.96	46.10	24.39	3.65	3.46	48.19
LN_{15}	boot.	91.18	92.46	89.67	73.39	93.93	92.64	18.05	67.55	4.17	92.83	63.53	9.63	10.87	69.35
GAM_1	asym.	5.58	6.13	6.72	5.21	5.47	5.04	4.82	5.12	4.82	17.29	5.33	5.07	6.03	5.82
	boot.	4.81	4.92	5.10	4.83	5.06	4.93	4.84	4.88	4.85	4.90	4.67	4.83	5.13	4.74
GAM_2	asym.	5.84	5.77	6.80	5.08	5.14	5.13	4.95	4.72	4.84	16.85	5.32	5.15	6.44	6.26
	boot.	4.92	4.67	5.01	4.74	4.73	5.11	5.06	4.62	4.98	4.36	4.74	5.00	5.08	4.88
GAM_3	asym.	6.70	6.91	8.27	5.65	5.64	5.04	5.08	5.44	4.89	15.77	6.16	5.44	7.13	7.42
	boot.	4.95	5.07	5.04	5.05	5.12	5.07	5.10	5.14	5.17	4.60	4.84	5.16	4.77	5.01
GAM_4	boot.	50.98	51.15	45.07	59.47	59.59	60.33	52.07	59.74	22.79	36.11	51.81	15.91	15.89	56.81
GAM_5	boot.	83.19	82.73	78.18	78.27	88.92	83.73	11.76	72.46	47.53	84.16	84.00	61.59	60.87	78.29
GAM_6	boot.	85.10	81.32	79.12	91.50	72.02	82.84	10.86	80.31	66.63	56.41	86.14	89.32	88.55	95.35
GAM_7	boot.	95.47	94.33	92.43	53.94	74.55	64.12	40.41	54.09	35.76	95.39	95.46	19.15	21.17	49.23
GAM_8	boot.	85.14	81.84	79.21	8.80	33.97	16.83	6.92	8.51	7.75	88.37	85.96	6.62	5.97	6.46
GAM_9	boot.	82.95	83.24	78.07	84.58	89.36	87.13	22.61	83.55	3.80	80.47	83.76	4.57	4.73	83.92
GAM_{10}	boot.	62.27	60.63	55.95	70.53	63.55	67.20	35.48	69.22	3.92	42.86	62.66	4.47	4.21	69.84
GAM_{11}	boot.	95.39	95.35	93.02	93.34	97.18	95.68	15.92	92.95	3.91	95.54	95.51	4.73	4.89	92.48
GAM_{12}	boot.	96.78	96.14	94.59	47.00	91.43	77.35	25.36	43.71	31.37	98.68	97.38	29.71	36.02	44.80
GAM_{13}	boot.	49.51	49.50	43.60	56.53	57.60	57.70	41.16	56.14	16.37	40.50	50.05	12.65	12.15	54.50
GAM_{14}	boot.	46.84	46.53	41.31	54.36	53.91	54.70	46.38	54.45	13.80	32.23	47.29	9.23	9.10	51.95
GAM_{15}	boot.	82.73	82.60	77.82	72.62	83.92	78.98	29.12	71.67	12.04	85.26	83.20	8.72	9.87	69.89

* NOTE: the Monte Carlo error is 0.218 (%) under the null models LN_1-LN_3 and GAM_1-GAM_3.

Table 3.11: Results from bootstrap and permutation methods for WTS1 and WTS2 when data are generated from a mixture model with parameter settings given in Table 3.1. The abbreviations "boot." and "perm." denote the bootstrap and permutation procedures, respectively.

Model	$(n_0, n_1, n_2) = (20, 20, 20)$				$(n_0, n_1, n_2) = (50, 50, 50)$				$(n_0, n_1, n_2) = (100, 100, 100)$				$(n_0, n_1, n_2) = (50, 100, 150)$			
	WTS1		WTS2		WTS1		WTS2		WTS1		WTS2		WTS1		WTS2	
	boot.	perm.	boot.	perm.	boot.	perm.	boot.	perm.	boot.	perm.	boot.	perm.	boot.	perm.	boot.	perm.
LN_1	4.39	4.77	4.66	4.76	4.58	4.50	4.64	4.80	4.83	5.02	4.91	5.02	4.84	4.99	4.84	5.02
LN_2	4.98	4.90	4.64	4.88	4.97	5.33	5.14	5.27	4.87	5.39	5.11	5.17	5.06	5.20	4.92	4.97
LN_3	5.78	5.70	5.79	5.78	4.91	4.76	4.65	5.00	4.84	5.21	5.00	5.07	4.83	5.32	5.33	5.53
LN_4	9.33	9.34	14.93	15.48	12.95	12.83	36.45	36.99	19.81	19.83	69.83	70.05	10.12	9.72	55.89	56.10
LN_5	13.83	14.74	11.86	13.08	33.17	34.69	28.58	29.62	56.36	58.08	50.60	51.42	58.95	60.00	54.60	55.28
LN_6	5.87	5.42	4.73	5.13	7.96	8.78	7.29	8.00	18.87	20.64	16.11	16.88	18.74	19.87	17.53	18.76
LN_7	4.30	4.70	16.64	17.33	4.92	5.66	35.84	36.66	6.43	6.78	65.30	66.04	3.57	3.72	54.65	55.55
LN_8	9.87	8.86	8.96	9.30	10.84	11.53	11.77	12.37	11.45	12.08	11.91	12.15	17.30	17.48	16.99	17.38
LN_9	4.63	5.69	14.21	14.93	5.31	5.71	34.82	35.86	4.75	4.97	66.40	66.92	3.75	3.70	55.49	56.11
LN_{10}	7.48	8.02	14.08	14.43	7.50	7.63	31.54	32.78	7.37	7.57	61.99	62.39	2.33	2.73	48.60	49.64
LN_{11}	4.62	5.33	16.74	17.98	5.23	5.68	39.56	40.49	5.37	5.90	69.99	70.52	7.58	7.97	62.82	63.06
LN_{12}	10.43	10.90	9.87	10.82	17.66	18.55	16.81	17.32	26.85	26.70	24.46	24.89	32.33	32.40	31.05	31.51
LN_{13}	5.73	6.45	13.24	14.00	6.99	7.64	33.10	33.90	7.98	7.94	61.13	63.36	3.30	3.11	50.19	50.89
LN_{14}	6.83	7.87	14.03	15.07	9.19	9.88	33.67	34.59	11.20	10.62	62.34	62.69	3.46	3.69	48.19	49.18
LN_{15}	4.82	5.61	18.41	19.64	6.23	6.53	44.32	45.05	7.84	8.11	76.04	76.61	10.87	11.07	69.35	69.97
GAM_1	5.14	5.32	5.18	5.22	4.72	4.96	4.98	4.98	5.17	4.95	5.06	5.05	5.13	4.93	4.74	4.84
GAM_2	5.36	5.04	5.03	5.13	5.02	5.20	4.72	4.88	4.76	5.05	5.07	4.97	5.08	4.97	4.88	4.90
GAM_3	6.06	5.77	5.10	5.46	4.71	5.05	4.77	4.95	4.79	5.04	5.24	5.42	4.77	4.94	5.01	4.91
GAM_4	9.70	8.83	11.84	12.24	14.96	15.01	36.24	36.20	25.22	25.01	69.10	69.13	15.89	15.91	56.81	56.69
GAM_5	15.46	14.88	22.20	23.00	34.87	35.52	51.47	51.68	64.77	65.17	84.17	84.21	60.87	61.23	78.29	78.55
GAM_6	19.36	18.88	25.88	26.48	57.77	59.49	72.73	73.52	92.24	92.38	97.61	97.62	88.55	88.71	95.35	95.45
GAM_7	8.03	7.58	12.39	12.88	15.86	16.45	31.19	31.12	34.20	33.55	58.97	59.14	21.17	21.04	49.23	49.50
GAM_8	6.24	6.44	6.40	6.20	5.33	5.47	5.41	5.43	4.94	5.11	5.43	5.45	5.97	6.64	6.46	6.49
GAM_9	5.13	5.24	20.73	20.99	5.03	5.23	55.44	55.57	4.85	4.99	89.11	89.15	4.73	4.65	83.92	83.99
GAM_{10}	5.65	5.86	15.14	15.34	5.44	5.88	41.08	41.26	5.23	5.09	77.71	77.83	4.21	4.44	69.84	69.78
GAM_{11}	5.36	4.72	27.26	27.89	4.97	5.36	68.60	68.12	5.02	4.66	95.41	95.40	4.89	5.02	92.48	92.27
GAM_{12}	12.16	12.80	16.00	16.16	20.66	21.29	27.74	27.97	35.50	35.29	47.59	47.60	36.02	35.45	44.80	44.78
GAM_{13}	8.19	8.77	13.54	13.39	11.45	11.62	33.45	33.49	19.20	18.56	63.98	63.85	12.15	12.26	54.50	54.65
GAM_{14}	8.27	7.54	11.91	12.09	10.94	11.32	32.34	32.64	16.16	16.21	63.80	63.67	9.10	8.83	51.95	51.99
GAM_{15}	6.67	5.88	17.11	17.63	8.47	8.17	45.22	45.22	11.99	11.70	78.48	78.56	9.87	9.53	69.89	70.14

Chapter 4

Semiparametric inference on the means of multiple nonnegative distributions with excess zero observations

4.1 Introduction

In this chapter, we answer the scientific question Q2 outlined in Section 1.1 of Chapter 1. That is, we would like to make inferences about the means of multiple nonnegative distributions with excess zero observations. Recall that we have $m+1$ independent samples as follows:

$$x_{i1}, \ldots, x_{in_i} \sim F_i(x) = \nu_i I(x=0) + (1-\nu_i) I(x>0) G_i(x), \quad i = 0, \ldots, m, \tag{4.1}$$

where n_i is the ith group's sample size, $I(\cdot)$ is an indicator function and the $G_i(\cdot)$'s are cumulative distribution functions with common support which may be continuous or discrete. Again, in this chapter, we concentrate on continuous distributions $G_i(\cdot)$'s whose support consists of all nonnegative real numbers; but we proposed, in Section 7.1, ways that the method can be applied to discrete distributions. Under (4.1), the mean of each $F_i(x)$ can be expressed as

$$\mu_i = \int_0^\infty x dF_i(x) = (1 - \nu_i) \int_0^\infty x dG_i(x), \quad i = 0, \ldots, m.$$

Our interest is to make inferences about $\mu_0, \ldots, \mu_m$. These include testing the null hypothesis $\mu_0 = \cdots = \mu_m$, and constructing confidence intervals for $\mu_i - \mu_j$ and μ_i/μ_j, for $i \neq j$.

The mean of a population with excess zeros has been considered an important summary quantity. For example, in fishery and health economics studies, the population total often has a crucial scientific meaning. The mean can provide information for recovering the population total, for example, for the total egg production of Atlantic mackerel (Pennington, 1983), and the total expenditure of patients (Chen and Zhou, 2006).

Inference on the means of two, or more, populations with excess zeros has been considered one of the most important, and fundamental, problems in many applications. For example, Tu and Zhou (1999) showed that testing the mean equality of several of these populations is a question of great importance in medical cost data analysis. Also, Zhou and Tu (2000) argued that confidence intervals for the ratio of mean diagnostic charges in two health groups can provide useful information on the magnitude of the relative difference between the two groups.

A natural way to make inference on these means is by using fully parametric models. Tu and Zhou (1999) and Zhou and Tu (1999) proposed to model G_i's by the log-normal distribution, based on which they developed a Wald and a likelihood ratio test, for testing the overall mean equality. Confidence intervals for the two-sample mean ratio and mean difference have been considered in Zhou and Tu (2000) and Chen and Zhou (2006), when the positive data in both samples follow log-normal distributions. Although the log-normal distributions are quite natural for modelling the positive observations, other parametric models, such as the gamma distribution, have also been argued to be suitable in applications (Marazzi et al., 1998; Nixon and Thompson, 2004). However, as concluded in Nixon and Thompson (2004), "when sample sizes are not large, different parametric models that fit the data equally well can lead to substantially different inferences". This fact may pose an issue of model robustness in the fully parametric approach.

Another approach is to use the nonparametric methods ignoring the mixture structure (4.1). As we mentioned in Chapter 3, the nonparametric ANOVA-type statistic (ATS) and the Wald-type permutation statistic (WTPS) are two representative methods for comparing the means of multiple samples. The ATS, proposed by Brunner et al. (1997), is an extension of the classical ANOVA F-test for heteroscedastic factorial designs. Brunner et al. (1997) suggested using an F-distribution with random degrees of freedom to approximate the finite sample distribution of the ATS. It can be shown that the ATS is equivalent to the Welch two-sample t-test (Welch, 1938) when $m = 1$. More recently, the WTPS was proposed by Pauly et al. (2015) for testing a linear hypothesis about the means without any distributional assumptions, under very general heteroscedastic factorial designs. In practice, the homoscedastic variance assumption is usually difficult to justify for multiple groups of observations with excess zeros; see for example, Zhou and Tu (1999) and Section 4.5. Hence, it is appropriate to directly apply the ATS and WTPS methods. The empirical likelihood method has also received considerable interest in dealing with such problems. Chen et al. (2003) and Chen and Qin (2003) used it to construct the confidence interval for the mean of a population with excess zeros. For the two-sample case, Taylor and Pollard (2009) considered a test for the means, and Kang et al. (2010) and Wu and Yan (2012) studied the constructions of confidence interval for the mean difference, by using the empirical likelihood.

In many applications, multiple populations may naturally share some common characteristics. It is therefore desirable to borrow efficiency across similar populations to improve the inferential results. At the same time, we might also want that inferences do not rely on any specific distributional assumption. In a similar way to Chapter 3, we propose to model the distributions of G_i's by the semiparametric DRM such that

$$dG_i(x) = \exp\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x)\} dG_0(x), \quad i = 0, \ldots, m \tag{4.2}$$

for a non-trivial, pre-specified, basis function $\mathbf{q}(x)$ of dimension d, and corresponding unknown parameters α_i and $\boldsymbol{\beta}_i$ ($\alpha_0 = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$). In this chapter, we propose a unified framework, based on an empirical likelihood ratio (ELR) statistic, for making inferences on the means of multiple nonnegative distributions under (4.1) and (4.2). Software implementing the proposed ELR for testing overall mean equality, with basis function $\mathbf{q}(x) = \log(x)$

in the DRM, has been developed in the `R` language (R Development Core Team, 2014) and is supplemented in the Appendix A.2 at the end of this thesis.

We note that deriving the asymptotic distribution of the proposed ELR statistic is technically challenging. Due to its non-standard mixture structure (4.1), the summations in the definition of the ELR (see Section 4.2) are over random numbers, i.e., the number of positive observations in each group. Hence, standard large sample theory may not be directly applied. In addition, we have to deal with a biased sampling problem (Qin, 1993) induced by (4.2) together with estimating equations. Unlike Chapter 3, we do not have a simple form for the profile empirical likelihood or dual empirical likelihood. This makes the theoretical derivation more complicated; see Qin et al. (2015). After some technical work, we show that the ELR enjoys a simple χ^2-type limiting distribution.

The structure of this chapter is as follows. In Section 4.2, we formulate the research problem, construct the empirical likelihood ratio statistic, and study its asymptotic properties. A numerical implementation is discussed in Section 4.3. Simulation results are reported in Section 4.4, and a real data set is analyzed in Section 4.5. For the convenience of presentation, proofs are given in Section 4.6.

4.2 Empirical likelihood inference under the DRM

4.2.1 Notation and problem setup

Let us first recall and introduce some notation. Recall that, as defined in Chapter 3, n_{i0} and n_{i1} denote the (random) numbers of zero and positive observations for the ith sample, for $i = 0, \ldots, m$. Define $n_{\cdot 0} = \sum_{i=0}^{m} n_{i0}$ and $n_{\cdot 1} = \sum_{i=0}^{m} n_{i1}$ the total zero and nonzero sample sizes, and let $n = \sum_{i=0}^{m} n_i$ denote the total sample size. Without loss of generality, we use first n_{i1} observations $x_{i1}, \ldots, x_{in_{i1}}$ to denote the positive observations in the ith sample for $i = 0, \ldots, m$.

Let $\boldsymbol{\mu} = (\mu_0, \ldots, \mu_m)^\top$ be the mean vector of the $m + 1$ groups. The main goal of this section is to develop a test for the following general linear hypothesis about the means

$$H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{d}, \tag{4.3}$$

where the $p \times (m+1)$ matrix $\mathbf{C}$ and $p \times 1$ vector $\mathbf{d}$ have real, non-random, entries that are completely specified under the null hypothesis and do not depend on sample sizes. We assume that $\mathbf{C}$ has full row rank so that $rank(\mathbf{C}) = p$ (with $p \leq m+1$).

We comment that formulation (4.3) is very flexible and includes many special cases. For example, when

$$\mathbf{C} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \cdots & 1 \end{pmatrix}_{m \times (m+1)}, \quad \mathbf{d} = \mathbf{0}_{m \times 1}, \tag{4.4}$$

then the hypothesis (4.3) becomes

$$H_0^* : \mu_0 = \mu_1 = \cdots = \mu_m. \tag{4.5}$$

In our numerical studies, in Section 4.4, we focus on this special, though important, hypothesis (4.5).

4.2.2 Empirical likelihood ratio

For a compact presentation, we use vector notation. Let $\boldsymbol{\nu} = (\nu_0, \ldots, \nu_m)^\top$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_0^\top, \ldots, \boldsymbol{\theta}_m^\top)^\top$ with $\boldsymbol{\theta}_i = (\alpha_i, \boldsymbol{\beta}_i^\top)^\top$ for $i = 0, \ldots, m$. Also, let $\boldsymbol{\omega}(x; \boldsymbol{\theta}) = \big(\omega_1(x; \boldsymbol{\theta}_1), \ldots, \omega_m(x; \boldsymbol{\theta}_m)\big)^\top$ with $\omega_i(x; \boldsymbol{\theta}_i) = \exp\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x)\}$ for $i = 0, \ldots, m$.

Under the DRM (4.2) for the G_i's, we can refine the definition of the means based on the pooled positive samples

$$\mu_i = (1 - \nu_i) \int_0^\infty x \omega_i(x; \boldsymbol{\theta}_i) dG_0(x), \qquad i = 0, \ldots, m. \tag{4.6}$$

Under the general null hypothesis H_0 in (4.3), we have $\mathbf{C}\boldsymbol{\mu} - \mathbf{d} = \mathbf{0}$, or equivalently

$$E_0 \{\mathbf{g}(X; \boldsymbol{\nu}, \boldsymbol{\theta})\} = \mathbf{0}_{p \times 1},$$

where $X \sim G_0(x)$ and E_0 means that the expectation is taken under $G_0(x)$, and

$$\mathbf{g}(x;\boldsymbol{\nu},\boldsymbol{\theta}) = \begin{pmatrix} g_1(x;\boldsymbol{\nu},\boldsymbol{\theta}) \\ \cdots \\ g_p(x;\boldsymbol{\nu},\boldsymbol{\theta}) \end{pmatrix} = \mathbf{C} \begin{pmatrix} (1-\nu_0)x \\ (1-\nu_1)x\omega_1(x;\boldsymbol{\theta}_1) \\ \cdots \\ (1-\nu_m)x\omega_m(x;\boldsymbol{\theta}_m) \end{pmatrix} - \mathbf{d}. \tag{4.7}$$

Therefore, the information about the means in null hypothesis H_0 can come into a p-dimensional unbiased estimating equation (4.7). For parameters estimated through unbiased estimating equations, the empirical likelihood method has been shown to provide an effective inference platform (Qin and Lawless, 1994). Based on the construction of unbiased estimating equation (4.7), we proceed to develop inference procedures using the empirical likelihood method.

Along the lines of empirical likelihood (see Section 2.2), we restrict the form of baseline distribution G_0 to be

$$G_0(x) = \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} p_{ij} I(x_{ij} \le x).$$

Recall that, in Chapter 3, given multiple groups of samples from (4.1) in which the G_i's satisfy the DRM (4.2), the empirical log-likelihood function can be written as

$$\tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0) = \sum_{i=0}^{m} \log\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\} + \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} \left\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij}) + \log(p_{ij})\right\}.$$

We always have the following set of natural constraints:

$$\mathcal{C}_1 = \left\{(\boldsymbol{\nu},\boldsymbol{\theta},G_0):\ \nu_i \in (0,1),\ \ p_{ij} > 0,\ \ \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} p_{ij} = 1,\ \ \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} p_{ij}\{\boldsymbol{\omega}(x_{ij},\boldsymbol{\theta}) - \mathbf{1}\} = \mathbf{0}_{m\times 1}\right\}.$$

Under the general null hypothesis H_0 in (4.3), we also have the following set of constraints:

$$\mathcal{C}_2 = \left\{(\boldsymbol{\nu},\boldsymbol{\theta},G_0):\ \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} p_{ij}\mathbf{g}(x_{ij};\boldsymbol{\nu},\boldsymbol{\theta}) = \mathbf{0}_{p\times 1}\right\}.$$

The empirical likelihood ratio (ELR) statistic for testing the general null hypothesis given in (4.3) is then defined via

$$R_n = 2\left\{\sup_{(\boldsymbol{\nu},\boldsymbol{\theta},G_0)\in\mathcal{C}_1} \tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0) - \sup_{(\boldsymbol{\nu},\boldsymbol{\theta},G_0)\in\mathcal{C}_1\cap\,\mathcal{C}_2} \tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)\right\}. \tag{4.8}$$

For the convenience of presentation, the numerical evaluation of R_n is discussed in Section 4.3.

4.2.3 Large sample property

In this section, we study the asymptotic distribution of the ELR statistic, R_n, for the general hypothesis testing problem in (4.3) under (4.1) and (4.2).

Suppose that the true value of $(\boldsymbol{\nu}^\top, \boldsymbol{\theta}^\top)^\top$ is $(\boldsymbol{\nu}^{*\top}, \boldsymbol{\theta}^{*\top})^\top$ under the null hypothesis H_0. Deriving the asymptotic distribution of R_n relies on the following regularity conditions.

R1. $\nu_i^* \in (0, 1)$ for $i = 0, \ldots, m$.

R2. $\lim_{\min\{n_0,\ldots,n_m\}\to\infty} n_i/n \to \rho_i^*$, where $\rho_i^* \in (0, 1)$ for $i = 0, \ldots, m$.

R3. $\int \left(1, \mathbf{q}^\top(x)\right)^\top \left(1, \mathbf{q}^\top(x)\right) dG_i(x)$ exists and is positive definite for $i = 0, \ldots, m$.

R4. $\int \exp\left\{\boldsymbol{\beta}_i^\top \mathbf{q}(x)\right\} dG_i(x) < \infty$ in a neighbourhood of $\boldsymbol{\theta}^*$.

R5. The matrix $\mathbf{U}$ defined in (4.31), in Section 4.6, is positive definite.

R6. $\|\partial \mathbf{g}(x; \boldsymbol{\nu}, \boldsymbol{\theta})/\partial \boldsymbol{\eta}\|$ and $\|\mathbf{g}(x; \boldsymbol{\nu}, \boldsymbol{\theta})\|^3$ are bounded by some integrable function of x in a neighbourhood of $(\boldsymbol{\nu}^{*\top}, \boldsymbol{\theta}^{*\top})^\top$, where $\boldsymbol{\eta} = (\boldsymbol{\nu}^\top, \boldsymbol{\theta}^\top)^\top$ and $\|\cdot\|$ denotes Euclidean norm.

Condition R1 states that the parameter $\boldsymbol{\nu}^*$ is an interior point of the parameter space of $\boldsymbol{\nu}$. Condition R2 assumes that the ratio of each group sample size to n converges to a constant as $\min\{n_0, \ldots, n_m\} \to \infty$. For simplicity, and convenience of presentation, we write $\rho_i^* = n_i/n$ and assume that it is a constant. This does not affect our technical development. Under Conditions R1 and R2, there is no need to distinguish the stochastic orders with respect to n or n_i. Condition R3 is an identifiability condition, and it ensures that the components of $\{1, \mathbf{q}^\top(x) I(x > 0)\}$ are linearly independent under all $G_i(x)$'s, and hence $\mathbf{q}(x)$ can not be a constant function. Conditions R3–R6 guarantee that a quandratic approximation of R_n is applicable. The following theorem defines the asymptotic null distribution of R_n under the general null hypothesis H_0 in (4.3).

Theorem 4.1 *Suppose we have $m+1$ groups of samples of the form (4.1) and condition (4.2) is satisfied. Assume, also, that the regularity conditions R1–R6 hold. Under the null hypothesis H_0, given in (4.3), we have*

$$R_n \to \chi^2_p,$$

in distribution as $n \to \infty$, where χ^2_p is a chi-squared random variable with p degrees of freedom, and $p = rank(\mathbf{C})$ for some full rank $\mathbf{C}$ in H_0.

For convenience of presentation, the proof of Theorem 4.1 is given in Section 4.6. Here we make three remarks about Theorem 4.1.

(a) As a direct consequence of Theorem 4.1, the ELR test for the overall equality of $m+1$ group means, (4.5), has a limiting chi-squared distribution with m degrees of freedom, since the rank of $\mathbf{C}$ in (4.4) is m.

(b) The mean differences and ratios are two quantities commonly used to measure the magnitudes of relative differences among the group means. The result of Theorem 4.1 is also useful for the construction of confidence interval, or region, for the mean differences and ratios. As an illustration, suppose we are interested in constructing the confidence interval for the mean difference $\delta = \mu_1 - \mu_0$ in the two-sample problem. The unbiased estimating equation (4.7) can be replaced by

$$d(x; \boldsymbol{\nu}, \boldsymbol{\theta}, \delta) = (1-\nu_1)x\omega_1(x; \boldsymbol{\theta}_1) - (1-\nu_0)x - \delta.$$

Then, the ELR, defined in (4.8), becomes a function of δ, since this parameter δ is incorporated through $d(x; \boldsymbol{\nu}, \boldsymbol{\theta}, \delta)$ in the constraint set $\mathcal{C}_2$. We denote it as $R_n(\delta)$. It follows that the 95% ELR confidence interval for δ can be constructed as $\{\delta : R_n(\delta) \le \chi^2_{1,0.95}\}$, where $\chi^2_{1,0.95}$ denotes the 95th quantile of the χ^2_1 distribution.

(c) A Wald-type statistic may also be constructed based on the normal approximation to $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})$ defined in Section 4.3. However, such a statistic is not invariant to transformations (Critchley et al., 1996). For example, mean differences and mean ratios are two different nonlinear transformations of $(\boldsymbol{\nu}, \boldsymbol{\theta})$. The Wald-type statistics for testing mean ratios equal to one, and for testing mean differences equal to zero, could lead to two different conclusions.

4.3 Numerical implementation

In this section, we discuss numerical calculation of R_n, defined in (4.8), for some given $\mathbf{C}$ and $\mathbf{d}$.

We first discuss how to calculate $\sup_{(\boldsymbol{\nu},\boldsymbol{\theta},G_0)\in\mathcal{C}_1}\tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)$ in R_n. Note that the optimization problem of maximizing $\tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)$ subject to $\mathcal{C}_1$ for given $(\boldsymbol{\nu},\boldsymbol{\theta})$ is identical to the one discussed in Section 3.2. Following a similar profiling procedure as used in Section 2.2 and the results in Proposition 2.1, we have

$$\sup_{(\boldsymbol{\nu},\boldsymbol{\theta},G_0)\in\mathcal{C}_1}\tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)=\sup_{\boldsymbol{\nu},\boldsymbol{\theta}}\ell_A(\boldsymbol{\nu},\boldsymbol{\theta})-n_{\cdot 1}\log(n_{\cdot 1}), \tag{4.9}$$

where

$$\begin{aligned}\ell_A(\boldsymbol{\nu},\boldsymbol{\theta}) \;=\; & \sum_{i=0}^{m}\log\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\}+\sum_{i=1}^{m}\sum_{j=1}^{n_{i1}}\left\{\alpha_i+\boldsymbol{\beta}_i^{\top}\mathbf{q}(x_{ij})\right\}\\ & -\sum_{i=0}^{m}\sum_{j=1}^{n_{i1}}\log\left[\rho_0+\sum_{r=1}^{m}\rho_r\exp\{\alpha_r+\boldsymbol{\beta}_r^{\top}\mathbf{q}(x_{ij})\}\right],\end{aligned} \tag{4.10}$$

with $\rho_r=n_{r1}/n_{\cdot 1}$ for $r=0,\ldots,m$. Note that, here, we use $\ell_A(\boldsymbol{\nu},\boldsymbol{\theta})$ to denote the dual empirical log-likelihood to emphasize that it is defined under the alternative hypothesis. The numerical calculation of $(\hat{\boldsymbol{\nu}},\hat{\boldsymbol{\theta}})=\arg\sup_{\boldsymbol{\nu},\boldsymbol{\theta}}\ell_A(\boldsymbol{\nu},\boldsymbol{\theta})$ and $\ell_A(\hat{\boldsymbol{\nu}},\hat{\boldsymbol{\theta}})=\sup_{\boldsymbol{\nu},\boldsymbol{\theta}}\ell_A(\boldsymbol{\nu},\boldsymbol{\theta})$ can be solved straightforwardly via the connection with logistic regression, as discussed in Section 2.2.

We next discuss how to calculate $\sup_{(\boldsymbol{\nu},\boldsymbol{\theta},G_0)\in\mathcal{C}_1\cap\,\mathcal{C}_2}\tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)$ in R_n. We start with the profiling procedure of $\tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)$ by profiling out the infinite dimensional parameter G_0. First, we set up the Lagrangian function. For given $(\boldsymbol{\nu},\boldsymbol{\theta})$, define

$$\Psi(G_0,\boldsymbol{\lambda},\mathbf{t})=\tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)+\sum_{i=0}^{m}\sum_{j=1}^{n_{i1}}p_{ij}\boldsymbol{\lambda}^{\top}\{\boldsymbol{\omega}(x_{ij},\boldsymbol{\theta})-\mathbf{1}\}+\sum_{i=0}^{m}\sum_{j=1}^{n_{i1}}p_{ij}\mathbf{t}^{\top}\mathbf{g}(x_{ij};\boldsymbol{\nu},\boldsymbol{\theta}),$$

where $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_m)^{\top}$ and $\mathbf{t}=(t_1,\ldots,t_p)^{\top}$ are corresponding Lagrangian multipliers. The point $\{p_{i1},\ldots,p_{in_i}: i=0,\ldots,m\}$ that maximize $\tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)$ must be a stationary point of Ψ satisfying

$$\frac{\partial\Psi(G_0,\boldsymbol{\lambda},\mathbf{t})}{\partial p_{ij}}=0,\quad \frac{\partial\Psi(G_0,\boldsymbol{\lambda},\mathbf{t})}{\partial\lambda_i}=0,\quad \text{and}\quad \frac{\partial\Psi(G_0,\boldsymbol{\lambda},\mathbf{t})}{\partial t_i}=0. \tag{4.11}$$

It follows from (4.11) that, for fixed $(\boldsymbol{\nu}, \boldsymbol{\theta})$, $\tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0)$ attains its maximum at

$$p_{ij} = \frac{1}{n_{\cdot 1}} \cdot \frac{1}{1 + \boldsymbol{\lambda}^\top \{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \mathbf{1}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta})}, \tag{4.12}$$

where the Lagrange multipliers, $\boldsymbol{\lambda}$ and $\mathbf{t}$, solve following equations

$$\frac{1}{n_{\cdot 1}} \sum_{i=0}^{m} \sum_{j=1}^{n_{i1}} \frac{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \mathbf{1}}{1 + \boldsymbol{\lambda}^\top \{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \mathbf{1}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta})} = \mathbf{0}_{m\times 1}, \tag{4.13}$$

and

$$\frac{1}{n_{\cdot 1}} \sum_{i=0}^{m} \sum_{j=1}^{n_{i1}} \frac{\mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta})}{1 + \boldsymbol{\lambda}^\top \{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \mathbf{1}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta})} = \mathbf{0}_{p\times 1}. \tag{4.14}$$

Therefore, using (4.12) to profile out p_{ij}, the profile empirical log-likelihood function of $(\boldsymbol{\nu}, \boldsymbol{\theta})$ under the null hypothesis H_0 given in (4.3) can be written as $\ell_N(\boldsymbol{\nu}, \boldsymbol{\theta}) - n_{\cdot 1}\log(n_{\cdot 1})$ with

$$\begin{aligned} \ell_N(\boldsymbol{\nu}, \boldsymbol{\theta}) &= \sum_{i=0}^{m} \log\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\} + \sum_{i=1}^{m}\sum_{j=1}^{n_{i1}} \left\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij})\right\} \\ &\quad - \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} \log\Big[1 + \boldsymbol{\lambda}^\top \{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \mathbf{1}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta})\Big]. \end{aligned} \tag{4.15}$$

Then, it follows that

$$\sup_{(\boldsymbol{\nu}, \boldsymbol{\theta},\, G_0)\in \mathcal{C}_1 \cap\, \mathcal{C}_2} \tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) = \sup_{\boldsymbol{\nu}, \boldsymbol{\theta}} \ell_N(\boldsymbol{\nu}, \boldsymbol{\theta}) - n_{\cdot 1}\log(n_{\cdot 1}). \tag{4.16}$$

Let $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}) = \arg\sup_{\boldsymbol{\nu}, \boldsymbol{\theta}} \ell_N(\boldsymbol{\nu}, \boldsymbol{\theta})$ be the maximum EL estimate of $(\boldsymbol{\nu}, \boldsymbol{\theta})$ under H_0 in (4.3). Hence, to calculate $\sup_{(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0)\in \mathcal{C}_1 \cap\, \mathcal{C}_2} \tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0)$, it is sufficient to obtain $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})$.

Unfortunately, the numerical calculation of $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})$ may not be an easy task since there are no analytical solutions for $\boldsymbol{\lambda}$ and $\mathbf{t}$ in the definition of $\ell_N(\boldsymbol{\nu}, \boldsymbol{\theta})$ in (4.15).

Let $\boldsymbol{\psi} = (\boldsymbol{\nu}^\top, \boldsymbol{\theta}^\top, \boldsymbol{\lambda}^\top, \mathbf{t}^\top)^\top$ and define

$$\ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t}) = \sum_{i=0}^{m} \log\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\} + \sum_{i=1}^{m}\sum_{j=1}^{n_{i1}} \left\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij})\right\}$$

$$-\sum_{i=0}^{m}\sum_{j=1}^{n_{i1}}\log\Big[1+\boldsymbol{\lambda}^\top\{\boldsymbol{\omega}(x_{ij},\boldsymbol{\theta})-\mathbf{1}\}+\mathbf{t}^\top\mathbf{g}(x_{ij};\boldsymbol{\nu},\boldsymbol{\theta})\Big]. \quad (4.17)$$

Then $\ell_N(\boldsymbol{\nu},\boldsymbol{\theta})=\ell(\boldsymbol{\nu},\boldsymbol{\theta},\boldsymbol{\lambda},\mathbf{t})$ with $\boldsymbol{\lambda}$ and $\mathbf{t}$ being the solutions of (4.13) and (4.14). Note that, $\boldsymbol{\lambda}$ and $\mathbf{t}$ in $\ell_N(\boldsymbol{\nu},\boldsymbol{\theta})$ are actually functions of $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$.

Let $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\mathbf{t}}$ be the solutions of (4.13) and (4.14) when $(\boldsymbol{\nu},\boldsymbol{\theta})$ is replaced by $(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})$. Hence, $\ell(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}},\tilde{\boldsymbol{\lambda}},\tilde{\mathbf{t}})=\ell_N(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})$. Further let $\tilde{\boldsymbol{\psi}}=(\tilde{\boldsymbol{\nu}}^\top,\tilde{\boldsymbol{\theta}}^\top,\tilde{\boldsymbol{\lambda}}^\top,\tilde{\mathbf{t}}^\top)^\top$. We summarize some key properties of using $\ell(\boldsymbol{\nu},\boldsymbol{\theta},\boldsymbol{\lambda},\mathbf{t})$ to find $(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})=\arg\sup_{\boldsymbol{\nu},\boldsymbol{\theta}}\ell_N(\boldsymbol{\nu},\boldsymbol{\theta})$ and $\ell_N(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})=\sup_{\boldsymbol{\nu},\boldsymbol{\theta}}\ell_N(\boldsymbol{\nu},\boldsymbol{\theta})$ in the following proposition.

Proposition 4.1 *For $\ell(\boldsymbol{\nu},\boldsymbol{\theta},\boldsymbol{\lambda},\mathbf{t})$, defined in (4.17), and $(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})=\arg\sup_{\boldsymbol{\nu},\boldsymbol{\theta}}\ell_N(\boldsymbol{\nu},\boldsymbol{\theta})$, then $\tilde{\boldsymbol{\psi}}$ is a stationary point of $\ell(\boldsymbol{\nu},\boldsymbol{\theta},\boldsymbol{\lambda},\mathbf{t})$. That is, $\tilde{\boldsymbol{\psi}}$ is a solution of*

$$\frac{\partial\ell(\boldsymbol{\nu},\boldsymbol{\theta},\boldsymbol{\lambda},\mathbf{t})}{\partial\boldsymbol{\psi}}=\mathbf{0}.$$

Proof. The result can be proved by the similar arguments as used in the proof of Proposition 2.1. We sketch the key steps.

First, when $\ell_N(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})$ is maximized, the following equations are satisfied:

$$\frac{\partial\ell_N(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\nu}}=\mathbf{0},\qquad\frac{\partial\ell_N(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}=\mathbf{0}.$$

Further, note that since $\boldsymbol{\lambda}$ and $\mathbf{t}$ in $\ell_N(\boldsymbol{\nu},\boldsymbol{\theta})$ are the solutions of (4.13) and (4.14), they are functions of $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$.

For illustration, we only show

$$\begin{aligned}\mathbf{0} &= \frac{\partial\ell_N(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_i}\\ &= \frac{\partial\ell(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}},\tilde{\boldsymbol{\lambda}},\tilde{\mathbf{t}})}{\partial\boldsymbol{\beta}_i}+\sum_{r=1}^{m}\frac{\partial\ell(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}},\tilde{\boldsymbol{\lambda}},\tilde{\mathbf{t}})}{\partial\lambda_r}\frac{\partial\lambda_r}{\partial\boldsymbol{\beta}_i}\Big|_{(\boldsymbol{\nu},\boldsymbol{\theta})=(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})}+\sum_{s=1}^{p}\frac{\partial\ell(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}},\tilde{\boldsymbol{\lambda}},\tilde{\mathbf{t}})}{\partial t_s}\frac{\partial t_s}{\partial\boldsymbol{\beta}_i}\Big|_{(\boldsymbol{\nu},\boldsymbol{\theta})=(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})}\\ &= \frac{\partial\ell(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}},\tilde{\boldsymbol{\lambda}},\tilde{\mathbf{t}})}{\partial\boldsymbol{\beta}_i}-0\cdot\frac{\partial\lambda_r}{\boldsymbol{\beta}_i}\Big|_{(\boldsymbol{\nu},\boldsymbol{\theta})=(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})}-0\cdot\frac{\partial t_s}{\boldsymbol{\beta}_i}\Big|_{(\boldsymbol{\nu},\boldsymbol{\theta})=(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})}\end{aligned}$$

$$= \frac{\partial \ell(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}})}{\partial \boldsymbol{\beta}_i}, \tag{4.18}$$

where the second last line is followed from (4.13) and (4.14).

One can similarly verify as (4.18) that

$$\frac{\partial \ell_N(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\eta}} = \frac{\partial \ell(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}})}{\partial \boldsymbol{\eta}} = \mathbf{0}.$$

Hence $\tilde{\boldsymbol{\psi}}$ is a stationary point of $\ell(\boldsymbol{\theta})$. This completes the proof. □

With the result of Proposition 4.1, $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})$ can be obtained by solving for a stationary point, in fact a saddlepoint of $\ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t})$, over the space of $\boldsymbol{\psi}$. To numerically calculate $\tilde{\boldsymbol{\psi}}$, we minimize the sum of squares of $\partial \ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t})/\partial \boldsymbol{\psi}$ by using the built-in `nlminb` function in `R`. Once we obtain $\tilde{\boldsymbol{\psi}}$, we calculate $\sup_{(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) \in \mathcal{C}_1 \cap \mathcal{C}_2} \tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0)$ by (4.16).

Combining (4.9) and (4.16), we finished the numeric calculation of R_n by

$$R_n = 2 \left\{ \ell_A(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}) - \ell_N(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}) \right\}.$$

We have also written `R` functions to calculate R_n for testing overall mean equality, with basis function $\mathbf{q}(x) = \log(x)$ in the DRM, and they are available in the Appendix A.2 of this thesis.

4.4 Simulation studies

In this section, we use Monte Carlo simulation to evaluate the finite-sample performance of the proposed ELR test for testing the overall mean equality, that is an important case covered by the hypothesis testing problem (4.3). The ELR test statistic is calculated by the procedure in Section 4.3 with the forms of $\mathbf{C}$ and $\mathbf{d}$ given in (4.4).

We fix the number of groups under comparison to be $m + 1 = 2$ or $m + 1 = 3$. We compare the type I error rates and the power of the proposed ELR test with the ATS of Brunner et al. (1997) and WTPS of Pauly et al. (2015). Note that the classical ANOVA F-test is designed for the assumption that the variances of $F_i(x)$ are homogenous, which may not be satisfied in our setup. Recall that both the ATS and WTPS do not require such an assumption. Hence in our comparison, we only include the ATS and WTPS.

For each test, the type I error rate and the power at the 5% significance level are calculated based on 10,000 and 2,000 repetitions, respectively. The computations for the ATS and WTPS methods use the `R` package "`GFD`" (Friedrich et al., 2016). Following the suggestion in Pauly et al. (2015), we use 10,000 permutation samples (the default number in "`GFD`") to calculate the type I error rates and the power of the WTPS.

The random observations are generated, conditional on all $\hat{\nu}_i$'s $\neq 0$ or 1, from (4.1) with all the G_i's being log-normal, or all the G_i's being gamma distributions. Note that if any $\hat{\nu}_i$'s $= 0$ or 1, then some test statistics may not be well defined. This is not a problem in practice. However, note that, when any true zero proportion ν_i is too close to the boundary 0 or 1, Anaya-Izquierdo et al. (2014) found that boundary effects can dominate the sampling distribution of $\hat{\nu}_i$. Hence, in our simulation settings, the true ν_i's are considered to be between 0.3 and 0.7. A diagnostic tool was proposed in Anaya-Izquierdo et al. (2014) which defines how far ν_i is required to be from the boundary so that first order asymptotics remain adequate.

In the following, we use $\text{LN}(a_i, b_i)$ to denote a log-normal distribution with mean a_i and variance b_i both with respect to the log scale (i.e. mean and variance of the associated normal random variable), and $\text{GAM}(a_i, b_i)$ to denote a gamma distribution with shape parameter a_i and scale parameter b_i.

The parameter settings under the null hypothesis (LN_1–LN_6 and GAM_1–GAM_6) that all the means are equal, and the alternative hypothesis (LN_7–LN_{15} and GAM_7–GAM_{15}) are given in Table 4.1. Note that in the following we use the same model notation for two- and three-sample comparisons when no confusion is caused. We consider the case with equal sample sizes by setting (n_0, n_1) to be $(50, 50)$ and $(100, 100)$ for the two-sample comparison, and (n_0, n_1, n_2) to be $(50, 50, 50)$ and $(100, 100, 100)$ for the three-sample comparison. We also consider the case with unequal sample sizes by setting (n_0, n_1) to be $(50, 150)$ and $(150, 50)$ for the two-sample comparison, and (n_0, n_1, n_2) to be $(50, 150, 100)$ and $(150, 50, 100)$ for the three-sample comparison. With the parameter settings in Table 4.1, these combinations of unequal sample sizes correspond to two cases where increasing the sample sizes is related with increasing variances (positive pairing) or with decreasing variances (negative pairing).

Table 4.1: Parameter settings for simulation studies. In the second column, each LN_1–LN_{16} and each GAM_1–GAM_{16} denote mixture models whose continuous parts follow the distributions $LN(a_i, b_i)$ and $GAM(a_i, b_i)$, respectively, for $i = 0, 1$ under two-sample comparison, or for $i = 0, 1, 2$ under three-sample comparison.

Scenario	Model	(ν_0, ν_1, ν_2)	(a_0, a_1, a_2)	(b_0, b_1, b_2)	Means	Variances
Scenario I (null)	LN_1	(0.3, 0.3, 0.3)	(0.00, 0.00, 0.00)	(1.00, 1.00, 1.00)	(1.15, 1.15, 1.15)	(3.84, 3.84, 3.84)
	LN_2	(0.7, 0.7, 0.7)	(0.00, 0.00, 0.00)	(1.00, 1.00, 1.00)	(0.49, 0.49, 0.49)	(1.97, 1.97, 1.97)
Scenario II (null)	LN_3	(0.3, 0.5, 0.4)	(0.33, 0.66, 0.48)	(1.00, 1.00, 1.00)	(1.60 ,1.60, 1.60)	(7.38, 11.36, 9.04)
	LN_4	(0.5, 0.7, 0.6)	(0.37, 0.89, 0.60)	(1.00, 1.00, 1.00)	(1.20, 1.20, 1.20)	(6.39, 11.61, 8.35)
Scenario III (null)	LN_5	(0.3, 0.5, 0.4)	(0.05, 0.29, 0.16)	(0.80, 1.00, 0.90)	(1.10, 1.10, 1.10)	(2.64, 5.37, 3.75)
	LN_6	(0.5, 0.7, 0.6)	(0.00, 0.50, 0.25)	(0.94, 0.96, 0.89)	(0.80, 0.80, 0.80)	(2.64, 4.94, 3.24)
Scenario I (alternative)	LN_7	(0.5, 0.3, 0.4)	(0.00, 0.00, 0.00)	(1.00, 1.00, 1.00)	(0.82, 1.15, 0.99)	(3.01, 3.84, 3.45)
	LN_8	(0.7, 0.5, 0.6)	(0.00, 0.00, 0.00)	(1.00, 1.00, 1.00)	(0.49, 0.82, 0.66)	(1.97, 3.01, 2.52)
	LN_9	(0.6, 0.4, 0.5)	(0.00, 0.00, 0.00)	(1.00, 1.00, 1.00)	(0.66, 0.99, 0.82)	(2.52, 3.45, 3.01)
Scenario II (alternative)	LN_{10}	(0.3, 0.3, 0.3)	(0.00, 0.50, 0.25)	(1.00, 1.00, 1.00)	(1.15, 1.90, 1.48)	(3.84, 10.44, 6.33)
	LN_{11}	(0.7, 0.7, 0.7)	(0.00, 0.75, 0.50)	(1.00, 1.00, 1.00)	(0.49, 1.05, 0.82)	(1.97, 8.84, 5.36)
	LN_{12}	(0.4, 0.6, 0.5)	(0.00, 1.00, 0.50)	(1.00, 1.00, 1.00)	(0.99, 1.79, 1.36)	(3.45, 18.63, 8.20)
Scenario III (alternative)	LN_{13}	(0.3, 0.3, 0.3)	(0.00, 0.50, 0.25)	(1.00, 0.80, 0.90)	(1.15, 1.72, 1.41)	(3.84, 6.46, 4.99)
	LN_{14}	(0.7, 0.7, 0.7)	(0.00, 0.75, 0.50)	(1.00, 0.80, 0.90)	(0.49, 0.95, 0.78)	(1.97, 5.76, 4.33)
	LN_{15}	(0.6, 0.4, 0.5)	(0.00, 0.50, 0.25)	(1.00, 0.60, 0.80)	(0.66, 1.34, 0.96)	(2.52, 3.63, 3.17)
Scenario I (null)	GAM_1	(0.3, 0.3, 0.3)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(0.70, 0.70, 0.70)	(0.91, 0.91, 0.91)
	GAM_2	(0.7, 0.7, 0.7)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(0.30, 0.30, 0.30)	(0.51, 0.51, 0.51)
Scenario II (null)	GAM_3	(0.3, 0.5, 0.4)	(1.43, 2.00, 1.67)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(1.43, 2.00, 1.67)
	GAM_4	(0.5, 0.7, 0.6)	(2.00, 3.33, 2.50)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(2.00, 3.33, 2.50)
Scenario III (null)	GAM_5	(0.3, 0.5, 0.4)	(1.71, 1.20, 1.33)	(1.00, 2.00, 1.50)	(1.20, 1.20, 1.20)	(1.82, 3.84, 2.76)
	GAM_6	(0.5, 0.7, 0.6)	(2.00, 1.50, 1.75)	(1.00, 2.22, 1.43)	(1.00, 1.00, 1.00)	(2.00, 4.56, 2.93)
Scenario I (alternative)	GAM_7	(0.5, 0.3, 0.4)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(0.50, 0.70, 0.60)	(0.75, 0.91, 0.84)
	GAM_8	(0.7, 0.5, 0.6)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(0.30, 0.50, 0.40)	(0.51, 0.75, 0.64)
	GAM_9	(0.6, 0.4, 0.5)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(0.40, 0.60, 0.50)	(0.64, 0.84, 0.75)
Scenario II (alternative)	GAM_{10}	(0.3, 0.3, 0.3)	(1.00, 2.00, 1.50)	(1.00, 1.00, 1.00)	(0.70, 1.40, 1.05)	(0.91, 2.24, 1.52)
	GAM_{11}	(0.7, 0.7, 0.7)	(1.00, 2.00, 1.50)	(1.00, 1.00, 1.00)	(0.30, 0.60, 0.45)	(0.51, 1.44, 0.92)
	GAM_{12}	(0.4, 0.6, 0.5)	(1.00, 2.50, 1.50)	(1.00, 1.00, 1.00)	(0.60, 1.00, 0.75)	(0.84, 2.50, 1.31)
Scenario III (alternative)	GAM_{13}	(0.3, 0.3, 0.3)	(1.50, 1.00, 1.25)	(1.00, 2.00, 1.50)	(1.05, 1.40, 1.31)	(1.52, 3.64, 2.71)
	GAM_{14}	(0.7, 0.7, 0.7)	(1.75, 1.25, 1.50)	(1.00, 2.00, 1.50)	(0.53, 0.75, 0.68)	(1.17, 2.81, 2.08)
	GAM_{15}	(0.6, 0.4, 0.5)	(2.00, 1.00, 1.50)	(1.00, 2.00, 1.50)	(0.80, 1.20, 1.13)	(1.76, 3.36, 2.95)

To evaluate the performance of the ELR test with respect to the choices of user-specified basis function $\mathbf{q}(x)$ in a DRM, we consider the following three scenarios that may be encountered in practice:

- Scenario I: all the distributions G_i's are homogenous, and thus any basis function is correctly specified under the models LN_1, LN_2, LN_7–LN_9, and GAM_1, GAM_2, GAM_7–GAM_9.

- Scenario II: the parameters a_i's are not equal while the parameters b_i's are held constant, and thus the basis function $\mathbf{q}(x) = \log(x)$ is correctly specified under the models LN_3, LN_4, LN_{10}–LN_{12}, and GAM_3, GAM_4, GAM_{10}–GAM_{12}.

- Scenario III: all the parameters a_i's and b_i's are not equal, and thus the basis function $\mathbf{q}(x) = \{\log(x), \log^2(x)\}^\top$ is correctly specified under the LN models LN_5, LN_6, LN_{13}–LN_{15}, and the basis function $\mathbf{q}(x) = \{x, \log(x)\}^\top$ is correctly specified under the GAM models GAM_5, GAM_6, GAM_{13}–GAM_{15}.

In the following comparisons, the simulation results are discussed by the above three scenarios.

4.4.1 Scenario I

The simulated type I error rates of the ELR, ATS, and WTPS under Scenario I are summarized in Tables 4.2–4.3, and the simulated power under the same scenario are plotted in Figure 4.1. Here, the DRM (4.2) is correctly specified with any form of $\mathbf{q}(x)$. After experimenting with several forms of $\mathbf{q}(x)$, the basis function $\mathbf{q}(x) = \log(x)$ is recommended since the ELR with such basis function has the most accurate type I error and the largest power. Hence we only present the results under the basis function $\mathbf{q}(x) = \log(x)$.

Based on our simulation results, our major observations for both the two- and three-sample comparisons are summarized as follows.

(a) It can be seen from the results in Tables 4.2–4.3, that the proposed ELR test and the WTPS method well control the type I error rates close to their nominal level. However, the ATS method tends to be conservative for equal sample sizes; and when the sample sizes are unequal it tends to be liberal for the two-sample comparisons.

Table 4.2: Scenario I: simulated probabilities (%) of rejecting H_0^* when data are generated from $\text{LN}(a_i, b_i)$ according to the parameter settings given in Table 4.1. Here the ELR is defined under basis function $\mathbf{q}(x) = \log(x)$.

Model	Two-sample comparison				Three-sample comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
LN_1	(50, 50)	5.10	4.05	4.85	(50, 50, 50)	5.01	3.03	4.97
	(100, 100)	4.83	4.84	5.15	(100, 100, 100)	5.10	3.77	4.75
	(50, 150)	5.10	5.85	4.92	(50, 150, 100)	5.01	5.00	5.06
	(150, 50)	4.94	5.91	4.84	(150, 50, 100)	5.14	4.75	5.08
LN_2	(50, 50)	4.96	3.68	5.01	(50, 50, 50)	4.77	2.45	5.13
	(100, 100)	5.07	4.04	4.73	(100, 100, 100)	4.87	3.01	4.98
	(50, 150)	4.74	5.78	3.99	(50, 150, 100)	5.20	4.67	4.97
	(150, 50)	4.77	6.15	4.25	(150, 50, 100)	5.15	4.79	5.02

NOTE: the Monte Carlo error is 0.218 (%) under the null models LN_1–LN_2.

Table 4.3: Scenario I: simulated probabilities (%) of rejecting H_0^* when data are generated from $\text{GAM}(a_i, b_i)$ according to the parameter settings given in Table 4.1. Here the ELR is defined under basis function $\mathbf{q}(x) = \log(x)$.

Model	Two-sample comparison				Three-sample comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
GAM_1	(50, 50)	5.16	4.68	4.82	(50, 50, 50)	5.47	4.34	4.99
	(100, 100)	4.98	4.56	4.84	(100, 100, 100)	5.17	4.62	5.09
	(50, 150)	4.83	5.89	5.18	(50, 150, 100)	5.28	4.82	4.94
	(150, 50)	5.13	5.74	5.08	(150, 50, 100)	5.32	5.11	5.07
GAM_2	(50, 50)	5.28	4.50	4.99	(50, 50, 50)	5.22	3.20	4.84
	(100, 100)	5.02	4.94	5.16	(100, 100, 100)	5.10	4.15	5.15
	(50, 150)	5.28	6.06	4.74	(50, 150, 100)	5.14	5.30	4.78
	(150, 50)	5.06	6.12	4.88	(150, 50, 100)	5.18	5.15	4.97

NOTE: the Monte Carlo error is 0.218 (%) under the null models GAM_1–GAM_2.

(b) In terms of power, it can be observed from Figure 4.1, that the performance of the proposed ELR test seems to be less sensitive to the sample sizes than the other two tests. Although there is no uniformly dominant method for all the settings in Figure 4.1, the ELR test seems to be the most powerful for equal sample sizes and unequal sample sizes with a negative pairing. For the unequal sample sizes with positive pairing, the WTPS method may have some advantage over the ELR test for the LN models, and thus it may be favourable in this scenario, while recalling that the ATS method may have inflated type I error.

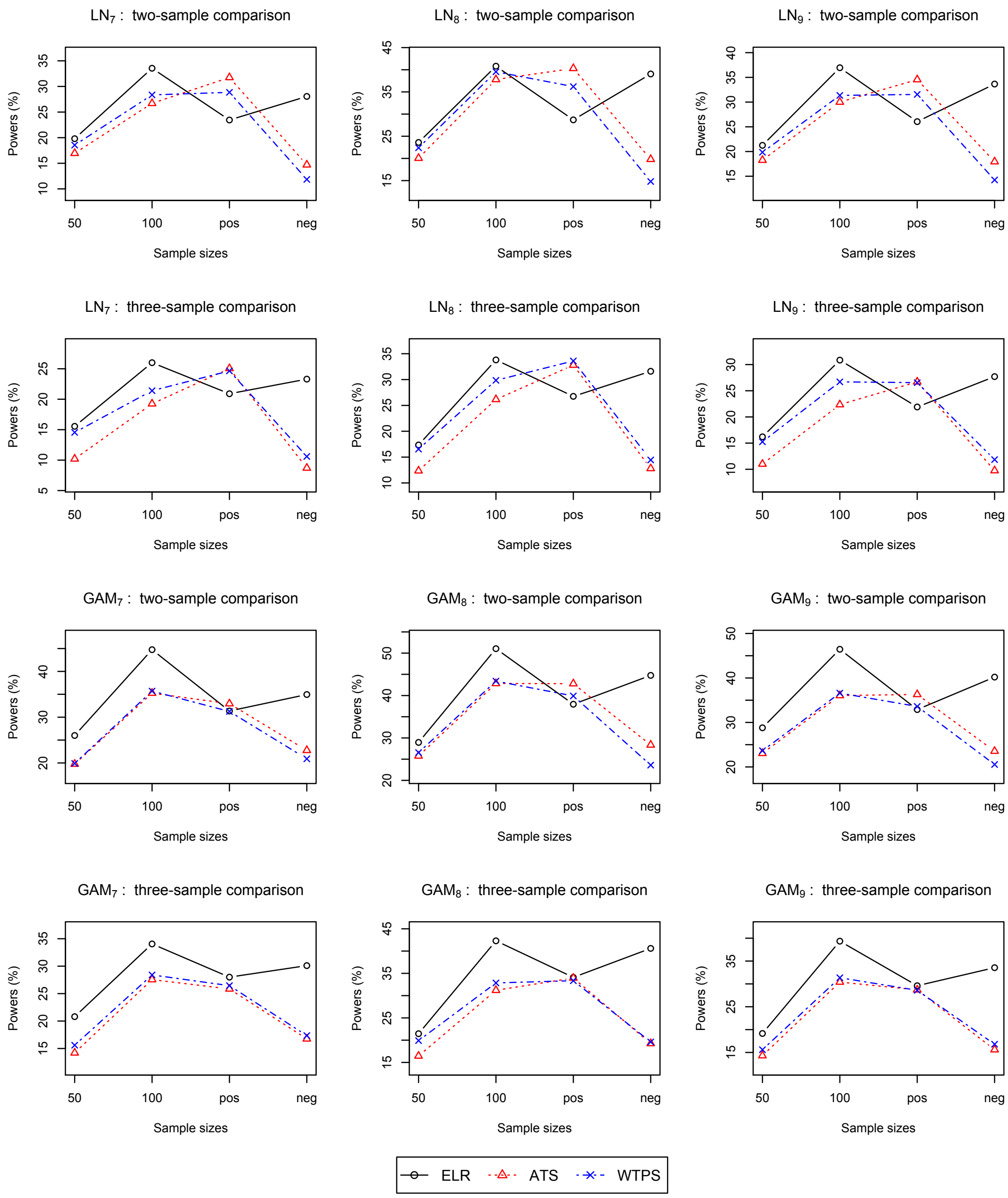

Figure 4.1: Scenario I: simulated power (%) of rejecting H_0^* at significance level 0.05 when data are generated from a log-normal mixture model with parameter settings given in Table 4.1. The ELR test is defined under $\mathbf{q}(x) = \log(x)$. The horizontal axis denotes combinations of sample sizes (n_0, n_1) equal to $(50, 50)$, $(100, 100)$, $(50, 150)$ and $(150, 50)$ for two-sample comparisons; and (n_0, n_1, n_2) equal to $(50, 50, 50)$, $(100, 100, 100)$, $(50, 150, 100)$ and $(150, 50, 100)$ for three-sample comparisons, from left to right.

4.4.2 Scenario II

The simulated type I error rates for Scenario II are summarized in Tables 4.4–4.5, and the simulated power for Scenario II are plotted in Figure 4.2.

Table 4.4: Scenario II: simulated probabilities (%) of rejecting H_0^* when data are generated from $\text{LN}(a_i, b_i)$ according to the parameter settings given in Table 4.1. Here the ELR is defined under correctly specified basis function $\mathbf{q}(x) = \log(x)$.

Model	Two-sample comparison				Three-sample comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
LN_3	$(50, 50)$	5.27	4.26	5.08	$(50, 50, 50)$	5.01	3.29	4.97
	$(100, 100)$	5.12	5.31	5.74	$(100, 100, 100)$	5.19	3.74	5.08
	$(50, 150)$	5.13	5.48	4.32	$(50, 150, 100)$	5.37	4.17	4.37
	$(150, 50)$	5.11	7.37	6.22	$(150, 50, 100)$	5.28	6.27	6.51
LN_4	$(50, 50)$	5.19	5.16	5.53	$(50, 50, 50)$	5.10	3.60	4.87
	$(100, 100)$	5.34	4.89	5.42	$(100, 100, 100)$	5.41	3.85	5.79
	$(50, 150)$	5.42	4.82	3.06	$(50, 150, 100)$	5.23	3.61	3.67
	$(150, 50)$	5.36	8.88	7.52	$(150, 50, 100)$	5.47	7.18	7.67

NOTE: the Monte Carlo error is 0.218 (%) under the null models LN_3–LN_4.

Table 4.5: Scenario II: simulation probabilities (%) of rejecting H_0^* when data are generated from $\text{GAM}(a_i, b_i)$ according to the parameter settings given in Table 4.1. Here the ELR is defined under correctly specified basis function $\mathbf{q}(x) = \log(x)$.

Model	Two-sample comparison				Three-sample comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
GAM_3	$(50, 50)$	5.00	4.64	4.83	$(50, 50, 50)$	5.36	4.41	4.96
	$(100, 100)$	4.84	4.57	4.69	$(100, 100, 100)$	5.16	4.79	5.10
	$(50, 150)$	5.03	5.60	5.14	$(50, 150, 100)$	4.91	4.84	4.85
	$(150, 50)$	5.11	5.43	5.16	$(150, 50, 100)$	5.09	5.10	4.83
GAM_4	$(50, 50)$	5.10	5.21	5.39	$(50, 50, 50)$	5.04	4.29	4.91
	$(100, 100)$	5.28	5.26	5.33	$(100, 100, 100)$	5.25	4.82	5.10
	$(50, 150)$	5.16	5.02	4.38	$(50, 150, 100)$	5.30	4.84	4.69
	$(150, 50)$	4.92	6.05	5.58	$(150, 50, 100)$	4.93	5.65	5.29

NOTE: the Monte Carlo error is 0.218 (%) under the null models GAM_3–GAM_4.

Based on our simulation results, our major observations for both the two- and three-sample comparisons are summarized as follows.

(c) In terms of type I error control, it can be seen from the results in Tables 4.4–4.5, that the ELR test under $\mathbf{q}(x) = \log(x)$ always retains error rates close to the nominal level, and the results seem insensitive to whether sample sizes are equal or not. On the other hand, the type I error rates of both ATS and WTPS methods seem to be sensitive to whether the sample sizes are equal or not. For equal sample sizes, both ATS and WTPS methods control the type I error satisfactory, although the ATS method may be conservative in some settings. For unequal sample sizes with positive pairing, both the ATS and WTPS methods are conservative in their type I error rates; however, with the negative pairing, they tend to have inflated type I error rates, particularly for the LN models.

(d) In terms of power, it can be observed from Figure 4.2, that the performance of the proposed ELR test is, again, not as sensitive to unequal sample sizes as the other two tests. Further, in all the settings in Figure 4.2, the ELR test is the most, or one of the most, powerful tests, for both the equal and unequal sample sizes under comparisons. In some special cases, the gain in power could be over 50%. Together with the observations from the type I error rates, the proposed ELR test under $\mathbf{q}(x) = \log(x)$ is much preferred in this scenario.

4.4.3 Scenario III

The simulated type I error rates for Scenario III are summarized in Tables 4.6–4.7, and the simulated power for Scenario III are plotted in Figure 4.3.

Based on our simulation results, our major observations for both two-sample and three-sample comparisons are summarized as follows.

(e) In terms of type I error control, it can be seen from the results in Tables 4.6–4.7 that the ELR test, under basis functions $\mathbf{q}(x) = \{\log(x), \log^2(x)\}^\top$ for LN models, or $\mathbf{q}(x) = \{x, \log(x)\}^\top$ for GAM models, may fail to keep the error under control, except for the unequal sample sizes with positive pairing. In this scenario, the WTPS method may also lead to inflated type I error rates in some settings. On the other hand, the ATS method seems to have overall good control of the type I error, except for the case of unequal sample sizes with negative pairing. Indeed, all three methods

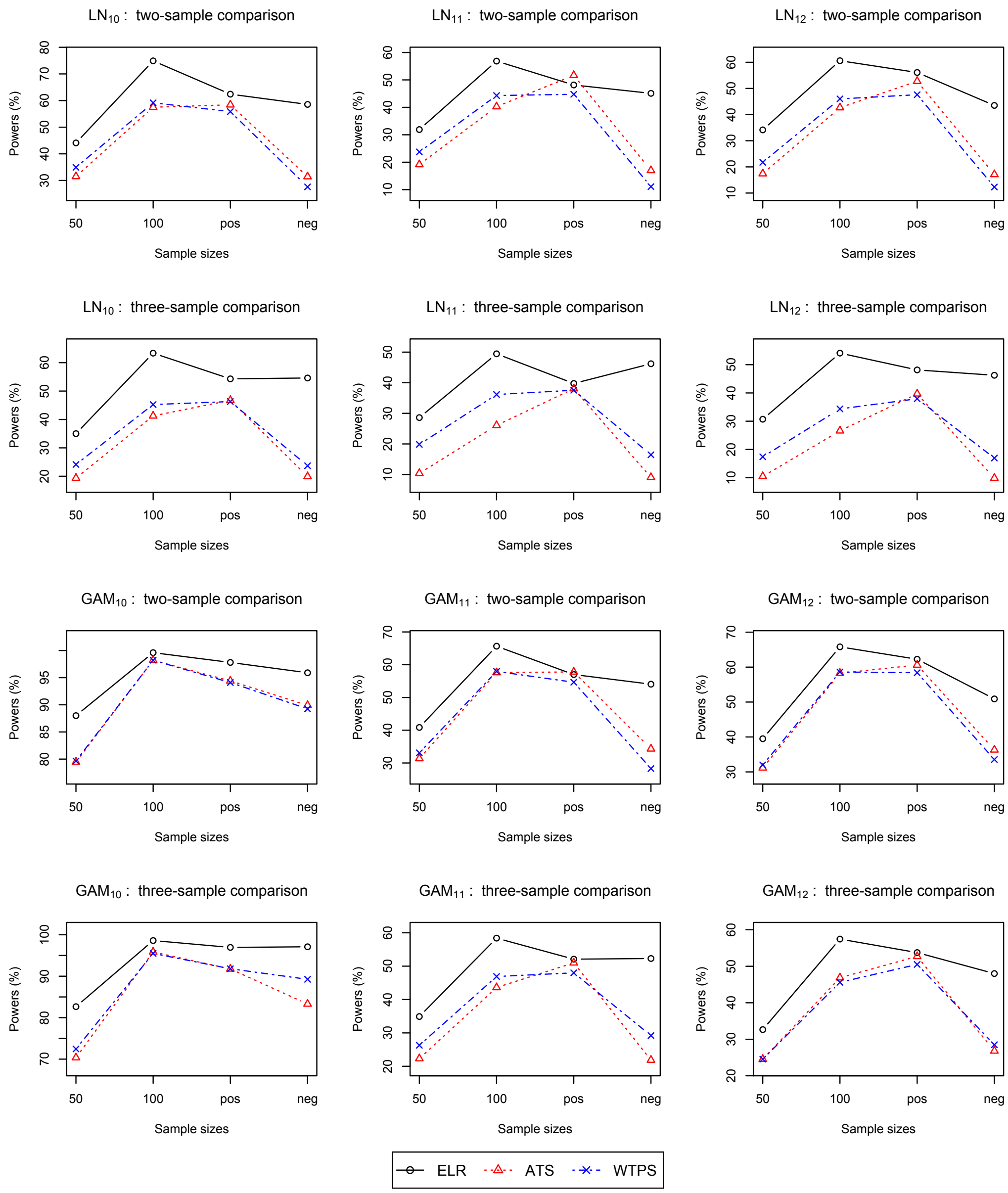

Figure 4.2: Scenario II: simulated power (%) of rejecting H_0^* at significance level 0.05 when data are generated from a log-normal mixture model with parameter settings given in Table 4.1. The ELR test is defined under $\mathbf{q}(x) = \log(x)$. The horizontal axis denotes combinations of sample sizes (n_0, n_1) equal to $(50, 50)$, $(100, 100)$, $(50, 150)$ and $(150, 50)$ for two-sample comparisons; and (n_0, n_1, n_2) equal to $(50, 50, 50)$, $(100, 100, 100)$, $(50, 150, 100)$ and $(150, 50, 100)$ for three-sample comparisons, from left to right.

Table 4.6: Scenario III: simulation probabilities (%) of rejecting H_0^* when data are generated from $\text{LN}(a_i, b_i)$ according to the parameter settings given in Table 4.1. Here the ELR is defined under correctly specified basis function $\mathbf{q}(x) = \{\log(x), \log^2(x)\}^\top$.

Model	Two-sample comparison				Three-sample comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
LN_5	$(50, 50)$	6.00	4.97	5.53	$(50, 50, 50)$	5.82	3.89	5.89
	$(100, 100)$	5.64	5.17	5.49	$(100, 100, 100)$	5.44	4.11	5.41
	$(50, 150)$	5.28	4.49	3.31	$(50, 150, 100)$	4.93	3.58	3.68
	$(150, 50)$	6.05	8.11	7.16	$(150, 50, 100)$	5.91	6.16	6.40
LN_6	$(50, 50)$	5.68	4.74	5.57	$(50, 50, 50)$	5.36	3.43	5.33
	$(100, 100)$	5.48	5.22	5.79	$(100, 100, 100)$	5.33	3.90	5.62
	$(50, 150)$	4.91	4.70	3.10	$(50, 150, 100)$	5.23	3.85	3.73
	$(150, 50)$	5.79	8.19	6.72	$(150, 50, 100)$	6.10	7.24	7.08

NOTE: the Monte Carlo error is 0.218 (%) under the null models LN_5–LN_6.

Table 4.7: Scenario III: simulation probabilities (%) of rejecting H_0^* when data are generated from $\text{GAM}(a_i, b_i)$ according to the parameter settings given in Table 4.1. Here the ELR is defined under correctly specified basis function $\mathbf{q}(x) = \{x, \log(x)\}^\top$.

Model	Two-sample comparison				Three-sample comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
GAM_5	$(50, 50)$	5.86	5.18	5.29	$(50, 50, 50)$	5.76	4.75	5.44
	$(100, 100)$	5.40	4.92	5.03	$(100, 100, 100)$	5.21	4.61	4.85
	$(50, 150)$	5.15	4.92	4.17	$(50, 150, 100)$	5.20	4.69	4.32
	$(150, 50)$	5.76	6.20	6.01	$(150, 50, 100)$	5.59	5.96	5.83
GAM_6	$(50, 50)$	5.91	5.46	5.63	$(50, 50, 50)$	5.93	4.60	5.53
	$(100, 100)$	5.82	5.51	5.61	$(100, 100, 100)$	5.46	4.99	5.50
	$(50, 150)$	5.42	4.96	3.90	$(50, 150, 100)$	5.22	4.45	4.02
	$(150, 50)$	6.24	7.15	6.67	$(150, 50, 100)$	5.83	6.96	6.53

NOTE: the Monte Carlo error is 0.218 (%) under the null models GAM_5–GAM_6.

under comparison lead to inflated type I error rates when the sample sizes are unequal with negative pairing. In general, for the cases of equal sample sizes and unequal sample sizes with positive pairing, the ATS is a suitable testing method in this scenario. In this scenario with more complex data distributions, the ELR and WTPS methods seem to need larger sample sizes than those considered to give adequate approximations.

(f) In terms of power, it can be observed from Figure 4.3 that the ATS method still has consistent performance in the settings of equal sample sizes and unequal sample sizes

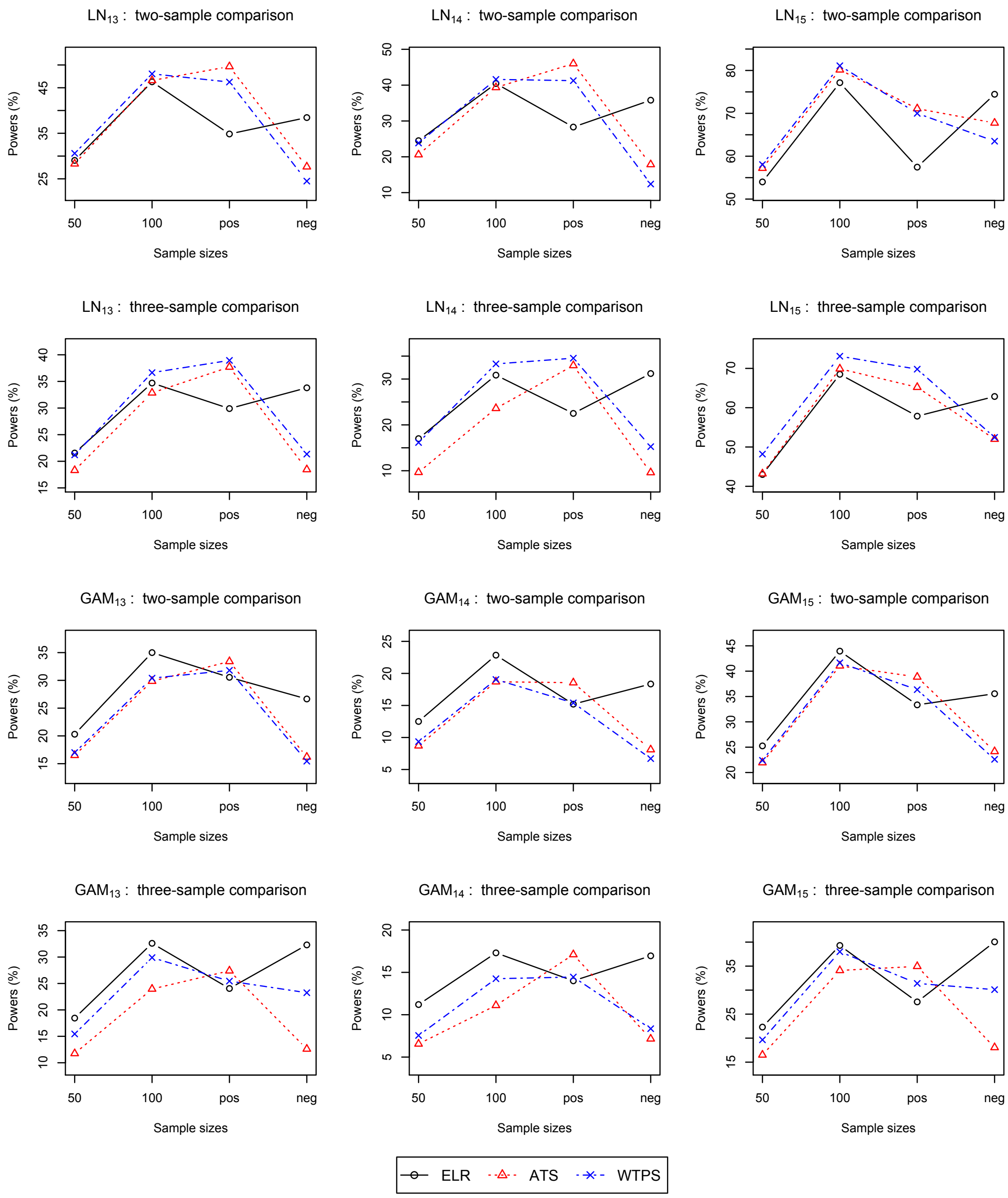

Figure 4.3: Scenario III: simulated power (%) of rejecting H_0^* at significance level 0.05 when data are generated from a log-normal mixture model with parameter settings given in Table 4.1. The ELR test is defined under $\mathbf{q}(x) = \{\log(x), \log^2(x)\}^\top$. The horizontal axis denotes combinations of sample sizes (n_0, n_1) equal to $(50, 50)$, $(100, 100)$, $(50, 150)$ and $(150, 50)$ for two-sample comparisons; and (n_0, n_1, n_2) equal to $(50, 50, 50)$, $(100, 100, 100)$, $(50, 150, 100)$ and $(150, 50, 100)$ for three-sample comparisons, from left to right.

with positive pairing. However, keep in mind that the ELR and WTPS methods both may fail to control the type I error rates. Therefore, no fair comparison can be made with ELR and WTPS methods.

In summary, for testing mean equality in the three scenarios considered, the proposed ELR test is robust against the assumption of parametric models that generate the data. The ELR test generally performs best in Scenario II with basis function $\mathbf{q}(x) = \log(x)$ correctly specified in a DRM. As we may expect, the failure of selecting an appropriate form of $\mathbf{q}(x)$ in a DRM may affect the control of type I error of the ELR test in some settings, and potential convergence problem in computation can also be an issue.

In our simulation studies, the numerical procedure for calculating ELR, discussed in Section 4.3, converges fast and is in general not sensitive to choices of initial values when the basis function $\mathbf{q}(x) = \log(x)$ is used in Scenarios I and II. However, when the basis function of increasing dimension are specified in Scenario III, we found that the computation may not be very stable and a good choice of initial value can be helpful. Besides, in our experience, the ELR with correctly specified basis function is also important for convergence in Scenarios II and III.

In practice, a comprehensive DRM selection strategy, as described in Section 2.3, would be encouraged at a preliminary stage of data analysis. We comment that such a preliminary DRM selection procedure is not too restrictive. Given the difficulty in identifying a suitable parametric model for many practical situations, the proposed ELR test can still be an attractive and robust semiparametric alternative approach.

Combining the conclusions of our simulation results and numerical experience, we recommend using the ELR test when the basis function $\mathbf{q}(x) = \log(x)$ is selected in the DRM.

We also comment that for the three-sample comparisons, the unequal sample sizes combinations presented in this section correspond to two extreme cases that are perfectly positive, or negative, pairing with the unequal variances. For these cases with positive and negative parings, the findings for ATS and WTPS methods are consistent with those in Vallejo et al. (2010) and Pauly et al. (2015). We do not report all possible patterns of unequal sample sizes and variances combinations in simulation. For other combinations that are somewhat in between those reported, the pattern of results do not exhibit large difference from those reported. In general, the ELR test has been observed to be more robust against unequal sample sizes when compared with the ATS and WTPS methods. In the next section, we examine a real data example in which the relation between sample sizes and variances are not dramatically positively or negatively correlated.

4.5 An illustrative real data example

In this section, we analyze a real data set from Koopmans (1981, p. 107). It arises in a biological study of the seasonal activity patterns of a species of field mice. The measurements are the average distances (in meters) travelled between captures by those mice at least twice in a given month. One of the objectives in this study is to discover if the mean measurements differ between the four seasons. In addition to the continuous positive measurements, there are substantial proportions of zero values, especially in the fall and winter data. Some summary statistics are:

- the sample estimate of $\boldsymbol{\nu}^\top$ is $(0.1765, 0.1111, 0.3704, 0.2941)$ with sample sizes $(17, 27, 27, 34)$;
- the sample means are $(26.9412, 30.8148, 13.0370, 15.0294)$;
- the sample variances are $(679.3088, 1118.926, 178.8832, 293.5446)$.

As discussed in Section 4.4, we need to select a basis function $\mathbf{q}(x)$ in a DRM that provides a reasonable fit to this data. We apply the AIC, discussed in Section 2.3, to select a basis function in the DRM for the positive data in this example. The five candidate basis functions are guided by Table 2.1. The results are given in Table 4.8. It can be seen that the DRM with $\mathbf{q}(x) = \log(x)$ has the smallest AIC among five commonly used basis functions, and hence it is recommended in this example. Furthermore, we have also applied the ELR test for homogeneity in distributions developed in Chapter 3 to this real data set. The bootstrap ELR test for homogeneity based on $\mathbf{q}(x) = \log(x)$ gives a p-value of 0.0402, with $10,000$ times bootstrap resampling. With this preliminary data analysis, it seems reasonable to categorize this real data example into the Scenario II considered in Section 4.4.2.

Table 4.8: AIC for five commonly used basis functions $\mathbf{q}(x)$ in a DRM for the positive field mice data.

$\mathbf{q}(x) =$	$\log(x)$	x	$\{\log(x), \log^2(x)\}^\top$	$\{x, \log(x)\}^\top$	$\{x, \log(x), \log^2(x)\}^\top$
AIC	219.63	220.45	224.19	224.05	228.86

We further fit this data by log-normal mixture models and gamma mixture models, under the null and alternative hypotheses, by the parametric maximum likelihood method. Here the null hypothesis is that the means of measurements for all four seasons are the same. The details are provided in Table 4.9. These fitted models will be used in our confirmative simulation later on.

Table 4.9: Fitted parameters for log-normal mixture model and gamma mixture model under the null and alternative hypotheses for field mice data. The models LN_{16} and GAM_{16} are fitted under the null hypothesis; and the models LN_{17} and GAM_{17} are fitted under the alternative hypothesis. The last two columns are the means and variances corresponding to each model, and $\mathbf{1}^\top = (1, 1, 1, 1)$.

Model	$(\nu_0, \nu_1, \nu_2, \nu_3)$	(a_0, a_1, a_2, a_3)	(b_0, b_1, b_2, b_3)	Means	Variances
LN_{16}	(0.28, 0.21, 0.28, 0.23)	(3.11, 3.03, 3.12, 3.05)	$0.41 \times \mathbf{1}^\top$	(19.95, 19.95, 19.95, 19.95)	(430.69, 362.72, 433.63, 376.22)
LN_{17}	(0.18, 0.11, 0.37, 0.29)	(3.29, 3.25, 2.91, 2.87)	$0.37 \times \mathbf{1}^\top$	(26.68, 27.49, 13.84, 14.99)	(539.26, 474.92, 248.75, 235.92)
GAM_{16}	(0.27, 0.20, 0.28, 0.23)	(2.32, 2.13, 2.35, 2.20)	$12.22 \times \mathbf{1}^\top$	(20.68, 20.68, 20.68, 20.68)	(410.53, 362.76, 417.95, 381.00)
GAM_{17}	(0.18, 0.11, 0.37, 0.29)	(2.88, 2.77, 2.10, 2.04)	$11.24 \times \mathbf{1}^\top$	(26.65, 27.67, 14.87, 16.22)	(451.68, 406.71, 297.12, 291.97)

We then applied the proposed ELR test, and other testing methods discussed in Section 4.4, for the mean equality in this real data. The observed test statistics and their corresponding p-values are reported in Table 4.10.

From the results in Table 4.10, the proposed ELR test as well as the other two tests all produce significant p-values at 5% significance level. It is worth emphasizing that the proposed ELR test provides the strongest evidence. Therefore, a nature followup concern is

Table 4.10: Test statistics, corresponding p-values, and confirmative simulation for field mice data. The ELR test is under basis function $\mathbf{q}(x) = \log(x)$ in a DRM.

Pair	Method	Field mice data		Confirmative simulation			
		Test statistic	p-value	LN_{16}	LN_{17}	GAM_{16}	GAM_{17}
All four seasons	ELR	12.40	0.0061	5.20	86.36	5.21	79.11
	ATS	3.17	0.0435	3.92	68.11	3.99	62.07
	WTPS	9.71	0.0388	4.83	76.58	4.67	66.88
Spring vs. Autumn	ELR	5.65	0.0175	5.05	59.35	5.29	53.19
	ATS	4.15	0.0542	4.36	47.32	4.61	42.50
	WTPS	4.15	0.0431	4.56	49.07	4.61	43.05
Spring vs. Winter	ELR	4.95	0.0261	4.91	57.24	5.23	45.97
	ATS	2.92	0.1009	4.69	42.15	4.90	34.24
	WTPS	2.92	0.0991	4.62	43.61	4.80	34.56
Summer vs. Autumn	ELR	6.56	0.0105	5.16	79.56	5.29	74.43
	ATS	6.58	0.0149	4.40	74.16	4.72	68.31
	WTPS	6.58	0.0071	4.88	74.92	4.91	68.41
Summer vs. Winter	ELR	6.62	0.0101	5.37	78.91	5.31	70.61
	ATS	4.98	0.0319	4.50	71.83	4.77	62.69
	WTPS	4.98	0.0262	4.80	72.96	4.92	62.91

to detect any potential pairwise mean differences. We further test for the two-sample mean equality for six pairwise combinations of this data. The testing results for the significant pairs are given in Table 4.10. We particularly highlight one pairwise comparison, that is, Spring vs. Winter. The proposed ELR test in this two-sample mean comparison gives a significant p-value of 0.0261 at the 5% significance level. In comparison, both the ATS and WTPS methods fail to detect the mean difference for this particular pair at the 5% significance level.

These results may be further verified by the simulation according to model settings in Table 4.9. The simulation results based on 10,000 repetitions are summarized in Table 4.10. It can be seem that all the simulation results demonstrate very good agreement with our real data analysis.

4.6 Proofs

This section contains a proof of Theorem 4.1. The regularity conditions provided in Section 4.2.3 will be used throughout this proof. We first define some useful notation before we proceed.

We emphasize that the definition of R_n in Section 4.3 involves the summation of random variables over random numbers, i.e., the numbers of zero and positive observations are random due to the mixture data structure. This fact prevents us from applying standard large sample theories directly. In our proofs, we use indicator variables to circumvent this technical difficulty. Let $\sum_{ij}$ denote the summation over the constant range of all n pooled observations (unless is otherwise specified), that is, $j \in \{1, \ldots, n_i\}$ and $i \in \{0, 1, \ldots, m\}$. For notational simplicity, let $\mathbf{Q}(x_{ij}) = \{1, \mathbf{q}^\top(x_{ij})\}^\top$, so that $\omega_r(x_{ij}; \boldsymbol{\theta}_r) = \exp\{\boldsymbol{\theta}_r^\top \mathbf{Q}(x_{ij})\}$, for $r = 0, \ldots, m$, where we have set $\boldsymbol{\theta}_0 = \mathbf{0}$.

Recall that $\boldsymbol{\eta} = (\boldsymbol{\nu}^\top, \boldsymbol{\theta}^\top)^\top$. Note that the function $\ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t})$ defined in Section 4.3 can be rewritten as

$$\begin{aligned}\ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t}) \;=\;& \sum_{i=0}^{m} \log\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\} + \sum_{ij} \left\{\boldsymbol{\theta}_i^\top \mathbf{Q}(x_{ij}) I(x_{ij} > 0)\right\} \\ & - \sum_{ij} \log\left[1 + \boldsymbol{\lambda}^\top\{\boldsymbol{\omega}(x_{ij}; \boldsymbol{\theta}) - \mathbf{1}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\eta})\right] I(x_{ij} > 0).\end{aligned}$$

For simplicity, we have used $\ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})$ to denote $\ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t})$, and $\mathbf{g}(x; \boldsymbol{\eta})$ to denote $\mathbf{g}(x; \boldsymbol{\nu}, \boldsymbol{\theta})$.

Suppose the true value of $\boldsymbol{\eta}$ is $\boldsymbol{\eta}^* = (\boldsymbol{\nu}^{*\top}, \boldsymbol{\theta}^{*\top})^\top$. Define

$$\lambda_r^* = \frac{\rho_r^*(1-\nu_r^*)}{\sum_{i=0}^m \rho_i^*(1-\nu_i^*)}, \qquad r = 1, \ldots, m.$$

Let $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_m^*)^\top$ and $\mathbf{t}^* = \mathbf{0}$. We comment that $\boldsymbol{\lambda}^*$ and $\mathbf{t}^*$ together have the following property: $\boldsymbol{\lambda}^*$ and $\mathbf{t}^*$ are the solution of

$$E\left\{\frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}, \mathbf{t})}{\partial \boldsymbol{\lambda}}\right\} = \mathbf{0}, \qquad E\left\{\frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}, \mathbf{t})}{\partial \mathbf{t}}\right\} = \mathbf{0}.$$

They play the role of "true values" of unknown parameters and hence are called the true values of $\boldsymbol{\lambda}$ and $\mathbf{t}$.

Recall that $\boldsymbol{\psi} = (\boldsymbol{\eta}^\top, \boldsymbol{\lambda}^\top, \mathbf{t}^\top)^\top$, and from Proposition 4.1, the estimator $\tilde{\boldsymbol{\psi}}$ of $\boldsymbol{\psi}$ under H_0 is the solution of

$$\frac{\partial \ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})}{\partial \boldsymbol{\psi}} = \mathbf{0}.$$

Recall that $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\nu}}^\top, \hat{\boldsymbol{\theta}}^\top)^\top$ is the maximum empirical likelihood estimator of $\boldsymbol{\eta}$ under the alternative model. Let $\hat{\boldsymbol{\lambda}}$ be the Lagrange multiplier $\boldsymbol{\lambda}$ correponding to $\hat{\boldsymbol{\eta}}$. Similarly to the arguments in Proposition 4.1, we have that $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\lambda}}$ are the solution of

$$\frac{\partial \ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{0})}{\partial \boldsymbol{\eta}} = \mathbf{0}, \qquad \frac{\partial \ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{0})}{\partial \boldsymbol{\lambda}} = \mathbf{0}.$$

Then the empirical likelihood ratio test statistic R_n can be rewritten as

$$R_n = 2\left\{\ell(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\lambda}}, \mathbf{0}) - \ell(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}})\right\}. \tag{4.19}$$

Next we apply the second-order Taylor expansion on $\ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})$ around $\boldsymbol{\psi} = \boldsymbol{\psi}^*$ with $\boldsymbol{\psi}^* = (\boldsymbol{\eta}^{*\top}, \boldsymbol{\lambda}^{*\top}, \mathbf{0}^\top)^\top$ to find the quadratic approximation of R_n, in which we need the first and second derivatives of $\ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})$ at $\boldsymbol{\psi} = \boldsymbol{\psi}^*$.

First derivatives of $\ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})$

We calculate the first derivatives of $\ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})$ as follows, for $r = 1, \ldots, m$ and $s = 1, \ldots, p$.

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})}{\partial \nu_0} &= \frac{n_{00}}{\nu_0} - \frac{n_{01}}{1-\nu_0} - \sum_{ij} \frac{\mathbf{t}^\top \partial \mathbf{g}(x_{ij}; \boldsymbol{\eta})/\partial \nu_0}{1 + \boldsymbol{\lambda}^\top \{\boldsymbol{\omega}(x_{ij}; \boldsymbol{\theta}) - \mathbf{1}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\eta})} I(x_{ij} > 0), \\
\frac{\partial \ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})}{\partial \nu_r} &= \frac{n_{r0}}{\nu_r} - \frac{n_{r1}}{1-\nu_r} - \sum_{ij} \frac{\mathbf{t}^\top \partial \mathbf{g}(x_{ij}; \boldsymbol{\eta})/\partial \nu_r}{1 + \boldsymbol{\lambda}^\top \{\boldsymbol{\omega}(x_{ij}; \boldsymbol{\theta}) - \mathbf{1}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\eta})} I(x_{ij} > 0), \\
\frac{\partial \ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})}{\partial \boldsymbol{\theta}_r} &= \sum_{j=1}^{n_r} \mathbf{Q}(x_{rj}) I(x_{rj} > 0) - \sum_{ij} \frac{\lambda_r \mathbf{Q}(x_{ij}) \omega_r(x_{ij}; \boldsymbol{\theta}_r) + \mathbf{t}^\top \partial \mathbf{g}(x_{ij}; \boldsymbol{\eta})/\partial \boldsymbol{\theta}_r}{1 + \boldsymbol{\lambda}^\top \{\boldsymbol{\omega}(x_{ij}; \boldsymbol{\theta}) - \mathbf{1}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\eta})} I(x_{ij} > 0), \\
\frac{\partial \ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})}{\partial \lambda_r} &= -\sum_{ij} \frac{\omega_r(x_{ij}; \boldsymbol{\theta}_r) - 1}{1 + \boldsymbol{\lambda}^\top \{\boldsymbol{\omega}(x_{ij}; \boldsymbol{\theta}) - \mathbf{1}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\eta})} I(x_{ij} > 0), \\
\frac{\partial \ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})}{\partial t_s} &= -\sum_{ij} \frac{g_s(x_{ij}; \boldsymbol{\eta})}{1 + \boldsymbol{\lambda}^\top \{\boldsymbol{\omega}(x_{ij}; \boldsymbol{\theta}) - \mathbf{1}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\eta})} I(x_{ij} > 0).
\end{aligned}$$

For a compact presentation, we introduce additional notation. Denote $\lambda_0^* = 1 - \sum_{i=1}^m \lambda_i^*$ and $h(x; \boldsymbol{\eta}^*) = \sum_{i=0}^m \lambda_i^* \omega_i(x; \boldsymbol{\theta}_i^*)$. Define

$$h_r(x; \boldsymbol{\eta}^*) = \lambda_r^* \omega_r(x; \boldsymbol{\theta}_r^*)/h(x; \boldsymbol{\eta}^*), \qquad r = 0, 1, \ldots, m.$$

Notice a useful fact that $\sum_{i=0}^m h_i(x; \boldsymbol{\eta}^*) = 1$. Further, we denote $\Delta^* = \sum_{i=0}^m \rho_i^*(1 - \nu_i^*)$, and hence $\lambda_r^* = \rho_r^*(1 - \nu_r^*)/\Delta^*$, for $r = 0, \ldots, m$. For a compact notation, let $\mathbf{h}(x_{ij}; \boldsymbol{\eta}^*) = \{h_1(x_{ij}; \boldsymbol{\eta}^*), \ldots, h_m(x_{ij}; \boldsymbol{\eta}^*)\}^\top$,

Define

$$\mathbf{S}_n = \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \boldsymbol{\psi}} = \begin{pmatrix} \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \boldsymbol{\nu}} \\ \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \boldsymbol{\theta}} \\ \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \boldsymbol{\lambda}} \\ \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \mathbf{t}} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_{n1} \\ \mathbf{S}_{n2} \\ \mathbf{S}_{n3} \\ \mathbf{S}_{n4} \end{pmatrix} \tag{4.20}$$

where $\mathbf{S}_{n1}$ is a $(m+1)$-dimensional vector, $\mathbf{S}_{n2}, \mathbf{S}_{n3}$ are m-dimensional vectors and $\mathbf{S}_{n4}$ is a p-dimensional vector with corresponding entries as follows:

$$\begin{aligned}
\mathbf{S}_{n1} &= \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \boldsymbol{\nu}} = \left(\frac{n_{00}}{\nu_0^*} - \frac{n_{01}}{1 - \nu_0^*}, \ldots, \frac{n_{m0}}{\nu_m^*} - \frac{n_{m1}}{1 - \nu_m^*} \right)^\top, \\
\mathbf{S}_{n2,r} &= \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \boldsymbol{\theta}_r} = \sum_{j=1}^{n_r} \mathbf{Q}(x_{rj}) I(x_{ij} > 0) - \sum_{ij} \mathbf{Q}(x_{ij}) h_r(x_{ij}; \boldsymbol{\eta}^*) I(x_{ij} > 0), \\
\mathbf{S}_{n3,r} &= \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \lambda_r} = -\sum_{ij} \frac{\omega_r(x_{ij}; \boldsymbol{\theta}_r^*) - 1}{h(x_{ij}; \boldsymbol{\eta}^*)} I(x_{ij} > 0), \\
\mathbf{S}_{n4,s} &= \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial t_s} = -\sum_{ij} \frac{g_s(x_{ij}; \boldsymbol{\eta}^*)}{h(x_{ij}; \boldsymbol{\eta}^*)} I(x_{ij} > 0),
\end{aligned}$$

for $r = 1, \ldots, m$ and $s = 1, \ldots, p$.

Second derivatives of $\ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})$

We next calculate the second derivatives of $\ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})$ and evaluate them at the true value.

We have

$$\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top} = \begin{pmatrix} \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\nu}\partial\boldsymbol{\nu}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\nu}\partial\boldsymbol{\theta}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\nu}\partial\boldsymbol{\lambda}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\nu}\partial\mathbf{t}^\top} \\ \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\nu}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\lambda}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\theta}\partial\mathbf{t}^\top} \\ \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\nu}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\theta}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\lambda}\partial\mathbf{t}^\top} \\ \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\mathbf{t}\partial\boldsymbol{\nu}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\mathbf{t}\partial\boldsymbol{\theta}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\mathbf{t}\partial\boldsymbol{\lambda}^\top} & \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\mathbf{t}\partial\mathbf{t}^\top} \end{pmatrix}$$

where

$$\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\nu}\partial\boldsymbol{\nu}^\top} &= \operatorname{diag}\left\{-\frac{n_{00}}{\nu_0^{*2}} - \frac{n_{01}}{(1-\nu_0^*)^2}, \ldots, -\frac{n_{m0}}{\nu_m^{*2}} - \frac{n_{m1}}{(1-\nu_m^*)^2}\right\}, \\
\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\nu}\partial\boldsymbol{\theta}^\top} &= \left\{\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\nu}^\top}\right\}^\top = \mathbf{0}, \\
\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\nu}\partial\boldsymbol{\lambda}^\top} &= \left\{\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\nu}^\top}\right\}^\top = \mathbf{0}, \\
\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\mathbf{t}\partial\boldsymbol{\nu}^\top} &= \left\{\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\nu}\partial\mathbf{t}^\top}\right\}^\top = -\sum_{ij} \frac{\partial\mathbf{g}(x_{ij};\boldsymbol{\eta}^*)/\partial\boldsymbol{\nu}^\top}{h(x_{ij};\boldsymbol{\eta}^*)} I(x_{ij} > 0), \\
\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} &= -\sum_{ij} \Big\{ \Big[\{\operatorname{diag}\{\mathbf{h}(x_{ij};\boldsymbol{\eta}^*)\} - \{\mathbf{h}(x_{ij};\boldsymbol{\eta}^*)\mathbf{h}^\top(x_{ij};\boldsymbol{\eta}^*)\}\Big] \otimes \{\mathbf{Q}(x_{ij})\mathbf{Q}^\top(x_{ij})\}\Big\} I(x_{ij} > 0), \\
\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\lambda}^\top} &= \left\{\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\theta}^\top}\right\}^\top \\
&= -\sum_{ij} \left\{ \left[\frac{\operatorname{diag}\{\boldsymbol{\omega}(x_{ij};\boldsymbol{\theta}^*)\} - \mathbf{h}(x_{ij};\boldsymbol{\eta}^*)\{\boldsymbol{\omega}(x_{ij};\boldsymbol{\theta}^*) - \mathbf{1}\}^\top}{h(x_{ij};\boldsymbol{\eta}^*)}\right] \otimes \mathbf{Q}(x_{ij})\right\} I(x_{ij} > 0), \\
\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\theta}\partial\mathbf{t}^\top} &= \left\{\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\mathbf{t}\partial\boldsymbol{\theta}^\top}\right\}^\top \\
&= -\sum_{ij} \frac{\partial\mathbf{g}^\top(x_{ij},\boldsymbol{\eta}^*)/\partial\boldsymbol{\theta} - \{\mathbf{h}(x_{ij};\boldsymbol{\theta}^*)\mathbf{g}^\top(x_{ij},\boldsymbol{\eta}^*)\} \otimes \mathbf{Q}(x_{ij})}{h(x_{ij};\boldsymbol{\eta}^*)} I(x_{ij} > 0), \\
\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top} &= -\sum_{ij} \frac{-\{\boldsymbol{\omega}(x_{ij};\boldsymbol{\theta}^*) - \mathbf{1}\}\{\boldsymbol{\omega}(x_{ij};\boldsymbol{\theta}^*) - \mathbf{1}\}^\top}{h(x_{ij};\boldsymbol{\eta}^*)^2} I(x_{ij} > 0), \\
\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\boldsymbol{\lambda}\partial\mathbf{t}^\top} &= \left\{\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\mathbf{t}\partial\boldsymbol{\lambda}^\top}\right\} = -\sum_{ij} \frac{-\{\boldsymbol{\omega}(x_{ij};\boldsymbol{\theta}^*) - \mathbf{1}\}\mathbf{g}^\top(x_{ij};\boldsymbol{\eta}^*)}{h(x_{ij};\boldsymbol{\eta}^*)^2} I(x_{ij} > 0), \\
\frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial\mathbf{t}\partial\mathbf{t}^\top} &= -\sum_{ij} \frac{-\mathbf{g}(x_{ij};\boldsymbol{\eta}^*)\mathbf{g}^\top(x_{ij};\boldsymbol{\eta}^*)}{h(x_{ij};\boldsymbol{\eta}^*)^2} I(x_{ij} > 0),
\end{aligned}$$

and $\otimes$ denotes the Kronecker product.

4.6.1 Some useful lemmas

In the proof of Theorem 4.1, we need the expectation of $\partial^2\ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top$, and the expectation and variance of $\mathbf{S}_n$. The following lemma is used to ease the calculation burden in our later proofs.

Lemma 4.1 *Suppose that f_1 and f_2 are two arbitrary vector-valued functions. Let X denote a random variable from $G_0(x)$. Then*
(a)

$$E\left\{\sum_{ij} f_1(x_{ij})I(x_{ij}>0)\right\} = n\Delta^* E_0\{h(X;\boldsymbol{\eta}^*)f_1(X)\}.$$

(b) For $r = 0, \ldots, m$

$$\begin{aligned} & Cov\left\{\sum_{j=1}^{n_r} f_1(x_{rj})I(x_{rj}>0), \sum_{ij} f_2(x_{ij})I(x_{ij}>0)\right\} \\ = \; & n\Delta^* E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)f_1(X)f_2^\top(X)\} \\ & - n\Delta^{*2}\rho_r^{*-1}E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)f_1(X)\}E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)f_2^\top(X)\}. \end{aligned}$$

(c)

$$\begin{aligned} & Cov\left\{\sum_{ij} f_1(x_{ij})I(x_{ij}>0), \sum_{ij} f_2(x_{ij})I(x_{ij}>0)\right\} \\ = \; & n\Delta^* E_0\{h(X;\boldsymbol{\eta}^*)f_1(X)f_2^\top(X)\} \\ & - n\Delta^{*2}\sum_{i=0}^{m}\rho_i^{*-1}E_0\{h(X;\boldsymbol{\eta}^*)h_i(X;\boldsymbol{\eta}^*)f_1(X)\}E_0\{h(X;\boldsymbol{\eta}^*)h_i(X;\boldsymbol{\eta}^*)f_2^\top(X)\}. \end{aligned}$$

Proof. For Part (a), we have

$$\begin{aligned} E\left\{\sum_{ij} f_1(x_{ij})I(x_{ij}>0)\right\} &= \sum_{i=0}^{m} n_i E\{f_1(x_{i1})I(x_{i1}>0)\} \\ &= \sum_{i=0}^{m} n_i(1-\nu_i)E_0\{\omega_i(X;\boldsymbol{\theta}_i^*)f_1(X)\} \end{aligned}$$

$$
\begin{aligned}
&= n\Delta^* \sum_{i=0}^{m} E_0\{\lambda_i^* \omega_i(X;\boldsymbol{\theta}_i^*) f_1(X)\} \\
&= n\Delta^* E_0\{h(X;\boldsymbol{\eta}^*) f_1(X)\}.
\end{aligned}
$$

This finishes Part (a).

For Part (b), we look at

$$
\begin{aligned}
& \mathrm{Cov}\left\{\sum_{j=1}^{n_r} f_1(x_{rj}) I(x_{rj} > 0), \sum_{ij} f_2(x_{ij}) I(x_{ij} > 0)\right\} \\
= \;& \mathrm{Cov}\left\{\sum_{j=1}^{n_r} f_1(x_{rj}) I(x_{rj} > 0), \sum_{j=1}^{n_r} f_2(x_{rj}) I(x_{rj} > 0)\right\} \\
= \;& n_r \mathrm{Cov}\left\{f_1(x_{r1}) I(x_{r1} > 0), f_2(x_{r1}) I(x_{r1} > 0)\right\} \\
= \;& n_r E\{f_1(x_{r1}) f_2^\top(x_{r1}) I(x_{r1} > 0)\} - n_r E\{f_1(x_{r1}) I(x_{r1} > 0)\} E\{f_2^\top(x_{r1}) I(x_{r1} > 0)\} \\
= \;& n\Delta^* E_0\{h(X;\boldsymbol{\eta}^*) h_r(X;\boldsymbol{\eta}^*) f_1(X) f_2^\top(X)\} \\
& - n\Delta^{*2} \rho_r^{*-1} E_0\{h(X;\boldsymbol{\eta}^*) h_r(X;\boldsymbol{\eta}^*) f_1(X)\} E_0\{h(X;\boldsymbol{\eta}^*) h_r(X;\boldsymbol{\eta}^*) f_2^\top(X)\}.
\end{aligned}
$$

This finishes Part (b).

For Part (c), we compare it to the result in Part (b) above. It is easy to find that

$$
\begin{aligned}
& \mathrm{Cov}\left\{\sum_{ij} f_1(x_{ij}) I(x_{ij} > 0), \sum_{ij} f_2(x_{ij}) I(x_{ij} > 0)\right\} \\
= \;& \sum_{r=0}^{m} \mathrm{Cov}\left\{\sum_{j=1}^{n_r} f_1(x_{rj}) I(x_{rj} > 0), \sum_{ij} f_2(x_{ij}) I(x_{ij} > 0)\right\} \\
= \;& n\Delta^* E_0\{h(X;\boldsymbol{\eta}^*) f_1(X) f_2^\top(X)\} \\
& - n\Delta^{*2} \sum_{r=0}^{m} \rho_r^{*-1} E_0\{h(X;\boldsymbol{\eta}^*) h_r(X;\boldsymbol{\eta}^*) f_1(X)\} E_0\{h(X;\boldsymbol{\eta}^*) h_r(X;\boldsymbol{\eta}^*) f_2^\top(X)\},
\end{aligned}
$$

where in the last line, we used a fact that $\sum_{i=0}^{m} h_i(x;\boldsymbol{\eta}^*) = 1$. This finishes Part (c). Hence, the proofs of Lemma 4.1 are completed. □

With the help of Lemma 4.1, we calculate the expectation of $\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top$. For notational simplicity, let $\mathbf{h} = \{h_1(X;\boldsymbol{\eta}^*), \ldots, h_m(X;\boldsymbol{\eta}^*)\}^\top$, $\boldsymbol{\omega} = \{\omega_1(X;\boldsymbol{\theta}_0^*), \ldots, \omega_m(X;\boldsymbol{\theta}_m^*)\}^\top$ and $\mathbf{g} = \{g_1(X;\boldsymbol{\eta}^*), \ldots, g_p(X;\boldsymbol{\eta}^*)\}^\top$.

Lemma 4.2 *With the form of $\partial^2\ell(\boldsymbol{\eta}^*,\boldsymbol{\lambda}^*,\mathbf{0})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top$ given above and under H_0 given in (4.3), we have*

$$-\frac{1}{n}E\left\{\frac{\partial^2\ell(\boldsymbol{\eta}^*,\boldsymbol{\lambda}^*,\mathbf{0})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top}\right\}=\mathbf{A}=\begin{pmatrix} A_{11} & \mathbf{0} & \mathbf{0} & A_{14}\\ \mathbf{0} & A_{22} & A_{23} & A_{24}\\ \mathbf{0} & A_{32} & -A_{33} & -A_{34}\\ A_{41} & A_{42} & -A_{43} & -A_{44}\end{pmatrix}$$

where

$$\begin{aligned}
A_{11} &= diag\left\{\frac{\rho_0^*}{\nu_0^*(1-\nu_0^*)},\ldots,\frac{\rho_m^*}{\nu_m^*(1-\nu_m^*)}\right\},\\
A_{14}^\top &= A_{41}=-\Delta^*\mathbf{C}\,diag(\boldsymbol{\mu})\,diag\left\{(\mathbf{1}-\boldsymbol{\nu}^*)^{-1}\right\},\\
A_{22} &= \Delta^*E_0\left[h(X;\boldsymbol{\eta}^*)\left\{diag(\mathbf{h})-(\mathbf{h}\mathbf{h}^\top)\right\}\otimes\left\{\mathbf{Q}(X)\mathbf{Q}^\top(X)\right\}\right],\\
A_{23} &= A_{32}^\top=\Delta^*E_0\left\{[diag(\boldsymbol{\omega})-\{\mathbf{h}(\boldsymbol{\omega}-\mathbf{1})^\top\}]\otimes\mathbf{Q}(X)\right\},\\
A_{24} &= A_{42}^\top=\Delta^*E_0\Big[\left\{diag(\boldsymbol{\omega})(\mathbf{0}_{m\times1},\mathbf{I}_{m\times m})\,diag(\mathbf{1}-\boldsymbol{\nu}^*)\mathbf{C}^\top X-(\mathbf{h}\mathbf{g}^\top)\right\}\otimes\mathbf{Q}(X)\Big],\\
A_{33} &= \Delta^*E_0\left\{\frac{(\boldsymbol{\omega}-\mathbf{1})(\boldsymbol{\omega}-\mathbf{1})^\top}{h(X;\boldsymbol{\eta}^*)}\right\},\\
A_{34} &= A_{43}^\top=\Delta^*E_0\left\{\frac{(\boldsymbol{\omega}-\mathbf{1})\mathbf{g}^\top}{h(X;\boldsymbol{\eta}^*)}\right\},\\
A_{44} &= \Delta^*E_0\left\{\frac{\mathbf{g}\mathbf{g}^\top}{h(X;\boldsymbol{\eta}^*)}\right\}.
\end{aligned}$$

Proof. The main idea for showing the results in this lemma is to apply the results in Lemma 4.1 to each block of $\partial^2\ell(\boldsymbol{\eta}^*,\boldsymbol{\lambda}^*,\mathbf{0})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top$. Because the proof for each block is either trivial or is very similar, here we only show how to find A_{24} as an illustration.

Recall that

$$\frac{\partial^2\ell(\boldsymbol{\eta}^*,\boldsymbol{\lambda}^*,\mathbf{0})}{\partial\boldsymbol{\theta}\partial\mathbf{t}^\top}=-\sum_{ij}\frac{\partial\mathbf{g}^\top(x_{ij},\boldsymbol{\eta}^*)/\partial\boldsymbol{\theta}-\{\mathbf{h}(x_{ij};\boldsymbol{\theta}^*)\mathbf{g}^\top(x_{ij},\boldsymbol{\eta}^*)\}\otimes\mathbf{Q}(x_{ij})}{h(x_{ij};\boldsymbol{\eta}^*)}I(x_{ij}>0),$$

Hence, by treating the term in summation as $f_1(x_{ij})I(x_{ij}>0)$ in part (a) of Lemma 4.1, we have

$$A_{24} = -\frac{1}{n}E\left\{\frac{\partial^2\ell(\boldsymbol{\eta}^*,\boldsymbol{\lambda}^*,\mathbf{0})}{\partial\boldsymbol{\theta}\partial\mathbf{t}^\top}\right\}$$

$$
\begin{aligned}
&= \Delta^* E_0\left[\left\{\partial \mathbf{g}^\top/\partial \boldsymbol{\theta} - (\mathbf{h}\mathbf{g}^\top)\otimes \mathbf{Q}(X)\right\}\right] \\
&= \Delta^* E_0\Big[\left\{\text{diag}(\boldsymbol{\omega})(\mathbf{0}_{m\times 1}, \mathbf{I}_{m\times m})\text{diag}(\mathbf{1}-\boldsymbol{\nu}^*)\mathbf{C}^\top X - (\mathbf{h}\mathbf{g}^\top)\right\}\otimes \mathbf{Q}(X)\Big].
\end{aligned}
$$

This finishes the proof. □

With the help of Lemma 4.1, we now study the properties of $\mathbf{S}_n$ defined in (4.20).

Lemma 4.3 *With the form of* $\mathbf{S}_n = \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \boldsymbol{\psi}}$ *defined in (4.20) and under* H_0 *given in (4.3), we have*

$$
\frac{1}{\sqrt{n}}\mathbf{S}_n \to N(\mathbf{0}, \mathbf{B})
$$

in distribution as $n \to \infty$*, where*

$$
\mathbf{B} = \begin{pmatrix} B_{11} & B_{12} & B_{13} & B_{14} \\ B_{21} & B_{22} & B_{23} & B_{24} \\ B_{31} & B_{32} & B_{33} & B_{34} \\ B_{41} & B_{42} & B_{43} & B_{44} \end{pmatrix},
$$

and

$$
\begin{aligned}
B_{11} &= A_{11}, \\
B_{12}^\top &= B_{21} = B_2 W, \\
B_{13}^\top &= B_{31} = -B_3 W, \\
B_{14}^\top &= B_{41} = -B_4 W, \\
B_{22} &= A_{22} - B_2 S B_2^\top, \\
B_{23} &= B_{32}^\top = B_2 S B_3^\top, \\
B_{24} &= B_{42}^\top = B_2 S B_4^\top, \\
B_{33} &= A_{33} - B_3 S B_3^\top, \\
B_{34} &= B_{43}^\top = A_{34} - B_3 S B_4^\top, \\
B_{44} &= A_{44} - B_4 S B_4^\top,
\end{aligned}
$$

and B_2*,* B_3*,* B_4*,* W*, and* S *are defined as*

$$
B_2 = \Delta^* E_0\left[h(X; \boldsymbol{\eta}^*)\left\{diag(\mathbf{h}) - (\mathbf{h}\mathbf{h}^\top)\right\}\otimes \mathbf{Q}(X)\right],
$$

$$
\begin{aligned}
B_3 &= \Delta^* E_0\{(\boldsymbol{\omega}-\mathbf{1})\mathbf{h}^\top\},\\
B_4 &= \Delta^* E_0\{\mathbf{g}\mathbf{h}^\top\},\\
W &= (\mathbf{1}_{m\times 1}, -\mathbf{I}_{m\times m})\, diag\{(\mathbf{1}-\boldsymbol{\nu}^*)^{-1}\}\\
S &= \{\rho_0^{*-1}\mathbf{1}_m\mathbf{1}_m^\top + diag(\rho_1^{*-1},\ldots,\rho_m^{*-1})\}.
\end{aligned}
$$

Proof. We first show that $E(\mathbf{S}_n) = \mathbf{0}$. The main idea is to apply the result of Part (a) of Lemma 4.1. We only verify $E(\mathbf{S}_{n2}) = \mathbf{0}$ since the other results are either trivial or very similar. For the rth segment of $\mathbf{S}_{n2}$ given in (4.20), we have

$$
\begin{aligned}
E(\mathbf{S}_{n2,r}) &= E\left\{\sum_{j=1}^{n_r}\mathbf{Q}(x_{rj})I(x_{ij}>0) - \sum_{ij}\mathbf{Q}(x_{ij})h_r(x_{ij};\boldsymbol{\eta}^*)I(x_{ij}>0)\right\}\\
&= E\left\{\sum_{j=1}^{n_r}\mathbf{Q}(x_{rj})I(x_{ij}>0)\right\} - E\left\{\sum_{ij}\mathbf{Q}(x_{ij})h_r(x_{ij};\boldsymbol{\eta}^*)I(x_{ij}>0)\right]\\
&= n\Delta^* E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\} - n\Delta^* E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}\\
&= \mathbf{0},
\end{aligned}
$$

where we have used Part (a) of Lemma 4.1, and a similar fact that

$$
E\left\{\sum_{j=1}^{n_r} f_1(x_{rj})I(x_{rj}>0)\right\} = n\Delta^* E_0\{\lambda_r^*\omega_r(X;\boldsymbol{\theta}_r)f_1(X)\} = n\Delta^* E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)f_1(X)\}.
$$

Next, we verify that $\mathrm{Var}(\mathbf{S}_n) = \mathbf{B}$. The key idea is to apply Parts (b)–(c) of Lemma 4.1 to the (r,s) entries or blocks of $\mathrm{Var}(\mathbf{S}_n)$. Again, in the following we will only show how to find B_{14} and B_{23} as an illustration, because the proof for other blocks are either trivial or are very similar.

First, we find B_{14}. For $r = 0,\ldots,m$ and $s = 1,\ldots,p$,

$$
\begin{aligned}
\frac{1}{n}\mathrm{Cov}(\mathbf{S}_{n1,r},\mathbf{S}_{n4,s}) &= \frac{1}{n}\mathrm{Cov}\left\{\frac{-n_{r1}}{\nu_r^*(1-\nu_r^*)}, -\sum_{ij}\frac{g_s(x_{ij};\boldsymbol{\eta}^*)}{h(x_{ij};\boldsymbol{\eta}^*)}I(x_{ij}>0)\right\}\\
&= \frac{1}{n\nu_r^*(1-\nu_r^*)}\mathrm{Cov}\left\{\sum_{j=1}^{n_r}I(x_{rj}>0), \sum_{ij}\frac{g_s(x_{ij};\boldsymbol{\eta}^*)}{h(x_{ij};\boldsymbol{\eta}^*)}I(x_{ij}>0)\right\}
\end{aligned}
$$

$$= \frac{\Delta^* \nu_r^*}{\nu_r^*(1-\nu_r^*)} E_0 \left\{ h_r(X;\boldsymbol{\eta}^*) g_s(X;\boldsymbol{\eta}^*) \right\},$$

where we have used a special case of part (b) when $f_1(x) = 1$ in Lemma 4.1. Hence, for $r = 1, \ldots, m$ and $s = 1, \ldots, p$, we have shown that

$$\frac{1}{n}\mathrm{Cov}(\mathbf{S}_{n1,r}, \mathbf{S}_{n4,s}) = \frac{1}{1-\nu_r^*} B_{4,sr}.$$

Further, we can simplify $\frac{1}{n}\mathrm{Cov}(\mathbf{S}_{n1,0}, \mathbf{S}_{n4,s})$ by recalling that $h_0(X;\boldsymbol{\eta}^*) = 1 - \sum_{i=1}^m h_i(x;\boldsymbol{\eta}^*)$. Then

$$\begin{aligned} \frac{1}{n}\mathrm{Cov}(\mathbf{S}_{n1,0}, \mathbf{S}_{n4,s}) &= \frac{\Delta^*}{1-\nu_0^*} E_0 \left\{ h_0(X;\boldsymbol{\eta}^*) g_s(X;\boldsymbol{\eta}^*) \right\}] \\ &= \frac{-1}{1-\nu_0^*} \sum_{r=1}^m B_{4,sr}. \end{aligned}$$

In summary,

$$B_{14} = \frac{1}{n}\mathrm{Cov}(\mathbf{S}_{n1}, \mathbf{S}_{n4}) = \left[-(1-\nu_0^*)^{-1} \mathbf{1}_m^\top B_4^\top, \mathrm{diag}\left\{ (1-\nu_1^*)^{-1}, \ldots, (1-\nu_m^*)^{-1} \right\} B_4^\top \right] = -W^\top B_4^\top.$$

This finishes the proof of B_{14}.

Next, we move on to find B_{23}. For $r = 1, \ldots, m$ and $s = 1, \ldots, m$,

$$\begin{aligned} &\frac{1}{n}\mathrm{Cov}(\mathbf{S}_{n2,r}, \mathbf{S}_{n3,s}) \\ = \; &\frac{1}{n}\mathrm{Cov}\left[\sum_{j=1}^{n_r} \mathbf{Q}(x_{rj}) I(x_{ij} > 0) - \sum_{ij} \mathbf{Q}(x_{ij}) h_r(x_{ij};\boldsymbol{\eta}^*) I(x_{ij} > 0), - \sum_{ij} \frac{\omega_s(x_{ij};\boldsymbol{\theta}_s^*) - 1}{h(x_{ij};\boldsymbol{\eta}^*)} I(x_{ij} > 0) \right] \\ = \; &\frac{1}{n}\mathrm{Cov}\left[\sum_{j=1}^{n_r} \mathbf{Q}(x_{rj}) I(x_{ij} > 0), - \sum_{ij} \frac{\omega_s(x_{ij};\boldsymbol{\theta}_s^*) - 1}{h(x_{ij};\boldsymbol{\eta}^*)} I(x_{ij} > 0) \right] \\ &- \frac{1}{n}\mathrm{Cov}\left[\sum_{ij} h_r(x_{ij};\boldsymbol{\eta}^*) \mathbf{Q}(x_{ij}) I(x_{ij} > 0), - \sum_{ij} \frac{\omega_s(x_{ij};\boldsymbol{\theta}_s^*) - 1}{h(x_{ij};\boldsymbol{\eta}^*)} I(x_{ij} > 0) \right]. \end{aligned} \tag{4.21}$$

Applying the part (b) of Lemma 4.1 to the first term in (4.21), it becomes

$$\frac{1}{n}\mathrm{Cov}\left[\sum_{j=1}^{n_r} \mathbf{Q}(x_{rj}) I(x_{ij} > 0), - \sum_{ij} \frac{\omega_s(x_{ij};\boldsymbol{\theta}_s^*) - 1}{h(x_{ij};\boldsymbol{\eta}^*)} I(x_{ij} > 0) \right]$$

$$
\begin{aligned}
&= \Delta^* E_0\left[-h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right] \\
&\quad -\Delta^{*2}\rho_r^{*-1}E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}E_0\left[-h_r(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right]. \qquad (4.22)
\end{aligned}
$$

Then applying the part (c) of Lemma 4.1 to the second term in (4.21), it becomes

$$
\begin{aligned}
&\frac{1}{n}\mathrm{Cov}\left[\sum_{ij}\mathbf{Q}(x_{ij})h_r(x_{ij};\boldsymbol{\eta}^*)I(x_{ij}>0), -\sum_{ij}\frac{\omega_s(x_{ij};\boldsymbol{\theta}_s^*)-1}{h(x_{ij};\boldsymbol{\eta}^*)}I(x_{ij}>0)\right] \\
= &-\Delta^* E_0\left[-h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right] \\
&+\Delta^{*2}\sum_{k=0}^{m}\rho_k^{*-1}E_0\{h(X;\boldsymbol{\eta}^*)h_k(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}E_0\left[-h_k(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right]. \qquad (4.23)
\end{aligned}
$$

Plugging in (4.22) and (4.23) to (4.21), we find

$$
\begin{aligned}
&\frac{1}{n}\mathrm{Cov}(\mathbf{S}_{n2,r},\mathbf{S}_{n3,s}) \\
= &\ \Delta^{*2}\rho_r^{*-1}E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}E_0\left[h_r(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right] \\
&-\Delta^{*2}\sum_{k=0}^{m}\rho_k^{*-1}E_0\{h(X;\boldsymbol{\eta}^*)h_k(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}E_0\left[h_k(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right]. \quad (4.24)
\end{aligned}
$$

To further simplify the result, recall that $h_0(X;\boldsymbol{\eta}^*) = 1-\sum_{i=1}^m h_i(x;\boldsymbol{\eta}^*)$. Then

$$
\begin{aligned}
&E_0\{h_0(X;\boldsymbol{\eta}^*)h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\} \\
= &\ E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\} - \sum_{i=1}^m E_0\{h_i(X;\boldsymbol{\eta}^*)h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}, \quad (4.25)
\end{aligned}
$$

and

$$
E_0\left[h_0(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right] = -\sum_{j=1}^m E_0\left[h_j(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right]. \qquad (4.26)
$$

Using (4.25) and (4.26) to replace the term in (4.24) for $k=0$, we find

$$
\begin{aligned}
&\Delta^{*2}\rho_0^{*-1}E_0\{h(X;\boldsymbol{\eta}^*)h_0(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}E_0\left[h_0(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right] \\
= &\ \Delta^{*2}\rho_0^{*-1}\left[E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\} - \sum_{i=1}^m E_0\{h_i(X;\boldsymbol{\eta}^*)h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}\right] \\
&\times\left\{-\sum_{j=1}^m E_0\left[h_j(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
= \quad & -\Delta^{*2}\rho_0^{*-1}E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}\left\{\sum_{j=1}^m E_0\left[h_j(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right]\right\} \\
& +\Delta^{*2}\rho_0^{*-1}\left[\sum_{i=1}^m E_0\{h_i(X;\boldsymbol{\eta}^*)h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}\right]\left\{\sum_{j=1}^m E_0\left[h_j(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right]\right\},
\end{aligned}
$$

which together with (4.24) leads to

$$
\begin{aligned}
& \frac{1}{n}\mathrm{Cov}(\mathbf{S}_{n2,r},\mathbf{S}_{n3,s}) \\
= \quad & \Delta^{*2}\rho_0^{*-1}E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}\left\{\sum_{j=1}^m E_0\left[h_j(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right]\right\} \\
& +\Delta^{*2}\rho_r^{*-1}E_0\{h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}E_0\left[h_r(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right] \\
& -\Delta^{*2}\rho_0^{*-1}\left[\sum_{i=1}^m E_0\{h_i(X;\boldsymbol{\eta}^*)h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}\right]\left\{\sum_{j=1}^m E_0\left[h_j(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right]\right\} \\
& -\Delta^{*2}\sum_{k=1}^m \rho_k^{*-1}E_0\{h(X;\boldsymbol{\eta}^*)h_k(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\}E_0\left[h_k(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right]. \qquad (4.27)
\end{aligned}
$$

It is very useful to introduce a group indicator I_{ij} such that $I_{ij}=1$ if $i=j$ and $I_{ij}=0$ otherwise. Hence, we simplify (4.27) to

$$
\begin{aligned}
& \frac{1}{n}\mathrm{Cov}(\mathbf{S}_{n2,r},\mathbf{S}_{n3,s}) \\
= \quad & \sum_{i=1}^m\sum_{j=1}^m(\rho_0^{*-1}+\rho_i^{*-1}I_{ij})E_0\left[h_i(X;\boldsymbol{\eta}^*)h(X;\boldsymbol{\eta}^*)I_{ir}\mathbf{Q}(X)\right]E_0\left[h_j(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right] \\
& -\sum_{i=1}^m\sum_{j=1}^m(\rho_0^{*-1}+\rho_i^{*-1}I_{ij})E_0\left[h_i(X;\boldsymbol{\eta}^*)h(X;\boldsymbol{\eta}^*)h_r(X;\boldsymbol{\eta}^*)\mathbf{Q}(X)\right]E_0\left[h_j(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right] \\
= & \sum_{i=1}^m\sum_{j=1}^m(\rho_0^{*-1}+\rho_i^{*-1}I_{ij})E_0\left[h_i(X;\boldsymbol{\eta}^*)h(X;\boldsymbol{\eta}^*)\{I_{ir}-h_r(X;\boldsymbol{\eta}^*)\}\mathbf{Q}(X)\right]E_0\left[h_j(X;\boldsymbol{\eta}^*)\{\omega_s(X;\boldsymbol{\theta}_s^*)-1\}\right] \\
= \quad & \sum_{i=1}^m\sum_{j=1}^m(\rho_0^{*-1}+\rho_i^{*-1}I_{ij})B_{2,ri}B_{3,sj}.
\end{aligned}
$$

In summary, we have

$$
B_{23}=B_2SB_3^{\top}.
$$

This finishes the proof of B_{23}.

Finally, recall that $\mathbf{S}_n$ in (4.20) is a sum of independent random vectors. Therefore, by the classical central limit theorem, we have

$$\frac{1}{\sqrt{n}}\mathbf{S}_n \to N(\mathbf{0}, \mathbf{B}),$$

in distribution, which finishes the proof of Lemma 4.3. □

4.6.2 Proof of Theorem 4.1

With above preparation, we now move back to derive the limiting distribution of R_n under the null hypothesis given in (4.3) where $H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{d}$. Using the similar arguments to those in the proofs of Lemma 1 and Theorem 1 of Qin and Lawless (1994), and Theorem 2.1 of Chen and Liu (2013), we have

$$\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}^* = O_p(n^{-1/2}), \quad \hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = O_p(n^{-1/2}), \quad \hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^* = O_p(n^{-1/2}).$$

Recall that in (4.19)

$$\begin{aligned} R_n &= 2\left\{\ell(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\lambda}}, \mathbf{0}) - \ell(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}})\right\} \\ &= 2\left\{\ell(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\lambda}}, \mathbf{0}) - \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})\right\} - 2\left\{\ell(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}}) - \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})\right\}. \end{aligned}$$

We proceed in three steps. In Step 1, we find the asymptotic approximation of $\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}^*$, which is used to find an approximation of $\ell(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}}) - \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})$. In Step 2, we find the asymptotic approximation of $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*$ and $\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*$, which is then used to find one for $\ell(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\lambda}}, \mathbf{0}) - \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})$. Combining Steps 1 and 2 together, we obtain a quadratic approximation of R_n. In Step 3, we derive the limiting distribution of the quadratic approximation.

We start with Step 1. Recall that

$$\frac{\partial \ell(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}})}{\partial \boldsymbol{\psi}} = \mathbf{0}.$$

Applying the first-order Taylor expansion to the above equation, we have

$$\mathbf{0} = \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \boldsymbol{\psi}} + \frac{\partial^2 \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^\top}(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) + O_p(1).$$

Recall that

$$\mathbf{S}_n = \frac{\partial \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})}{\partial \boldsymbol{\psi}}.$$

Using Lemma 4.2, we further have

$$\mathbf{0} = \mathbf{S}_n - n\mathbf{A}(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) + O_p(1),$$

which implies that

$$\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}^* = (n\mathbf{A})^{-1}\mathbf{S}_n + o_p(n^{-1/2}).$$

Applying the second-order Taylor expansion to $\ell(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}}) - \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})$ and using Lemma 4.2, we get

$$\ell(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}}) - \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0}) = \mathbf{S}_n^\top(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) - \frac{1}{2}(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}^*)^\top(n\mathbf{A})(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) + o_p(1).$$

With the approximation of $\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}^*$, we obtain an approximation of $\ell(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}}) - \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})$ as

$$\ell(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}}) - \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0}) = \frac{1}{2}\mathbf{S}_n^\top(n\mathbf{A})^{-1}\mathbf{S}_n + o_p(1). \tag{4.28}$$

In Step 2, we partition the matrix $\mathbf{A}$ into blocks and denote them by

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\Lambda}_0 & \boldsymbol{\Lambda}_1 \\ \boldsymbol{\Lambda}_1^\top & \boldsymbol{\Lambda}_2 \end{pmatrix} = \left(\begin{array}{ccc:c} A_{11} & \mathbf{0} & \mathbf{0} & A_{14} \\ \mathbf{0} & A_{22} & A_{23} & A_{24} \\ \mathbf{0} & A_{32} & -A_{33} & -A_{34} \\ \hdashline A_{41} & A_{42} & -A_{43} & -A_{44} \end{array}\right).$$

Using the similar steps as in Step 1, we obtain the approximation of $\ell(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}}) - \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0})$ as

$$\ell(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\lambda}}, \mathbf{0}) - \ell(\boldsymbol{\eta}^*, \boldsymbol{\lambda}^*, \mathbf{0}) = \frac{1}{2}\mathbf{S}_{n,-4}^\top(n\boldsymbol{\Lambda}_0)^{-1}\mathbf{S}_{n,-4} + o_p(1), \tag{4.29}$$

where $\mathbf{S}_{n,-4} = (\mathbf{S}_{n1}^\top, \mathbf{S}_{n2}^\top, \mathbf{S}_{n3}^\top)^\top$.

Combining (4.28) and (4.29), we obtain the approximation of R_n as

$$R_n = \mathbf{S}_{n,-4}^\top (n\boldsymbol{\Lambda}_0)^{-1}\mathbf{S}_{n,-4} - \mathbf{S}_n^\top (n\mathbf{A})^{-1}\mathbf{S}_n + o_p(1).$$

By the inverse of block matrix, we have

$$\mathbf{A}^{-1} = \begin{pmatrix} \boldsymbol{\Lambda}_0^{-1} + \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_1\mathbf{A}_{2.1}^{-1}\boldsymbol{\Lambda}_1^\top\boldsymbol{\Lambda}_0^{-1} & -\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_1\mathbf{A}_{2.1}^{-1} \\ -\mathbf{A}_{2.1}^{-1}\boldsymbol{\Lambda}_1^\top\boldsymbol{\Lambda}_0^{-1} & \mathbf{A}_{2.1}^{-1} \end{pmatrix},$$

where $\mathbf{A}_{2.1} = \boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1^\top\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_1$. After some algebra, we find that

$$R_n = -\frac{1}{n}\left(\mathbf{S}_{n4} - \boldsymbol{\Lambda}_1^\top\boldsymbol{\Lambda}_0^{-1}\mathbf{S}_{n,-4}\right)^\top (\boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1^\top\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_1)^{-1}\left(\mathbf{S}_{n4} - \boldsymbol{\Lambda}_1^\top\boldsymbol{\Lambda}_0^{-1}\mathbf{S}_{n,-4}\right) + o_p(1). \quad (4.30)$$

From Lemma 4.3, we have

$$n^{-1/2}\mathbf{S}_n \to N(\mathbf{0}, \mathbf{B})$$

in distribution. As a result,

$$n^{-1/2}\left(\mathbf{S}_{n4} - \boldsymbol{\Lambda}_1^\top\boldsymbol{\Lambda}_0^{-1}\mathbf{S}_{n,-4}\right) = n^{-1/2}(-\boldsymbol{\Lambda}_1^\top\boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})\mathbf{S}_n \to N(\mathbf{0}, \mathbf{U})$$

in distribution, where $\mathbf{U} = (-\boldsymbol{\Lambda}_1^\top\boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})\mathbf{B}(-\boldsymbol{\Lambda}_1^\top\boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top$. Hence, to show the chi-squared limiting distribution of R_n, the remaining task is to argue that

$$\mathbf{U} = \boldsymbol{\Lambda}_1^\top\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_1 - \boldsymbol{\Lambda}_2. \quad (4.31)$$

Next, we present a lemma, which study the relationship between $\boldsymbol{\Lambda}_0^{-1}$ and $\mathbf{B}$.

Lemma 4.4 *Denote* $\boldsymbol{\Lambda}_0^{-1}$ *by*

$$\boldsymbol{\Lambda}_0^{-1} = \begin{pmatrix} A_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_{22} & A_{23} \\ \mathbf{0} & A_{32} & -A_{33} \end{pmatrix}^{-1} = \begin{pmatrix} A^{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A^{22} & A^{23} \\ \mathbf{0} & A^{32} & -A^{33} \end{pmatrix}.$$

Then we have

(a)

$$A^{11} = A_{11}^{-1},$$

$$A^{22} = A_{22}^{-1} - A^{23}A_{32}A_{22}^{-1} = A_{22}^{-1} - A^{23}(A^{33})^{-1}A^{32}, \quad (4.32)$$
$$A^{32} = (A^{23})^\top = \frac{1}{\Delta^*}\{\mathbf{I}_{m\times m} \otimes \mathbf{e}^\top\}, \quad (4.33)$$
$$A^{33} = \frac{1}{\Delta^*}\{diag(\boldsymbol{\lambda}^*) - \boldsymbol{\lambda}^*\boldsymbol{\lambda}^{*\top}\}, \quad (4.34)$$

where $\mathbf{e} = (1, 0, \ldots, 0)^\top$ *is a vector with first element being* 1 *and others zero, and the length of* $\mathbf{e}$ *is the same as that of* $\mathbf{Q}(X)$*; and*
(b)

$$B_2 = \Delta^* A_{22}A^{23} = \Delta^* A_{23}A^{33}, \quad (4.35)$$
$$B_3 = \Delta^*\mathbf{I}_{m\times m} - \Delta^* A_{32}A^{23} = \Delta^* A_{33}A^{33}, \quad (4.36)$$
$$B_4 = \Delta^* V - \Delta^* A_{42}A^{23} = \Delta^* A_{43}A^{33}, \quad (4.37)$$

where $V = \mathbf{C}\,diag(\mathbf{1} - \boldsymbol{\nu}^*)(\mathbf{0}_{m\times 1}, \mathbf{I}_{m\times m})^\top E_0\left\{diag(\boldsymbol{\omega})X\right\}$ *appears in (4.37), for simplicity.*

Proof. Recall that from Lemma 4.2, we have found

$$A_{22} = \Delta^* E_0\left[h(X;\boldsymbol{\eta}^*)\left\{\text{diag}(\mathbf{h}) - (\mathbf{h}\mathbf{h}^\top)\right\} \otimes \{\mathbf{Q}(X)\mathbf{Q}^\top(X)\}\right],$$
$$A_{32} = A_{23}^\top = \Delta^* E_0\left\{[\text{diag}(\boldsymbol{\omega}) - \{(\boldsymbol{\omega} - \mathbf{1})\mathbf{h}^\top\}] \otimes \mathbf{Q}^\top(X)\right\},$$
$$A_{33} = \Delta^* E_0\left\{\frac{(\boldsymbol{\omega} - \mathbf{1})(\boldsymbol{\omega} - \mathbf{1})^\top}{h(X;\boldsymbol{\eta}^*)}\right\}.$$

For part (a), $A^{11} = A_{11}^{-1}$ is obvious. For the rest, since the inverse matrix of $\boldsymbol{\Lambda}_0$ is unique, we only need to verify that

$$A_{22}A^{22} + A_{23}A^{32} = \mathbf{I}_{md\times md}, \quad (4.38)$$
$$A_{22}A^{23} - A_{23}A^{33} = \mathbf{0}_{md\times m}, \quad (4.39)$$
$$A_{32}A^{22} - A_{33}A^{32} = \mathbf{0}_{m\times md}, \quad (4.40)$$
$$A_{32}A^{23} + A_{33}A^{33} = \mathbf{I}_{m\times m}, \quad (4.41)$$

with the forms of $A^{22} = A_{22}^{-1} - A^{23}A_{32}A_{22}^{-1}$, A^{23}, A^{32}, and A^{33} given in (4.33)–(4.34).

We first establish some useful relationships. By the property of Kronecker product and the fact that $\mathbf{Q}^\top(X)\mathbf{e} = 1$, we have

$$A_{22}A^{23} = \Delta^* E_0\left[h(X;\boldsymbol{\eta}^*)\left\{\text{diag}(\mathbf{h}) - (\mathbf{h}\mathbf{h}^\top)\right\} \otimes \{\mathbf{Q}(X)\mathbf{Q}^\top(X)\}\right]\frac{1}{\Delta^*}\{\mathbf{I}_{m\times m} \otimes \mathbf{e}\}$$

$$
\begin{aligned}
&= E_0\left[h(X;\boldsymbol{\eta}^*)\left\{\text{diag}(\mathbf{h}) - (\mathbf{h}\mathbf{h}^\top)\right\} \otimes \{\mathbf{Q}(X)\}\right] \\
&= B_2/\Delta^*. \qquad (4.42)
\end{aligned}
$$

Similarly, we have

$$
A_{32}A^{23} = \mathbf{I}_{m\times m} - B_3/\Delta^*. \qquad (4.43)
$$

Next, by the properties of the Kronecker product and the definitions of $\boldsymbol{\omega}$, $\mathbf{h}$, and $\boldsymbol{\lambda}^*$, we have

$$
\begin{aligned}
A_{23}A^{33} &= \Delta^* E_0\left\{[\text{diag}(\boldsymbol{\omega}) - \{\mathbf{h}(\boldsymbol{\omega}-\mathbf{1})^\top\}] \otimes \mathbf{Q}(X)\right\} \frac{1}{\Delta^*}\{\text{diag}(\boldsymbol{\lambda}^*) - \boldsymbol{\lambda}^*\boldsymbol{\lambda}^{*\top}\} \\
&= E_0\Big\{[\text{diag}(\boldsymbol{\omega})\text{diag}(\boldsymbol{\lambda}^*) - \text{diag}(\boldsymbol{\omega})\boldsymbol{\lambda}^*\boldsymbol{\lambda}^{*\top} \\
&\qquad -\mathbf{h}(\boldsymbol{\omega}-\mathbf{1})^\top\text{diag}(\boldsymbol{\lambda}^*) + \mathbf{h}(\boldsymbol{\omega}-\mathbf{1})^\top\boldsymbol{\lambda}^*\boldsymbol{\lambda}^{*\top}] \otimes \mathbf{Q}(X)\Big\} \\
&= E_0\left\{[h(X;\boldsymbol{\eta}^*)\text{diag}(\mathbf{h}) - h(X;\boldsymbol{\eta}^*)\mathbf{h}\boldsymbol{\lambda}^{*\top} - h(X;\boldsymbol{\eta}^*)\mathbf{h}\mathbf{h}^\top + \mathbf{h}\boldsymbol{\lambda}^{*\top}\right. \\
&\qquad \left. + \{h(X;\boldsymbol{\eta}^*) - \lambda_0^*\}\mathbf{h}\boldsymbol{\lambda}^{*\top} - (1-\lambda_0^*)\mathbf{h}\boldsymbol{\lambda}^{*\top}] \otimes \mathbf{Q}(X)\right\} \\
&= E_0\left[h(X;\boldsymbol{\eta}^*)\left\{\text{diag}(\mathbf{h}) - (\mathbf{h}\mathbf{h}^\top)\right\} \otimes \{\mathbf{Q}(X)\}\right] \\
&= B_2/\Delta^*. \qquad (4.44)
\end{aligned}
$$

Also, we can similarly show

$$
A_{33}A^{33} = B_3/\Delta^*. \qquad (4.45)
$$

Then (4.42) and (4.44) together lead to

$$
A_{22}A^{23} = A_{23}A^{33}, \qquad (4.46)
$$

which implies that (4.39) is correct and the two forms of A^{22} in (4.32) are the same. Combining (4.43) and (4.45) gives

$$
\mathbf{I}_{m\times m} - \Delta^* A_{32}A^{23} = A_{33}A^{33}, \qquad (4.47)
$$

which implies that (4.41) is correct.

Next, (4.32) together with (4.46) give

$$
A_{22}A^{22} = A_{22}(A_{22}^{-1} - A^{23}(A^{33})^{-1}A^{32})
$$

$$
\begin{aligned}
&= \mathbf{I}_{md\times md} - A_{22}A^{23}(A^{33})^{-1}A^{32} \\
&= \mathbf{I}_{md\times md} - A_{23}A^{33}(A^{33})^{-1}A^{32} \\
&= \mathbf{I}_{md\times md} - A_{23}A^{32}.
\end{aligned}
$$

This verifies that (4.38) is correct.

Lastly, (4.32) together with (4.46) and (4.47) give

$$
\begin{aligned}
A_{32}A^{22} &= A_{32}(A_{22}^{-1} - A^{23}A_{32}A_{22}^{-1}) \\
&= A_{32}A_{22}^{-1} - A_{32}A^{23}A_{32}A_{22}^{-1} \\
&= A_{32}A_{22}^{-1} - (\mathbf{I}_{m\times m} - A_{33}A^{33})A_{32}A_{22}^{-1} \\
&= A_{33}A^{33}A_{32}A_{22}^{-1} \\
&= A_{33}A^{32}.
\end{aligned}
$$

This verifies that (4.40) is correct.

For Part (b), (4.35) follows from (4.42) and (4.44), and (4.36) follows from (4.43) and (4.45). Here we only need to further verify (4.37). Recall that from Lemma 4.2,

$$
\begin{aligned}
A_{42} &= \Delta^* E_0\Big[\mathbf{C}\mathrm{diag}(\mathbf{1}-\boldsymbol{\nu}^*)\left\{(\mathbf{0}_{m\times 1}, \mathbf{I}_{m\times m})^\top \mathrm{diag}(\boldsymbol{\omega})X - (\mathbf{g}\mathbf{h}^\top)\right\} \otimes \mathbf{Q}^\top(X)\Big], \\
A_{43} &= \Delta^* E_0\left\{\frac{\mathbf{g}(\boldsymbol{\omega}-\mathbf{1})^\top}{h(X;\boldsymbol{\eta}^*)}\right\}.
\end{aligned}
$$

Similar to (4.42), we have

$$
A_{42}A^{32} = E_0\Big\{\mathbf{C}\mathrm{diag}(\mathbf{1}-\boldsymbol{\nu}^*)(\mathbf{0}_{m\times 1}, \mathbf{I}_{m\times m})^\top \mathrm{diag}(\boldsymbol{\omega})X\Big\} - B_4/\Delta^*.
$$

It can be verified that $V = E_0\Big\{\mathbf{C}\mathrm{diag}(\mathbf{1}-\boldsymbol{\nu}^*)(\mathbf{0}_{m\times 1}, \mathbf{I}_{m\times m})^\top \mathrm{diag}(\boldsymbol{\omega})X\Big\}$. Hence

$$
A_{42}A^{32} = V - B_4/\Delta^* \tag{4.48}
$$

Also, similar to (4.42), we have

$$
A_{43}A^{33} = B_4/\Delta^*. \tag{4.49}
$$

Finally, (4.48) and (4.49) together imply (4.37). This completes the proof of Part (b), and hence that of Lemma 4.4. □

We now move back to verify (4.31) by following two steps. In Step 1, we find the form of $-\mathbf{\Lambda}_2+\mathbf{\Lambda}_1^\top\mathbf{\Lambda}_0^{-1}\mathbf{\Lambda}_1$. In Step 2, we identify the form of $\mathbf{U}=(-\mathbf{\Lambda}_1^\top\mathbf{\Lambda}_0^{-1},\mathbf{I}_{p\times p})\mathbf{B}(-\mathbf{\Lambda}_1^\top\mathbf{\Lambda}_0^{-1},\mathbf{I}_{p\times p})^\top$. Combining the two steps together, we can verify (4.31).

For Step 1, we start with

$$\mathbf{\Lambda}_0^{-1}\mathbf{\Lambda}_1 = \begin{pmatrix} A_{11}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A^{22} & A^{23} \\ \mathbf{0} & A^{32} & -A^{33} \end{pmatrix} \begin{pmatrix} A_{14} \\ A_{24} \\ -A_{34} \end{pmatrix} = \begin{pmatrix} A_{11}^{-1}A_{14} \\ A^{22}A_{24}-A^{23}A_{34} \\ A^{32}A_{24}+A^{33}A_{34} \end{pmatrix}.$$

By (4.37), we have

$$A^{32}A_{24}+A^{33}A_{34}=V^\top.$$

Using (4.32), we get

$$\begin{aligned} A^{22}A_{24}-A^{23}A_{34} &= \{A_{22}^{-1}-A^{23}(A^{33})^{-1}A^{32}\}A_{24}-A^{23}A_{34} \\ &= A_{22}^{-1}A_{24}-A^{23}(A^{33})^{-1}A^{32}A_{24}-A^{23}A_{34} \\ &= A_{22}^{-1}A_{24}-A^{23}(A^{33})^{-1}V^\top. \end{aligned}$$

Hence, we obtain

$$\mathbf{\Lambda}_0^{-1}\mathbf{\Lambda}_1 = \begin{pmatrix} A_{11}^{-1}A_{14} \\ A_{22}^{-1}A_{24}-A^{23}(A^{33})^{-1}V^\top \\ V^\top \end{pmatrix}. \tag{4.50}$$

With (4.50), $-\mathbf{\Lambda}_2+\mathbf{\Lambda}_1^\top\mathbf{\Lambda}_0^{-1}\mathbf{\Lambda}_1$ can be rewritten as

$$\begin{aligned} -\mathbf{\Lambda}_2+\mathbf{\Lambda}_1^\top\mathbf{\Lambda}_0^{-1}\mathbf{\Lambda}_1 &= A_{44}+A_{41}A_{11}^{-1}A_{14}+A_{42}\{A_{22}^{-1}A_{24}-A^{23}(A^{33})^{-1}V^\top\}-A_{43}V^\top \\ &= A_{44}+A_{41}A_{11}^{-1}A_{14}+A_{42}A_{22}^{-1}A_{24}-A_{42}A^{23}(A^{33})^{-1}V^\top-A_{43}V^\top \\ &= A_{44}+A_{41}A_{11}^{-1}A_{14}+A_{42}A_{22}^{-1}A_{24}-V(A^{33})^{-1}V^\top, \end{aligned} \tag{4.51}$$

where in the last equation we have used (4.37). This finishes the Step 1.

We now move to Step 2. With the form of $\mathbf{B}$ defined in Lemma 4.3, we can rewrite $\mathbf{B}$ as

$$\mathbf{B} = \begin{pmatrix} A_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_{33} & A_{34} \\ \mathbf{0} & \mathbf{0} & A_{43} & A_{44} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & W^\top B_2^\top & -W^\top B_3^\top & -W^\top B_4^\top \\ B_2 W & -B_2 S B_2^\top & B_2 S B_3^\top & B_2 S B_4^\top \\ -B_3 W & B_3 S B_2^\top & -B_3 S B_3^\top & -B_3 S B_4^\top \\ -B_4 W & B_4 S B_2^\top & -B_4 S B_3^\top & -B_4 S B_4^\top \end{pmatrix}. \quad (4.52)$$

Let $\mathbf{B}_{(1)}$ and $\mathbf{B}_{(2)}$ denote the first and second matrix on the right hand side of (4.52), respectively. Then

$$\begin{aligned} \mathbf{U} &= (-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})\mathbf{B}(-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top \\ &= (-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})\mathbf{B}_{(1)}(-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top + (-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})\mathbf{B}_{(2)}(-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top. \quad (4.53) \end{aligned}$$

With (4.50), after some algebra, we find that

$$\begin{aligned} & (-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})\mathbf{B}_{(1)}(-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top \\ = \; & A_{44} + A_{41}A_{11}^{-1}A_{14} + A_{42}A_{22}^{-1}A_{24} \\ & -V\{(A^{33})^{-1}A^{32}A_{24} + A_{34}\} - \{A_{42}A^{23}(A^{33})^{-1} + A_{43}\}V^\top \\ & +V\{(A^{33})^{-1}A^{32}A_{22}A^{23}(A^{33})^{-1} + A_{33}\}V^\top. \end{aligned} \quad (4.54)$$

By (4.37), we have

$$(A^{33})^{-1}A^{32}A_{24} + A_{34} = (A^{33})^{-1}V^\top.$$

By (4.35) and (4.36), we further have

$$(A^{33})^{-1}A^{32}A_{22}A^{23}(A^{33})^{-1} + A_{33} = (A^{33})^{-1}. \quad (4.55)$$

Combining (4.54)–(4.55), we obtain

$$\begin{aligned} & (-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})\mathbf{B}_{(1)}(-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top \\ = \; & A_{44} + A_{41}A_{11}^{-1}A_{14} + A_{42}A_{22}^{-1}A_{24} - V(A^{33})^{-1}V^\top. \end{aligned}$$

Next we find the form of $(-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})\mathbf{B}_{(2)}(-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top$. A key step is to argue that

$$(\mathbf{0}, B_2^\top, -B_3^\top, -B_4^\top)(-\boldsymbol{\Lambda}_1^\top \boldsymbol{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top = \mathbf{0}. \quad (4.56)$$

After some algebra, it can be shown that

$$(\mathbf{0}, B_2^\top, -B_3^\top, -B_4^\top)(-\mathbf{\Lambda}_1^\top \mathbf{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top = -B_2^\top A_{22}^{-1} A_{24} + B_2^\top A^{23}(A^{33})^{-1} V^\top + B_3^\top V^\top - B_4^\top.$$

With (4.35)–(4.37), $(\mathbf{0}, B_2^\top, -B_3^\top, -B_4^\top)(-\mathbf{\Lambda}_1^\top \mathbf{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top$ can be rewritten as

$$\begin{aligned}
& (\mathbf{0}, B_2^\top, -B_3^\top, -B_4^\top)(-\mathbf{\Lambda}_1^\top \mathbf{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top \\
= & -\Delta^* A^{32} A_{24} + \Delta^* A^{33} A_{32} A^{23} (A^{33})^{-1} V^\top + \Delta^* A^{33} A_{33} V^\top - \Delta^* V^\top + \Delta^* A^{32} A_{24} \\
= & \Delta^* \left\{ A^{33} A_{32} A^{23} (A^{33})^{-1} + A^{33} A_{33} - \mathbf{I}_{m\times m} \right\} V^\top \\
= & \mathbf{0},
\end{aligned}$$

where in the last equation we have used a fact implied by (4.36) that

$$A_{32} A^{23} (A^{33})^{-1} = (A^{33})^{-1} - A_{33}.$$

The finishes the proof of (4.56).

With (4.56), we can simplify $(-\mathbf{\Lambda}_1^\top \mathbf{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})\mathbf{B}_{(2)}(-\mathbf{\Lambda}_1^\top \mathbf{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top$ as

$$(-\mathbf{\Lambda}_1^\top \mathbf{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})\mathbf{B}_{(2)}(-\mathbf{\Lambda}_1^\top \mathbf{\Lambda}_0^{-1}, \mathbf{I}_{p\times p})^\top = (-\mathbf{\Lambda}_1^\top \mathbf{\Lambda}_0^{-1}, \mathbf{I}_{p\times p}) \begin{pmatrix} \mathbf{0} \\ -B_2 W A_{11}^{-1} A_{14} \\ B_3 W A_{11}^{-1} A_{14} \\ B_4 W A_{11}^{-1} A_{14} \end{pmatrix} = \mathbf{0}. \quad (4.57)$$

Combining (4.53)–(4.57) leads to

$$\mathbf{U} = A_{44} + A_{41} A_{11}^{-1} A_{14} + A_{42} A_{22}^{-1} A_{24} - V (A^{33})^{-1} V^\top, \quad (4.58)$$

which together with (4.51) implies that

$$\mathbf{U} = \mathbf{\Lambda}_1^\top \mathbf{\Lambda}_0^{-1} \mathbf{\Lambda}_1 - \mathbf{\Lambda}_2.$$

This finishes the Step 2.

Therefore, we conclude that the quadratic approximation (4.30) of R_n converges to χ_p^2 in distribution as $n \to \infty$, which completes the proof of Theorem 4.1. □

Part II

The bootstrap with inequality constraints

Chapter 5

Introduction to Part II

5.1 The bootstrap method

In this chapter, we begin by introducing some basic concepts concerning the bootstrap method. This has become a popular data resampling method for statistical inference after its introduction by Efron (1979). Part of its popularity comes from its widespread applicability and its simplicity, especially when the sampling distribution of a statistic is unknown or very difficult to calculate or approximate. Efron and Tibshirani (1993), Shao and Tu (1995), and Davison and Hinkley (1997) serve as excellent reference books on the bootstrap.

In Part II of this thesis, our study focuses on the application of the bootstrap method, under parametric model settings, but with an inequality constrained parameter space. When the data is known to follow a parametric model, it is natural to use the parametric bootstrap. Therefore, here we briefly outline the parametric bootstrap procedure for estimating the sampling distribution of a scalar statistic, and give the definition of its consistency. We focus our review on the case where the observations are independent and identically distributed (*i.i.d.*). It should also be mentioned that when talking about the bootstrap, we are referring to the standard n out of n bootstrap, where n is the total sample size.

Suppose we have a sequence of *i.i.d.* random observations $\mathbf{X} = \{X_1, \ldots, X_n\}$ from an unknown distribution P. In the parametric setting, the distribution $P = P(\theta)$ is completely known, except for an unknown parameter $\theta \in \Theta$. Let $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X})$ be an estimator of θ. Here and after, we use the simplified notation $\hat{\theta}_n$. Throughout Part II of this thesis, we concentrate on using the parametric maximum likelihood estimator $\hat{\theta}_n$ of θ based on the observations $\mathbf{X}$. Let $T_n = T_n(\mathbf{X})$ be a scalar statistic of interest, and let

$$H_n(x;\theta) = \Pr(T_n \leq x|\theta)$$

denote its sampling distribution. The bootstrap method can be readily used to estimate $H_n(x;\theta)$.

Let $\mathbf{X}^* = \{X_1^*, \ldots, X_n^*\}$ be a parametric bootstrap sample such that $\mathbf{X}^*$ is an random sample from the parametric model $P(\hat{\theta}_n)$ for a given observed value $\hat{\theta}_n$. With the bootstrap sample $\mathbf{X}^*$, we can similarly calculate $T_n^* = T_n(\mathbf{X}^*)$. Let

$$H_n^*(x;\hat{\theta}_n) = \Pr(T_n^* \leq x|\hat{\theta}_n)$$

be the conditional distribution of T_n^* for a given $\hat{\theta}_n$. Then $H_n^*(x;\hat{\theta}_n)$ is called the parametric *bootstrap distribution* of T_n, which can be used to estimate $H_n(x;\theta)$.

The following is a definition of the consistency of a bootstrap estimator (Shao and Tu, 1995, Section 3.1).

Definition 5.1 (Bootstrap consistency) *Suppose ρ is a metric distance on the space of all probability measures on $\mathbb{R}$. We say $H_n^*(x;\hat{\theta}_n)$ is ρ-consistent (or weakly ρ-consistent) if*

$$\rho\left(H_n^*(x;\hat{\theta}_n), H_n(x;\theta)\right) \to 0, \tag{5.1}$$

in probability under $P(\theta)$ as $n \to \infty$.

Recall that $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X})$ is a function of random observations $\mathbf{X}$. Hence, the bootstrap distribution $H_n^*(x;\hat{\theta}_n)$ is defined as a random distribution function since it is conditional on $\hat{\theta}_n$, although we usually treat it as a conditional distribution for a given value $\hat{\theta}_n$. Note that a distance ρ is required in this definition. In the literature, Kolmogorov-Smirnov-type distance is commonly used, while Mallow's distance is also sometimes used. The consistency of the bootstrap distribution is fundamental for further investigating the validity of various bootstrap based inference methods.

5.2 Bootstrap percentile confidence intervals

Confidence interval (CI) estimation has been an important research focus throughout the theoretical development of the bootstrap (Hall, 1988; DiCiccio and Efron, 1996). There are many types of bootstrap methods for constructing CIs of a parameter. For example, at least five common methods are described in Efron and Tibshirani (1993), and their asymptotic properties can be found in Hall (1988) and Shao and Tu (1995). In Part II of this thesis, we mainly focus on the method which is based on using the percentiles of the bootstrap distribution of its estimator, which is usually referred to as the *bootstrap percentile confidence interval* (Efron and Tibshirani, 1993). This type of bootstrap CI is simple to implement and thus is widely used in practice (*referring from Hall (1988): "Our enquiries of users indicate that the percentile method is used in more than half of cases, ..."*). For a scalar parameter, the standard procedure to construct CI based on bootstrap percentile method can be described as follows.

Suppose we consider the same setup as in Section 5.1. For a cumulative distribution function $F(x)$, the α^{th} quantile of $F(x)$ is defined as $F^{-1}(\alpha) = \inf\{x : F(x) \geq \alpha\}$.

Definition 5.2 (Bootstrap percentile confidence interval) *Let* $G_n^*(x; \hat{\theta}_n) = \Pr(\hat{\theta}_n^* \leq x|\hat{\theta}_n)$ *be the bootstrap distribution function of* $\hat{\theta}_n$. *Then a* $100(1-\alpha)\%$ *bootstrap percentile CI for* θ *is constructed as*

$$\left[q^*_{\alpha_1}, q^*_{1-\alpha_2}\right] = \left[G_n^{*-1}(\alpha_1; \hat{\theta}_n), G_n^{*-1}(1-\alpha_2; \hat{\theta}_n)\right], \tag{5.2}$$

where $G_n^{*-1}(\alpha_1; \hat{\theta}_n)$ *and* $G_n^{*-1}(1-\alpha_2; \hat{\theta}_n)$ *are the* α_1^{th} *and* $(1-\alpha_2)^{\text{th}}$ *quantiles of* $G_n^*(x; \hat{\theta}_n)$, $\alpha_1, \alpha_2 \in (0, 0.5)$ *and* $\alpha = \alpha_1 + \alpha_2$.

Note that the endpoints of a confidence interval defined in (5.2) are random functions of $\hat{\theta}_n$, since they are obtained by conditioning on $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X})$ as a function of random observations $\mathbf{X}$. Then the probability $\Pr\left(\theta \in [q^*_{\alpha_1}, q^*_{1-\alpha_2}]\right)$ is called the *coverage probability* of a confidence interval (5.2) for θ, where $\Pr(\cdot)$ indicates the probability distribution of $\mathbf{X}$ under $P(\theta)$. In general, the exact coverage probability of a CI depends on the sample size n, and may need to be calculated case by case. If the exact coverage probability of a CI is not easy to calculate, it is desirable to have its *asymptotic coverage probability*

$$\lim_{n\to\infty} \Pr\left(\theta \in [q^*_{\alpha_1}, q^*_{1-\alpha_2}]\right)$$

being the nominal coverage level $1-\alpha$. This asymptotic coverage probability provides us a useful approximation to the exact finite sample coverage probability of a CI. Hence, it may help us to understand the finite sample behaviour of a CI. The following definition gives the consistency of the bootstrap percentile confidence interval (Shao and Tu, 1995, Section 4.2).

Definition 5.3 (Consistency of bootstrap percentile confidence interval) *A bootstrap percentile confidence interval for θ defined in (5.2), with nominal coverage level $1-\alpha$, is said to be consistent if*

$$\Pr\left(\theta \in \left[q^*_{\alpha_1}, q^*_{1-\alpha_2}\right]\right) \to 1-\alpha, \tag{5.3}$$

under $P(\theta)$ as $n \to \infty$.

The next theorem reviews some sufficient conditions for the consistency of bootstrap percentile CI. Let $H_n(x;\theta)$ be the sampling distribution of $T_n = n^{1/2}(\hat{\theta}_n - \theta)$, and let $H^*_n(x;\hat{\theta}_n)$ be its bootstrap distribution conditional on $\hat{\theta}_n$.

Theorem 5.1 *Suppose that the following conditions are satisfied:*

(a) $H^*_n(x;\hat{\theta}_n)$ *is consistent with* $H_n(x;\theta)$*,*

(b) $\lim_{n\to\infty}\sup_x |H_n(x;\theta) - H(x)| = 0$ *for a continuous, and strictly increasing* $H(x)$*, where its density function is symmetric around zero,*

then the bootstrap percentile confidence interval defined in (5.2) is consistent.

Despite the vast literature on the bootstrap method, its applicability under inequality constrained inference problems seems to be not well understood. As we will show in the next chapter, the conditions in Theorem 5.1 may be violated in problems with inequality constraints, and thus the bootstrap percentile confidence interval may be inconsistent.

Chapter 6

Quantifying the local asymptotic coverage probabilities of bootstrap percentile confidence intervals for constrained parameters

6.1 Introduction

As shown by various examples in Chapter 1, in many applications there is useful prior knowledge available about the parameters of interest, which can be expressed in terms of linear inequality constraints. Research on developing statistical methods and theory under such constraints has a long history; see Robertson et al. (1988) and Silvapulle and Sen (2004). The general consensus is that, if the constraint information can be properly incorporated in the analysis, we can expect to obtain more accurate estimation and more powerful statistical tests. In this chapter, we concentrate on confidence intervals for the constrained parameters when there are linear inequality constraints on the parameters under parametric models. More specifically, we aim to answer the scientific questions Q3–Q5 outlined in Section 1.1 of Chapter 1.

For regular parametric models, there are mainly three types of methods to construct a confidence interval (CI) for an unknown parameter: the Wald-, score-, and likelihood-ratio-type. Under some regularity conditions, the maximum likelihood estimator (MLE) of the unknown parameter is consistent and asymptotically normal. This property ensures that the Wald-, score-, and likelihood-ratio-type statistics are asymptotical pivotal. That is, their limiting distributions do not depend on the unknown parameter. This nice property guarantees the consistency of the resultant confidence intervals based on their limiting distributions. When there are linear inequality constraints on the parameters, the true values of the parameters may lie on the boundary of the parameter space, which violates commonly used regularity conditions. In such situation, the limiting distribution of the MLE may not be normal. Instead, it becomes a complicated function of a normal random variable or multivariate normal random vector; see Chernoff (1954) and Self and Liang (1987). Because of that, the Wald-, score-, and likelihood-ratio-type statistics are no longer asymptotical pivotal. In some situations, their limiting distributions do not have a simple analytic form. We refer to Chernoff (1954), Self and Liang (1987) and Susko (2013) for results on the likelihood-ratio-type statistic, and Silvapulle and Silvapulle (1995) and Molenberghs and Verbeke (2007) for results on the score- and Wald-type statistics.

The bootstrap method reviewed in Chapter 5 can be intuitively applied to obtain the bootstrap distributions of these statistics, which can then be used to construct confidence intervals for constrained parameters. A natural question is how does the bootstrap method perform in these situations? Surprisingly, some pioneer work, including Andrews (1997, 2000), showed that the bootstrap distribution is inconsistent with the sampling distribution of the MLE of an unknown parameter when it is on the boundary. Drton and Williams (2011) discussed the application of the bootstrap method to hypothesis testing with the likelihood ratio test. They quantified the asymptotic size of the bootstrap likelihood ratio test under a boundary hypothesis and concluded that the size can be both below or above the nominal level. Peddada (1997) studied the construction of confidence intervals for order constrained parameters by bootstrapping the point estimator introduced in Hwang and Peddada (1994). Li et al. (2010) investigated several bootstrap methods for constructing confidence intervals for ordered binomial proportions via simulations and a real data example.

To the best of our knowledge, the asymptotic behaviour of bootstrap methods for confidence interval estimation under inequality constraints still seems unclear, especially when the true value of the parameter is near the boundary. The primary goal of this chapter concerns quantifying the asymptotic coverage probabilities of the bootstrap percentile CIs as reviewed in Section 5.2 for constrained parameters. We study the important one- and two-sample problems with data generated from distributions in natural exponential family.

This chapter is organized as follows. In Section 6.2, we first consider one- and two-sample problems with data generated from normal distributions. For the one-sample problem, the mean parameter is constrained to be nonnegative; and for the two-sample problem, the mean parameter for the first sample is constrained to be smaller or equal to that of the second sample. We quantify the exact coverage probabilities of the bootstrap percentile confidence intervals for both the one- and two-sample problems. Then, in Section 6.3, under a local asymptotic setup, these results are generalized to the constrained mean parameters of distributions in natural exponential family. For presentational convenience, the proofs in Section 6.3 are deferred to Section 6.4.

6.2 Results for the normal distributions

6.2.1 One-sample normal distribution

In this section, we consider a simple but generic setup motivated by Andrews (1997). The results under this setup will shed light on the more complicated situation discussed in Section 6.3.1.

Suppose $X_1, \ldots, X_n$ is a random sample from a normal distribution with mean θ and unit variance. The parameter space of θ is constrained to be $\mathcal{C}_1 = \{\theta : \theta \geq 0\}$. We aim to quantify the coverage probability of a bootstrap percentile confidence interval of θ as reviewed in Section 5.2.

We start with finding the form of the MLE $\hat{\theta}_n$ of θ, and the sampling distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$. Based on n random observations $X_1, \ldots, X_n$ from $N(\theta, 1)$, the log-likelihood function of θ, up to a constant not dependent on θ, is

$$l_n(\theta) = -n(\bar{X}_n - \theta)^2.$$

Then the MLE of θ is defined as

$$\hat{\theta}_n = \arg\max_{\theta \in \mathcal{C}_1} l_n(\theta).$$

In the next lemma, we discuss the closed form of $\hat{\theta}_n$ and the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$. Define $\bar{X}_n = \sum_{i=1}^n X_i/n$ and $y^+ = \max(y, 0)$.

Lemma 6.1 *Suppose $X_1, \ldots, X_n$ are i.i.d. with distribution $N(\theta, 1)$ with $\theta \geq 0$ and the true value of θ is θ_0. Then $\hat{\theta}_n = \bar{X}_n^+$ and*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \stackrel{d}{=} \max(Z, -\sqrt{n}\theta_0),$$

where Z is a standard normal random variable with zero mean and unit variance. Here $\stackrel{d}{=}$ denotes "has the same distribution as".

Proof. The MLE of θ can be equivalently defined as

$$\hat{\theta}_n = \arg\min_{\theta \geq 0} (\bar{X}_n - \theta)^2.$$

It can be easily verified that $\hat{\theta}_n = \bar{X}_n^+$. Hence

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \max\{\sqrt{n}(\bar{X}_n - \theta_0), -\sqrt{n}\theta_0\},$$

which has the same distribution as $\max(Z, -\sqrt{n}\theta_0)$. This finishes the proof. □

Lemma 6.1 implies that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ has the same distribution as Z^+ when $\theta_0 = 0$. Its distribution function is not continuous, strictly increasing, and symmetric. That is, Condition (b) in Theorem 5.1 does not hold. Hence, the bootstrap percentile confidence interval reviewed in Section 5.2 may not be consistent when $\theta_0 = 0$. This is confirmed in the next proposition.

For the convenience of our presentation, we define some notation. Let $X_1^*, \ldots, X_n^*$ denote the bootstrap sample from $N(\hat{\theta}_n, 1)$, for given $\hat{\theta}_n$, and let $\hat{\theta}_n^* = \bar{X}_n^{*+}$ denote the maximum likelihood estimator of θ based on the bootstrap sample, where $\bar{X}_n^* = \sum_{i=1}^n X_i^*/n$. Further let

$$G_n^*(x;\hat{\theta}_n) = \Pr(\hat{\theta}_n^* \leq x|\hat{\theta}_n)$$

be the bootstrap distribution function of $\hat{\theta}_n$, and q_α^* be the α^{th} quantile of $G_n^*(x;\hat{\theta}_n)$. From Lemma 6.1, we note that $\hat{\theta}_n$ is actually a function of random observations $X_1,\ldots,X_n$. Hence, the bootstrap distribution $G_n^*(x;\hat{\theta}_n)$ and its quantile q_α^* are defined as random functions in terms of $\hat{\theta}_n$.

Proposition 6.1 *Suppose we consider the same setup and assumptions as in Lemma 6.1. Then*

$$\Pr\left(\theta_0 \in [q_{\alpha_1}^*, q_{1-\alpha_2}^*]\right) = \begin{cases} 1-\alpha_1-\alpha_2, & \text{if } \sqrt{n}\theta_0 > \Phi^{-1}(1-\alpha_2) \\ 1-\alpha_1, & \text{if } \sqrt{n}\theta_0 \leq \Phi^{-1}(1-\alpha_2) \end{cases},$$

where $\Pr(\cdot)$ *indicates the probability under the distribution of* $\hat{\theta}_n$ *given in Lemma 6.1,* $\Phi(\cdot)$ *is the cumulative distribution function of a standard normal random variable, and* $\alpha_1,\alpha_2 \in (0,0.5)$ *with* $\alpha = \alpha_1+\alpha_2 \in (0,1)$.

Proof. Note that

$$G_n^*(x;\hat{\theta}_n) = \begin{cases} 0 & \text{if } x<0 \\ 1-\Pr(\bar{X}_n^* > x|\hat{\theta}_n) & \text{if } x \geq 0 \end{cases} = \begin{cases} 0 & \text{if } x<0 \\ \Phi\left\{\sqrt{n}(x-\hat{\theta}_n)\right\} & \text{if } x\geq 0 \end{cases}.$$

For a given true value $\theta_0 \geq 0$, we have

$$\begin{aligned}\Pr\left(\theta_0 \in [q_{\alpha_1}^*, q_{1-\alpha_2}^*]\right) &= \Pr\left\{\alpha_1 \leq G_n^*(\theta_0;\hat{\theta}_n) \leq 1-\alpha_2\right\} \\ &= \Pr\left[\alpha_1 \leq \Phi\left\{\sqrt{n}(\theta_0-\hat{\theta}_n)\right\} \leq 1-\alpha_2\right].\end{aligned}$$

Then by the property of $\Phi(\cdot)$, the above equation becomes

$$\begin{aligned}\Pr\left(\theta_0 \in [q_{\alpha_1}^*, q_{1-\alpha_2}^*]\right) &= \Pr\left\{\Phi^{-1}(\alpha_2) \leq \sqrt{n}(\hat{\theta}_n-\theta_0) \leq \Phi^{-1}(1-\alpha_1)\right\} \\ &= \Pr\left\{\sqrt{n}(\hat{\theta}_n-\theta_0) \leq \Phi^{-1}(1-\alpha_1)\right\} - \Pr\left\{\sqrt{n}(\hat{\theta}_n-\theta_0) < \Phi^{-1}(\alpha_2)\right\}.\end{aligned}$$

By Lemma 6.1, we have

$$\begin{aligned}\Pr\left(\theta_0 \in [q^*_{\alpha_1}, q^*_{1-\alpha_2}]\right) &= \Phi\{\Phi^{-1}(1-\alpha_1)\} - \Pr\left\{\sqrt{n}(\hat{\theta}_n - \theta_0) < \Phi^{-1}(\alpha_2)\right\} \\ &= \begin{cases} 1-\alpha_1-\alpha_2, & \text{if } \sqrt{n}\theta_0 > \Phi^{-1}(1-\alpha_2) \\ 1-\alpha_1, & \text{if } \sqrt{n}\theta_0 \le \Phi^{-1}(1-\alpha_2) \end{cases}.\end{aligned}$$

This finishes the proof. □

We comment that the exact coverage probability in Proposition 6.1 is a left-continuous, piecewise constant, function of θ_0 with a jump at $\Phi^{-1}(1-\alpha_2)/\sqrt{n}$, for given n and α_2 values. In Figure 6.1, we plot the exact coverage probabilities versus the true value of θ_0 for different sample sizes n at level $\alpha = 0.10$ with $\alpha_1 = \alpha_2 = \alpha/2$ as an illustration. However, we note that different choices of α_1 and α_2 should give different graphs.

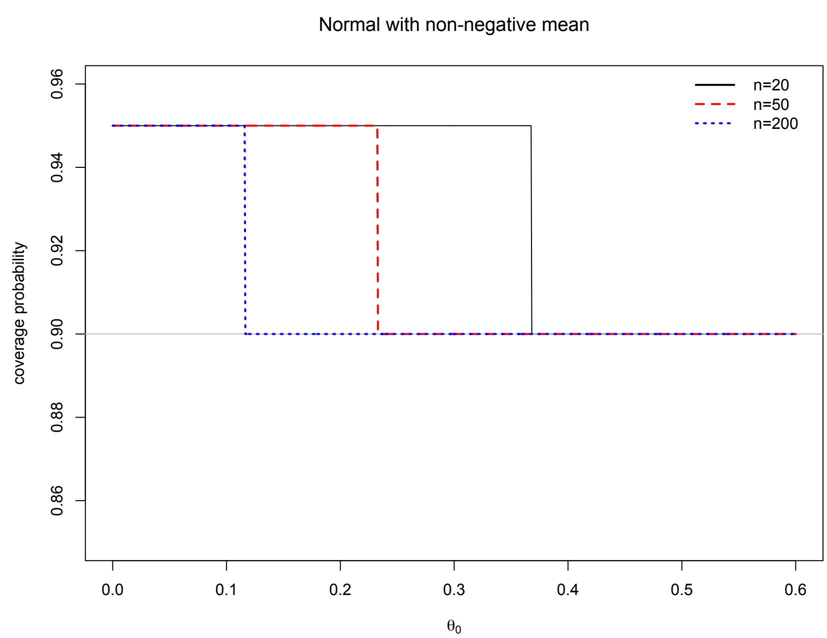

Figure 6.1: Coverage probability of 90% bootstrap percentile CI for the non-negative mean of univariate normal distribution with $\sigma^2 = 1$ and $\alpha_1 = \alpha_2 = 0.05$.

In general, we can observe from Figure 6.1, that the bootstrap percentile CI behaves conservatively in terms of its coverage probability. Specifically, when the true value θ_0 is on, or close to, the boundary, a non-regular behaviour, that the bootstrap percentile CI over covers the true parameter, can happen. Note that the degree of closeness of θ_0 to the boundary crucially depends on the sample size by an order of $\sqrt{n}$.

6.2.2 Two-sample normal distributions

The two-sample problem for the means is of great importance in statistical inference. In this section, we consider the case of two independent normal populations with equal and known variances, both set to be one, where the two means are known to satisfy an ordering constraint.

Suppose we have independent observations X_{ij}'s such that $X_{ij} \sim N(\theta_i, 1)$ for $j = 1, \ldots, n_i$ and $i = 1, 2$. The parameter space of (θ_1, θ_2) is constrained to be $\mathcal{C}_2 = \{(\theta_1, \theta_2) : \theta_2 \geq \theta_1\}$. We aim to quantify the coverage probabilities of bootstrap percentile CIs for θ_1, θ_2, and their difference $\Delta = \theta_2 - \theta_1$.

As a first step, we identify the form of the MLE of $(\theta_1, \theta_2, \Delta)$. Let $n = n_1 + n_2$. The log-likelihood function of (θ_1, θ_2), up to a constant not dependent on the unknown parameters, is

$$l_n(\theta_1, \theta_2) = -n_1(\bar{X}_{n1} - \theta_1)^2 - n_2(\bar{X}_{n2} - \theta_2)^2,$$

where $\bar{X}_{ni} = \sum_{j=1}^{n_i} X_{ij}/n_i$, $i = 1, 2$. The MLEs of θ_1 and θ_2 are defined as

$$(\hat{\theta}_{n1}, \hat{\theta}_{n2}) = \arg \max_{(\theta_1, \theta_2) \in \mathcal{C}_2} l_n(\theta_1, \theta_2) = \arg \min_{\theta_1 \leq \theta_2} \left\{ n_1(\bar{X}_{n1} - \theta_1)^2 + n_2(\bar{X}_{n2} - \theta_2)^2 \right\}$$

and the MLE of Δ is $\hat{\Delta}_n = \hat{\theta}_{n2} - \hat{\theta}_{n1}$.

Suppose the true value of (θ_1, θ_2) is $(\theta_{10}, \theta_{20})$ and let $\Delta_0 = \theta_{20} - \theta_{10}$. In the next lemma, we summarize the explicit form of $(\hat{\theta}_{n1}, \hat{\theta}_{n2}, \hat{\Delta}_n)$ and the marginal sampling distributions of $\sqrt{n}(\hat{\theta}_{n1} - \theta_{10})$, $\sqrt{n}(\hat{\theta}_{n2} - \theta_{20})$, and $\sqrt{n}(\hat{\Delta}_n - \Delta_0)$. For the convenience of presentation, we let $\omega = n_1/n \in (0, 1)$.

Lemma 6.2 *Suppose X_{ij}'s are independent, $X_{ij} \sim N(\theta_i, 1)$ for $j = 1, \ldots, n_i$ and $i = 1, 2$, and the true value of $(\theta_1, \theta_2, \Delta)$ is $(\theta_{10}, \theta_{20}, \Delta_0)$. Then* *(a) the MLE of (θ_1, θ_2), subject to the constraint in $\mathcal{C}_2$, is*

$$\hat{\theta}_{n1} = \min\{\bar{X}_{n1}, \omega\bar{X}_{n1} + (1-\omega)\bar{X}_{n2}\}, \quad \hat{\theta}_{n2} = \max\{\bar{X}_{n2}, \omega\bar{X}_{n1} + (1-\omega)\bar{X}_{n2}\},$$

and the MLE of Δ is $\hat{\Delta}_n = (\bar{X}_{n2} - \bar{X}_{n1})^+$;

(b)

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_{n1} - \theta_{10}) &\overset{d}{=} \min\left\{\sqrt{\frac{1}{\omega}}Z_1, \sqrt{\omega}Z_1 + \sqrt{1-\omega}Z_2 + (1-\omega)\sqrt{n}\Delta_0\right\}, \\
\sqrt{n}(\hat{\theta}_{n2} - \theta_{20}) &\overset{d}{=} \max\left\{\sqrt{\frac{1}{1-\omega}}Z_2, \sqrt{\omega}Z_1 + \sqrt{1-\omega}Z_2 - \omega\sqrt{n}\Delta_0\right\}, \\
\sqrt{n}(\hat{\Delta}_n - \Delta_0) &\overset{d}{=} \max\left\{\sqrt{\frac{1}{1-\omega}}Z_2 - \sqrt{\frac{1}{\omega}}Z_1, -\sqrt{n}\Delta_0\right\},
\end{aligned}$$

where Z_1, Z_2 are two independent standard normal random variables.

Proof. We first consider Part (a). To identify the forms of $(\hat{\theta}_{n1}, \hat{\theta}_{n2})$, we consider the following sample sizes dependent reparameterization

$$\theta_1 = \eta - (1-\omega)\Delta, \quad \theta_2 = \eta + \omega\Delta,$$

or equivalently

$$\eta = \omega\theta_1 + (1-\omega)\theta_2, \quad \Delta = \theta_2 - \theta_1.$$

Under the above reparameterization, the constraint $\theta_1 \le \theta_2$ then becomes $\Delta \ge 0$, while η is free of restriction.

Recall the log-likelihood function $l_n(\theta_1, \theta_2)$. Then the MLE of (η, Δ) is

$$(\hat{\eta}_n, \hat{\Delta}_n) = \arg\min_{\eta,\Delta}\left[n_1\left\{\bar{X}_{n1} - \eta + (1-\omega)\Delta\right\}^2 + n_2(\bar{X}_{n2} - \eta - \omega\Delta)^2\right]$$

subject to the constraint $\Delta \ge 0$. After some algebra, we find

$$(\hat{\eta}_n, \hat{\Delta}_n) = \arg\min_{\eta,\Delta}\left[n\left\{\eta - \omega\bar{X}_{n1} - (1-\omega)\bar{X}_{n2}\right\}^2 + n\omega(1-\omega)\{\Delta - (\bar{X}_{n2} - \bar{X}_{n1})\}^2\right]$$

subject to the constraint $\Delta \geq 0$. Hence

$$\hat{\eta}_n = \omega \bar{X}_{n1} + (1-\omega)\bar{X}_{n2}, \quad \hat{\Delta}_n = (\bar{X}_{n2} - \bar{X}_{n1})^+,$$

which implies that

$$\begin{aligned} \hat{\theta}_{n1} &= \hat{\eta}_n - (1-\omega)\hat{\Delta}_n = \min\{\bar{X}_{n1}, \omega\bar{X}_{n1} + (1-\omega)\bar{X}_{n2}\}, \\ \hat{\theta}_{n2} &= \hat{\eta}_n + \omega\hat{\Delta}_n = \max\{\bar{X}_{n2}, \omega\bar{X}_{n1} + (1-\omega)\bar{X}_{n2}\}. \end{aligned}$$

This finishes the proof of Part (a).

We now come to Part (b). It can be easily verify that

$$\begin{aligned} &\sqrt{n}(\hat{\theta}_1 - \theta_{10}) \\ = &\min\left\{\sqrt{\frac{1}{\omega}}\sqrt{n_1}(\bar{X}_{n1} - \theta_{10}), \sqrt{\omega}\sqrt{n_1}(\bar{X}_{n1} - \theta_{10}) + \sqrt{1-\omega}\sqrt{n_2}(\bar{X}_{n2} - \theta_{20}) + (1-\omega)\sqrt{n}\Delta_0\right\}, \end{aligned}$$

which has the same distribution as $\min\left\{\sqrt{\frac{1}{\omega}}Z_1, \sqrt{\omega}Z_1 + \sqrt{1-\omega}Z_2 + (1-\omega)\sqrt{n}\Delta_0\right\}$. The sampling distributions of $\sqrt{n}(\hat{\theta}_{n2} - \theta_{20})$, and $\sqrt{n}(\hat{\Delta}_n - \Delta_0)$ can be similarly derived, and are thus omitted here. This finishes the proof of Part (b). $\square$

Lemma 6.2 implies that the sampling distributions of $\sqrt{n}(\hat{\theta}_{n1} - \theta_{10})$ and $\sqrt{n}(\hat{\theta}_{n2} - \theta_{20})$ are continuous, and strictly increasing, but their densities not symmetric, and the sampling distribution of $\sqrt{n}(\hat{\Delta}_n - \Delta_0)$ is not continuous and strictly increasing, and its density is not symmetric, when $\Delta_0 = 0$. That is, Condition (b) in Theorem 5.1 does not hold for $\sqrt{n}(\hat{\theta}_{n1} - \theta_{10})$, $\sqrt{n}(\hat{\theta}_{n2} - \theta_{20})$, and $\sqrt{n}(\hat{\Delta}_n - \Delta_0)$. Hence, the bootstrap percentile confidence intervals for θ_1, θ_2, and Δ may not be consistent when $\Delta_0 = 0$. This is confirmed in in the next proposition.

For the convenience of presentation, we define some notation. Let X_{ij}^*'s be the bootstrap sample such that $X_{11}^*, \ldots, X_{1n_1}^* \sim N(\hat{\theta}_{n1}, 1)$ for given $\hat{\theta}_{n1}$, and independently, $X_{21}^*, \ldots, X_{2n_2}^* \sim N(\hat{\theta}_{n2}, 1)$ for given $\hat{\theta}_{n2}$. Define $\bar{X}_{ni}^* = \sum_{j=1}^{n_i} X_{ij}^*/n_i$, $i = 1, 2$. Further, let

$$\hat{\theta}_{n1}^* = \min\{\bar{X}_{n1}^*, \omega\bar{X}_{n1}^* + (1-\omega)\bar{X}_{n2}^*\}, \quad \hat{\theta}_{n2}^* = \max\{\bar{X}_{n2}^*, \omega\bar{X}_{n1}^* + (1-\omega)\bar{X}_{n2}^*\}, \quad \hat{\Delta}_n^* = (\bar{X}_{n2}^* - \bar{X}_{n1}^*)^+$$

be the MLEs of θ_1, θ_2, and Δ, respectively, based on the bootstrap sample. Denote the bootstrap distributions of $\hat{\theta}_{n1}$, $\hat{\theta}_{n2}$, and $\hat{\Delta}_n$ respectively by

$$G_{n1}^*(x; \hat{\theta}_{n1}, \hat{\theta}_{n2}) = \Pr(\hat{\theta}_{n1}^* \leq x | \hat{\theta}_{n1}, \hat{\theta}_{n2}),$$

$$
\begin{aligned}
G^*_{n2}(x;\hat{\theta}_{n1},\hat{\theta}_{n2}) &= \Pr(\hat{\theta}^*_{n2} \le x|\hat{\theta}_{n1},\hat{\theta}_{n2}),\\
G^*_{n,\Delta}(x;\hat{\theta}_{n1},\hat{\theta}_{n2}) &= \Pr(\hat{\Delta}^*_{n} \le x|\hat{\theta}_{n1},\hat{\theta}_{n2}),
\end{aligned}
$$

and their α^{th} quantiles by $q^*_{1,\alpha}$, $q^*_{2,\alpha}$, and $q^*_{\Delta,\alpha}$. From Part (a) of Lemma 6.2, we note that $\hat{\theta}_{n1}$ and $\hat{\theta}_{n2}$ are both functions of random observations X_{ij}'s. Hence, as defined above, the bootstrap distributions $G^*_{n1}(x;\hat{\theta}_{n1},\hat{\theta}_{n2})$, $G^*_{n2}(x;\hat{\theta}_{n1},\hat{\theta}_{n2})$, and $G^*_{n,\Delta}(x;\hat{\theta}_{n1},\hat{\theta}_{n2})$, and their corresponding quantiles $q^*_{1,\alpha}$, $q^*_{2,\alpha}$, and $q^*_{\Delta,\alpha}$ are random functions in terms of $(\hat{\theta}_{n1},\hat{\theta}_{n2})$.

Further, let $\boldsymbol{\Phi}_{(\mathbf{0},\boldsymbol{\Sigma})}(x,y)$ denote the joint cumulative distribution function (cdf) of a bivariate normal random vector with mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, and $F_{12}(x,y)$ denote the joint cdf of $\sqrt{n}(\hat{\theta}_{n1}-\theta_{10})$ and $\sqrt{n}(\hat{\theta}_{n2}-\theta_{20})$. That is, from Part (b) of Lemma 6.2, $F_{12}(x,y)$ is the joint cdf of

$$
\min\left\{\sqrt{\frac{1}{\omega}}Z_1, \sqrt{\omega}Z_1+\sqrt{1-\omega}Z_2+(1-\omega)\sqrt{n}\Delta_0\right\},
$$

and

$$
\max\left\{\sqrt{\frac{1}{1-\omega}}Z_2, \sqrt{\omega}Z_1+\sqrt{1-\omega}Z_2-\omega\sqrt{n}\Delta_0\right\}.
$$

Proposition 6.2 *Suppose we consider the same setup and assumptions as in Lemma 6.2. For* $\alpha_1,\alpha_2\in(0,0.5)$ *and* $\alpha_1+\alpha_2=\alpha$*, we have*

(a)

$$
\Pr\left(\Delta_0\in[q^*_{\Delta,\alpha_1},q^*_{\Delta,1-\alpha_2}]\right) = \begin{cases} 1-\alpha_1-\alpha_2, & \text{if } \sqrt{n}\Delta_0 > \Phi^{-1}(1-\alpha_2)/\sqrt{\omega(1-\omega)} \\ 1-\alpha_1, & \text{if } \sqrt{n}\Delta_0 \le \Phi^{-1}(1-\alpha_2)/\sqrt{\omega(1-\omega)} \end{cases};
$$

(b)

$$
\Pr\left(\theta_{10}\in[q^*_{1,\alpha_1},q^*_{1,1-\alpha_2}]\right) = \iint I\{\alpha_1\le g_1(x,y)\le 1-\alpha_2\}dF_{12}(x,y),
$$

where

$$
g_1(x,y)=\Phi\left\{-C_{11}(x)\right\}+\Phi\left\{-C_{12}(x,y)\right\}-\boldsymbol{\Phi}_{(\mathbf{0},\boldsymbol{\Lambda}_1)}\left\{-C_{11}(x),-C_{12}(x,y)\right\},
$$

with $C_{11}(x)=\sqrt{\omega}x$*,* $C_{12}(x,y)=\omega x+(1-\omega)y+\sqrt{n}(1-\omega)\Delta_0$*, and*

$$
\boldsymbol{\Lambda}_1=\begin{pmatrix} 1 & \sqrt{\omega} \\ \sqrt{\omega} & 1 \end{pmatrix};
$$

(c)

$$\Pr\left(\theta_{20} \in [q^*_{2,\alpha_1}, q^*_{2,1-\alpha_2}]\right) = \iint I\{\alpha_1 \le g_2(x,y) \le 1-\alpha_2\} dF_{12}(x,y),$$

where

$$g_2(x,y) = \mathbf{\Phi}_{(\mathbf{0},\mathbf{\Lambda}_2)}\left\{-C_{21}(y), -C_{22}(x,y)\right\},$$

with $C_{21}(y) = \sqrt{1-\omega}y$, $C_{22}(x,y) = \omega x + (1-\omega)y - \sqrt{n}\omega\Delta_0$, *and*

$$\mathbf{\Lambda}_2 = \begin{pmatrix} 1 & \sqrt{1-\omega} \\ \sqrt{1-\omega} & 1 \end{pmatrix}.$$

Proof. For Part (a), the proof is similar to that of Proposition 6.1.

For Part (b), we first find $G^*_{n1}(x;\hat{\theta}_{n1},\hat{\theta}_{n2})$ as

$$\begin{aligned} G^*_{n1}(x;\hat{\theta}_{n1},\hat{\theta}_{n2}) &= \Pr\left\{\bar{X}^*_{n1} \le x \ \text{ or } \ \omega\bar{X}^*_{n1} + (1-\omega)\bar{X}^*_{n2} \le x|\hat{\theta}_{n1},\hat{\theta}_{n2}\right\} \\ &= \Pr(\bar{X}^*_{n1} \le x|\hat{\theta}_{n1},\hat{\theta}_{n2}) + \Pr\left\{\omega\bar{X}^*_{n1} + (1-\omega)\bar{X}^*_{n2} \le x|\hat{\theta}_{n1},\hat{\theta}_{n2}\right\} \\ &\quad -\Pr\left\{\bar{X}^*_{n1} \le x, \omega\bar{X}^*_{n1} + (1-\omega)\bar{X}^*_{n2} \le x|\hat{\theta}_{n1},\hat{\theta}_{n2}\right\}. \end{aligned}$$

Note that given $\hat{\theta}_{n1}$ and $\hat{\theta}_{n2}$,

$$\begin{pmatrix} \sqrt{n\omega}(\bar{X}^*_{n1} - \hat{\theta}_{n1}) \\ \sqrt{n}\{\omega\bar{X}^*_{n1} + (1-\omega)\bar{X}^*_{n1} - \omega\hat{\theta}_{n1} - (1-\omega)\hat{\theta}_{n2}\} \end{pmatrix} \sim N\left(\mathbf{0}, \Lambda_1\right).$$

Then

$$\begin{aligned} G^*_{n1}(x;\hat{\theta}_{n1},\hat{\theta}_{n2}) &= \Phi\{\sqrt{n\omega}(x-\hat{\theta}_{n1})\} + \Phi\{\sqrt{n}(x-\omega\hat{\theta}_{n1}-(1-\omega)\hat{\theta}_{n2})\} \\ &\quad -\mathbf{\Phi}_{(\mathbf{0},\mathbf{\Lambda}_1)}\left\{\sqrt{n\omega}(x-\hat{\theta}_{n1}), \sqrt{n}(x-\omega\hat{\theta}_{n1}-(1-\omega)\hat{\theta}_{n2})\right\}, \end{aligned}$$

which is a continuous and strictly increasing function. The coverage probability is

$$\Pr\left(\theta_{10} \in [q^*_{1,\alpha_1}, q^*_{1,1-\alpha_2}]\right) = \Pr\left\{\alpha_1 \le G^*_{n1}(\theta_{10};\hat{\theta}_{n1},\hat{\theta}_{n2}) \le 1-\alpha_2\right\}.$$

With $g_1(x,y)$ defined in Part (b), we further have

$$\Pr\left(\theta_{10} \in [q^*_{1,\alpha_1}, q^*_{1,1-\alpha_2}]\right) = \Pr\left[\alpha_1 \le g_1\left\{\sqrt{n}(\hat{\theta}_{n1}-\theta_{10}), \sqrt{n}(\hat{\theta}_{n2}-\theta_{20})\right\} \le 1-\alpha_2\right].$$

Together with the definition of $F_{12}(x, y)$, this finishes the proof of Part (b).

For Part (c), the proof is similar to that of Part (b), and is thus omitted. $\square$

It appears that explicit expressions for the coverage probabilities of θ_1 and θ_2 are not available. Fortunately, the coverage probabilities of θ_1 and θ_2 are written in terms of bivariate integrals that can be easily evaluated using numerical methods. In Figure 6.2, we plot the coverage probabilities for Δ, θ_1 and θ_2 versus the true mean difference Δ_0, in the case of sample sizes $n_1 = 25$ and $n_2 = 75$ at level $\alpha = 0.10$ with $\alpha_1 = \alpha_2 = 0.05$ as an illustration. Again, we note that different choices of α_1 and α_2 should give different graphs.

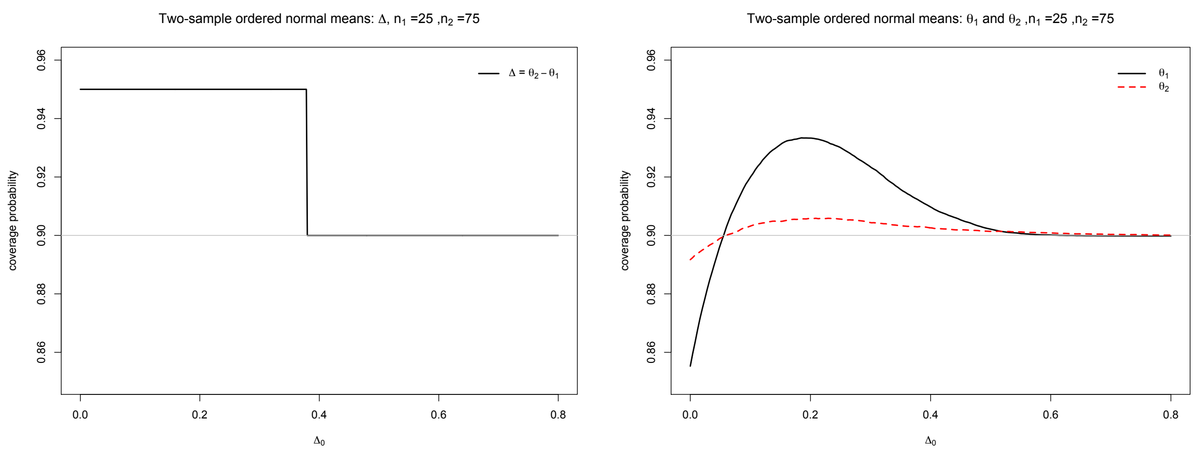

Figure 6.2: Coverage probabilities of of 90% bootstrap percentile CIs for two normal means with ordering constraint, and common $\sigma^2 = 1$ and $\alpha_1 = \alpha_2 = 0.05$. Left panel is for the mean difference Δ, and right panel is for the two-sample means θ_1 and θ_2.

From Figure 6.2, we observe that the coverage probabilities for θ_1 and θ_2 can be both greater or smaller than the nominal level, no matter what the values of θ_{10} and θ_{20} are, but only depend on their true difference Δ_0. These exact results show that the bootstrap percentile CIs for θ_1 and θ_2 are no longer always conservative, but can significantly under cover the corresponding true parameters when the constrained parameter is on, or close to, the boundary. On the other hand, when the two-sample mean difference Δ is the parameter of interest, the bootstrap percentile CI for Δ is still conservative.

6.3 Results for natural exponential family

In the preceding section, we examined the exact finite sample coverage probabilities of bootstrap percentile CIs under linear inequality constraints on the mean parameters when data is generated from normal population(s). In many applications (see Section 1.1), observations may come from distributions such as binomial or Poisson, in which their mean parameters are subject to linear inequality constraints. In this section, we generalize the results for normal distributions in Section 6.2 to the distributions belonging to the general class of natural exponential family (NEF). For the convenience of presentation, all the proofs in this section are deferred to Section 6.4.

6.3.1 One-sample NEF

In this section, we consider the one-sample problem with the linear inequality constraints on the mean parameter when data is generated from the natural exponential family of distributions.

Suppose $X_1, \ldots, X_n$ is an *i.i.d.* sample from a distribution in the natural exponential family with probability density function (pdf) or probability mass function (pmf)

$$f(x; \theta) = a(x) \exp \{\psi x - b(\psi)\}, \tag{6.1}$$

where ψ is the natural parameter, and $\theta = E(X_1) = b'(\psi)$ represents the mean parameter. It is assumed that $b(\cdot)$ is twice continuously differentiable with $b''(\psi)$ always positive. Let $\sigma^2 = b''(\psi)$ be the variance of X_1 under $f(x; \theta)$. The parameter space of the mean θ of X_1 is constrained by $\mathcal{C}_3 = \{\theta : \theta \geq d\}$ for some fixed boundary d. Under this general parametric setup, the explicit form of the coverage probability of the percentile confidence interval of θ is typically unavailable, or has to be derived case by case. Therefore, it is of interest to quantify the asymptotic coverage probability of the bootstrap percentile CI for θ.

Similarly to Section 6.2.1, we first find the form of the MLE of θ. Based on n random observations from (6.1), the log-likelihood function of θ, up to a constant not dependent on θ, is

$$l_n(\theta) = \psi \sum_{i=1}^{n} X_i - nb(\psi).$$

Then the MLE of θ is defined as

$$\hat{\theta}_n = \arg\max_{\theta \in \mathcal{C}_3} l_n(\theta).$$

The following lemma finds the closed form of $\hat{\theta}_n$.

Lemma 6.3 *Suppose $X_1, \ldots, X_n$ is an i.i.d. random sample from $f(x;\theta)$ defined in (6.1). The MLE of θ subject to the constraint $\mathcal{C}_3$ is $\hat{\theta}_n = \max(\bar{X}_n, d)$ where $\bar{X}_n = \sum_{i=1}^{n} X_i/n$.*

Let θ_0 be the true value of θ. Next, we investigate the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$, which plays an important role in deriving the asymptotic coverage probability of the bootstrap percentile CI of θ. As we have seen in Section 6.2.1, the coverage probability for θ depends on how close the true value θ_0 is to the boundary. The magnitude of "closeness" crucially depends on how large the sample size is. Treating θ_0 as a fixed value may not be helpful for understanding the subtle results in Proposition 6.1 and Figure 6.1. Hence, motivated by Proposition 6.1, we adopt a local asymptotic framework by allowing the true constrained parameter varying in a $n^{-1/2}$ neighbourhood of the boundary. More specifically, we let $\theta_0 = \theta_{0,n} = d + \tau n^{-1/2}$. The corresponding local parameter $\tau = n^{1/2}(\theta_{0,n} - d)$ controls the order of closeness of $\theta_{0,n}$ approaching to the boundary d. This setup helps capture the asymptotic distribution of the MLE in terms of τ, as stated in the next lemma.

Lemma 6.4 *Suppose $X_1, \ldots, X_n$ is a random sample from $f(x;\theta)$ defined in (6.1), and the true value of θ is $\theta_{0,n} = d + \tau n^{-1/2}$ with τ being a fixed nonnegative local parameter not depending on n. Let $\sigma_0^2 = b''(\psi_0)$ with $\psi_0 = b'^{-1}(d)$. Then*

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta_{0,n})}{\sigma_0} \overset{d}{\to} \max\left(Z, -\frac{\tau}{\sigma_0}\right),$$

as $n \to \infty$, where Z is a standard normal random variable. Here $\overset{d}{\to}$ denotes "converges in distribution".

Lemmas 6.3 and 6.4 together extend the results in Lemma 6.1 from the normal distribution to the class of distributions in NEF. With the help of Lemmas 6.3 and 6.4, we next quantify the local asymptotic coverage probability of the bootstrap percentile CI for θ in the next proposition. Let q^*_α be the α^{th} quantile of the bootstrap distribution of $\hat{\theta}_n$. Recall that the bootstrap distribution of $\hat{\theta}_n$ is similar to that considered in Section 6.2.1.

Theorem 6.1 *Suppose we consider the same setup and assumptions as in Lemma 6.4. Then*

$$\Pr\left(\theta_{0,n} \in [q^*_{\alpha_1}, q^*_{1-\alpha_2}]\right) \quad \rightarrow \quad \begin{cases} 1-\alpha_1-\alpha_2, & \text{if } \tau > \Phi^{-1}(1-\alpha_2)\sigma_0 \\ 1-\alpha_1, & \text{if } \tau < \Phi^{-1}(1-\alpha_2)\sigma_0 \end{cases},$$

as $n \to \infty$*, where* $\alpha = \alpha_1 + \alpha_2 \in (0,1)$ *with* $\alpha_1, \alpha_2 \in (0, 0.5)$.

This result generalizes the exact finite sample result in Proposition 6.1 for single normal distribution to cover the NEF of distributions. If such an approximation remains good for finite n, then this local asymptotic result provides us with information on how likely we are to have conservative conclusions, when $\theta_{0,n}$ is shrinking close to d.

To make our asymptotic results in Proposition 6.1 more appealing for practitioners, we illustrate with two commonly used, and important, distributions in the NEF. Recall the motivating examples in Chapter 1. The binomial proportion with boundary constraint finds important applications in genetic linkage analysis (Example 1.4). The Poisson rate with boundary constraint finds important applications in the signal with background noise problem (Example 1.3). Next, we compare the exact coverage probability and asymptotic coverage probability of θ under binomial and Poisson distributions.

Example 6.1 (One-sample binomial example) *Suppose* $X_1, \ldots, X_n$ *is a random sample from a Binomial*(m,p) *distribution with known* m *(for simplicity, let* $m = 1$*), and the mean* p *subject to constraint* $p \in [d, 1]$ *with* $0 < d < 1$*. For illustration, we consider* $d = 0.5$*. Under the this setup,* $\sigma_0^2 = d(1-d) = 0.25$*.*

Suppose the true value of p *is* $p_{0,n} = 0.5 + \tau n^{-1/2}$*. Applying Proposition 6.1, the local asymptotic coverage probability of the* $100(1-\alpha)\%$ *bootstrap percentile CI of* p *is*

$$\lim_{n\to\infty} \Pr\left(p_{0,n} \in [q^*_{\alpha_1}, q^*_{1-\alpha_2}]\right) \quad = \quad \begin{cases} 1-\alpha_1-\alpha_2, & \text{if } \tau > 0.5\Phi^{-1}(1-\alpha_2) \\ 1-\alpha_1, & \text{if } \tau < 0.5\Phi^{-1}(1-\alpha_2) \end{cases}.$$

The exact coverage probability of the $100(1-\alpha)\%$ *bootstrap percentile CI for* p *can also be calculated by noting the fact that a sum of independent binomial random variables still has a binomial distribution. Combining these results, in Figure 6.3, we plot the asymptotic and exact coverage probabilities of the bootstrap percentile CI as functions of* $p_0 = p_{0,n}$. *We note that different choices of* α_1 *and* α_2 *should give different graphs. For comparison, we also add the exact coverage probability of the bootstrap percentile CI for* p *without constraint.*

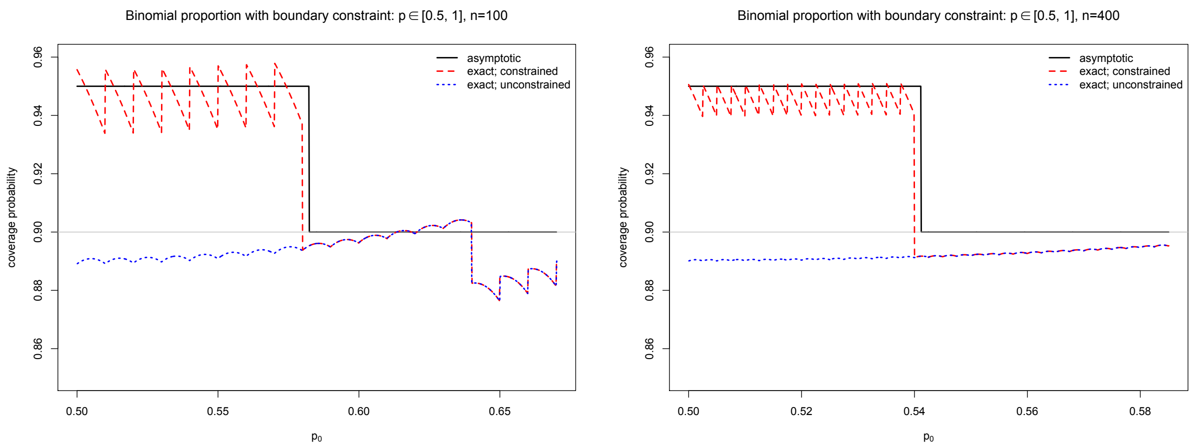

Figure 6.3: Asymptotic and exact coverage probabilities of 90% bootstrap percentile CI for the proportion p of univariate binomial distribution with $0.5 \le p \le 1$ and $\alpha_1 = \alpha_2 = 0.05$. Left panel is for $n = 100$, and right panel is for $n = 400$.

In Figure 6.3, we observe chaotic behaviour with oscillation phenomenon of the coverage probabilities due to the discrete nature of the binomial distribution. Hence, we can not always expect the quantified exact coverage probabilities to achieve the nominal level, even for the exact unconstrained case. In general, we can see a clear trend that the quantified asymptotic local coverage probability shows a close agreement with the exact coverage probability, as functions of p_0, *especially when the sample size increases.*

Example 6.2 (One-sample Poisson example) *Suppose* $X_1, \ldots, X_n$ *is a random sample from a Poisson*(λ) *distribution with mean* λ *subject to constraint* $\lambda \in [d, \infty)$ *with* $d > 0$. *For illustration, we consider* $d = 2$. *Under this setup,* $\sigma_0^2 = d$.

Suppose the true value of λ is $\lambda_{0,n} = 2 + \tau n^{-1/2}$. Applying Proposition 6.1, we have the local asymptotic coverage probability of the $100(1-\alpha)\%$ bootstrap percentile CI of λ is

$$\lim_{n\to\infty} \Pr\left(\lambda_{0,n} \in [q^*_{\alpha_1}, q^*_{1-\alpha_2}]\right) = \begin{cases} 1-\alpha_1-\alpha_2, & \textit{if } \tau > \Phi^{-1}(1-\alpha_2)\sqrt{2} \\ 1-\alpha_1, & \textit{if } \tau < \Phi^{-1}(1-\alpha_2)\sqrt{2} \end{cases}.$$

The exact coverage probability of the $100(1-\alpha)\%$ bootstrap percentile CI for λ can also be calculated by noting the fact that a sum of independent Poisson random variables is still has a Poisson distribution. In Figure 6.4, we plot the asymptotic and exact coverage probabilities as functions of the true parameter $\lambda_0 = \lambda_{0,n}$. We note that different choices of α_1 and α_2 should give different graphs. For comparison, we also add the exact coverage probabilities of the bootstrap percentile CI for λ without constraint.

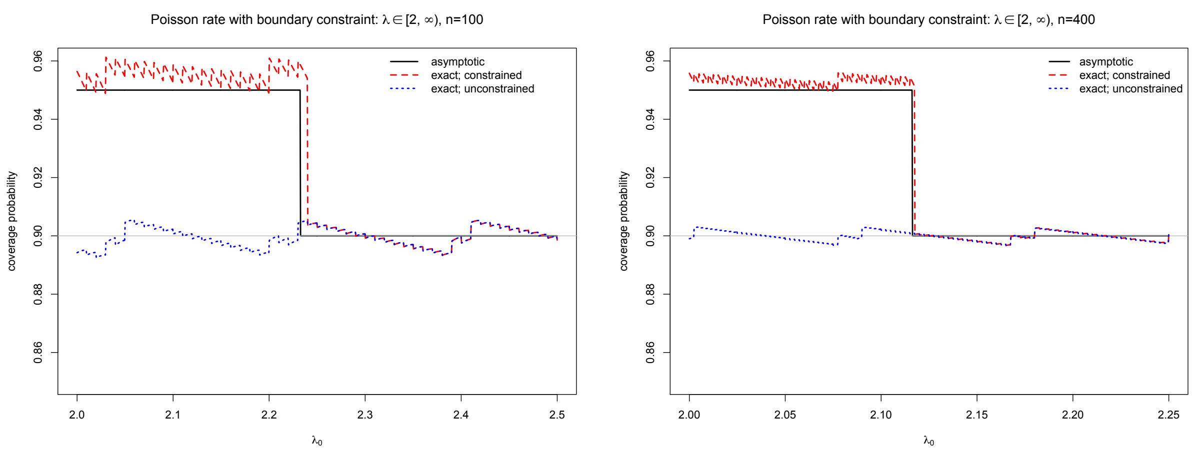

Figure 6.4: Asymptotic and exact coverage probabilities of 90% bootstrap percentile CI for the rate λ of univariate Poisson distribution with $\lambda \geq 2$ and $\alpha_1 = \alpha_2 = 0.05$. Left panel is for $n = 100$, and right panel is for $n = 400$.

Again, we can not always expect these exact coverage probabilities to achieve the nominal level due to the discrete nature of the Poisson distribution. In general, we can see a clear trend that the quantified asymptotic coverage probability shows close agreement with the exact finite sample coverage probability, as functions of λ_0, especially when the sample size increases.

6.3.2 Two-sample NEF

In this section, we generalize the results developed for the two-sample normal distributions in Section 6.2.2 to distributions in NEF.

Suppose we have *i.i.d.* observations $X_{11}, \ldots, X_{1n_1}$ from $f(x;\theta_1)$, and independently, we have *i.i.d.* observations $X_{21}, \ldots, X_{2n_2}$ from $f(x;\theta_2)$, with $f(x;\theta)$ belonging to the natural exponential family. That is, $f(x;\theta)$ satisfies (6.1) and θ still represents the mean parameter. The parameter space of the two means is defined to be $\mathcal{C}_4 = \{(\theta_1,\theta_2) : \theta_2 \geq \theta_1\}$. Our interest is to quantify the asymptotic coverage probabilities of the bootstrap percentile confidence intervals of θ_1, θ_2, and their difference Δ.

As a first step, we identify the form of MLE of $(\theta_1,\theta_2,\Delta)$. Let $n = n_1 + n_2$. Based on the n random observations, the log-likelihood function, up to a constant not dependent on the unknown parameters, is

$$l_n(\theta_1,\theta_2) = \sum_{i=1}^{2} \left\{ \psi_i \sum_{j=1}^{n_i} X_{ij} - nb(\psi_i) \right\}.$$

The MLE of (θ_1,θ_2) is defined as

$$(\hat{\theta}_{n1},\hat{\theta}_{n2}) = \arg \max_{(\theta_1,\theta_2)\in\mathcal{C}_4} l_n(\theta_1,\theta_2)$$

and the MLE of Δ is $\hat{\Delta}_n = \hat{\theta}_{n2} - \hat{\theta}_{n1}$. The following lemma finds the closed forms of the MLEs of θ_1, θ_2 and Δ. For the asymptotic purpose, we let $\omega = n_1/n$ and assume that $\omega \in (0,1)$ does not depend on n.

Lemma 6.5 *Suppose $X_{11}, \ldots, X_{1n_1}$ is a random sample from $f(x;\theta_1)$, and independently, $X_{21}, \ldots, X_{2n_2}$ is another random sample from $f(x;\theta_2)$, with $f(x;\theta)$ defined in (6.1). The MLEs of θ_1 and θ_2 subject to the constraint in $\mathcal{C}_4$ are*

$$\hat{\theta}_{n1} = \min\left\{\bar{X}_{n1}, \omega\bar{X}_{n1} + (1-\omega)\bar{X}_{n2}\right\}, \quad \hat{\theta}_{n2} = \max\left\{\bar{X}_{n2}, \omega\bar{X}_{n1} + (1-\omega)\bar{X}_{n2}\right\},$$

and the MLE of Δ is $\hat{\Delta}_n = (\bar{X}_{n2} - \bar{X}_{n1})^+$.

Lemma 6.5 generalizes Part (a) of Lemma 6.2 from normal distribution to general distributions in the NEF. We next generalize Part (b) of Lemma 6.2 to the class of distributions in the NEF by considering the limiting distributions of $\sqrt{n}(\hat{\theta}_{n1} - \theta_{10})$, $\sqrt{n}(\hat{\theta}_{n2} - \theta_{10})$, and $\sqrt{n}(\hat{\Delta}_n - \Delta_0)$, where $(\theta_{10}, \theta_{20})$ is the true value of (θ_1, θ_2) and $\Delta_0 = \theta_{20} - \theta_{10}$. In a similar way to the discussion in Section 6.3.1, fixing θ_{10} and θ_{20} may not be able to reveal the subtle results discovered in Proposition 6.2 and Figure 6.2. Instead, to gain a better understanding, we proceed by considering the following local asymptotic framework. Let

$$\theta_{10} = \theta_{10,n} = \eta_0 - (1-\omega)\Delta_{0,n}, \quad \text{and} \quad \theta_{20} = \theta_{20,n} = \eta_0 + \omega\Delta_{0,n},$$

where $\Delta_{0,n} = \delta n^{-1/2}$. Here η_0 is a fixed value, and δ is a fixed, nonnegative, local parameter not depending on n. Under this setup, we fix the true value of $\omega\theta_{10,n} + (1-\omega)\theta_{20,n}$, i.e. the overall mean of two samples, to be η_0, and allow the true value of constrained mean difference, $\Delta_{0,n} = \theta_{20,n} - \theta_{10,n}$, to vary in a $n^{-1/2}$ neighbourhood of 0. Note that this local setup is motivated from the observations in Proposition 6.2 and Figure 6.2: the non-regular behaviour of the coverage probabilities of the bootstrap percentile CIs for θ_1, θ_2, and Δ only occurs when Δ_0 is a $n^{-1/2}$ neighbourhood of 0.

Lemma 6.6 *Assume that $X_{11}, \ldots, X_{1n_1}$ is a random sample from $f(x;\theta_1)$, and independently, $X_{21}, \ldots, X_{2n_2}$ is another random sample from $f(x;\theta_2)$, with $f(x;\theta)$ defined in (6.1). Suppose the true value of (θ_1, θ_2) is $(\theta_{10,n}, \theta_{20,n}) \in \mathcal{C}_4$ with $\theta_{10,n} = \eta_0 - (1-\omega)\Delta_{0,n}$ and $\theta_{20,n} = \eta_0 + \omega\Delta_{0,n}$, where $\Delta_{0,n} = \delta n^{-1/2}$, η_0 is a fixed parameter, and δ is a fixed nonnegative local parameter not depending on n. Let $\sigma_0^2 = b''(\psi_0)$ with $\psi_0 = b'^{-1}(\eta_0)$. Then*

$$\begin{aligned}
\frac{n^{1/2}(\hat{\theta}_{n1} - \theta_{10,n})}{\sigma_0} &\xrightarrow{d} \min\left\{\sqrt{\frac{1}{\omega}}Z_1, \sqrt{\omega}Z_1 + \sqrt{1-\omega}Z_2 + \frac{(1-\omega)\delta}{\sigma_0}\right\}, \\
\frac{n^{1/2}(\hat{\theta}_{n2} - \theta_{20,n})}{\sigma_0} &\xrightarrow{d} \max\left\{\sqrt{\frac{1}{1-\omega}}Z_2, \sqrt{\omega}Z_1 + \sqrt{1-\omega}Z_2 - \frac{\omega\delta}{\sigma_0}\right\}, \\
\frac{n^{1/2}(\hat{\Delta}_n - \Delta_{0,n})}{\sigma_0} &\xrightarrow{d} \max\left\{\sqrt{\frac{1}{1-\omega}}Z_2 - \sqrt{\frac{1}{\omega}}Z_1, -\frac{\delta}{\sigma_0}\right\},
\end{aligned}$$

as $n \to \infty$, where Z_1, Z_2 are two independent standard normal random variables.

With the help of Lemma 6.6, we are able to quantify the local asymptotic coverage probabilities of the bootstrap percentile CIs for θ_1, θ_2, and Δ in the next theorem, which generalizes the results in Proposition 6.2.

For the convenience of presentation, we define some notation. Let $q^*_{1,\alpha}$, $q^*_{2,\alpha}$, and $q^*_{\Delta,\alpha}$ denote the α^{th} quantiles of the bootstrap distributions of $\hat{\theta}_{n1}$, $\hat{\theta}_{n2}$, and $\hat{\Delta}_n$, similar to those considered in Section 6.2.2. From Lemma 6.5, we note that $\hat{\theta}_{n1}$ and $\hat{\theta}_{n2}$ are both functions of random observations X_{ij}'s. Hence, the quantiles $q^*_{1,\alpha}$, $q^*_{2,\alpha}$, and $q^*_{\Delta,\alpha}$ are as defined random functions in terms of $(\hat{\theta}_{n1}, \hat{\theta}_{n2})$. Further let $F_{12}(x, y)$ denote the joint cumulative distribution function of

$$\min\left\{\sqrt{\frac{1}{\omega}}Z_1, \sqrt{\omega}Z_1 + \sqrt{1-\omega}Z_2 + \frac{(1-\omega)\delta}{\sigma_0}\right\},$$

and

$$\max\left\{\sqrt{\frac{1}{1-\omega}}Z_2, \sqrt{\omega}Z_1 + \sqrt{1-\omega}Z_2 - \frac{\omega\delta}{\sigma_0}\right\}.$$

That is, from Lemma 6.6, $F_{12}(x, y)$ is the joint limiting distribution of of $\sqrt{n}(\hat{\theta}_{n1} - \theta_{10,n})/\sigma_0$ and $\sqrt{n}(\hat{\theta}_{n2} - \theta_{20,n})/\sigma_0$.

Theorem 6.2 *Suppose we consider the same setup and assumptions as in Lemma 6.6. Then, for $\alpha_1, \alpha_2 \in (0, 0.5)$ with $\alpha = \alpha_1 + \alpha_2$, we have, as $n \to \infty$,*

(a)

$$\Pr\left(\Delta_{0,n} \in [q^*_{\Delta,\alpha_1}, q^*_{\Delta,1-\alpha_2}]\right) \;\to\; \begin{cases} 1-\alpha_1-\alpha_2, & \text{if } \delta > \Phi^{-1}(1-\alpha_2)\sigma_0/\sqrt{\omega(1-\omega)} \\ 1-\alpha_1, & \text{if } \delta < \Phi^{-1}(1-\alpha_2)\sigma_0/\sqrt{\omega(1-\omega)} \end{cases};$$

(b)

$$\Pr\left(\theta_{10,n} \in [q^*_{1,\alpha_1}, q^*_{1,1-\alpha_2}]\right) \;\to\; \iint I\{\alpha_1 \le g_1(x,y) \le 1-\alpha_2\} dF_{12}(x,y),$$

where

$$g_1(x,y) = \Phi\left\{-C_{11}(x)\right\} + \Phi\left\{-C_{12}(x,y)\right\} - \boldsymbol{\Phi}_{(\mathbf{0},\boldsymbol{\Lambda}_1)}\left\{-C_{11}(x), -C_{12}(x,y)\right\},$$

with $C_{11}(x) = \sqrt{\omega}x$, $C_{12}(x,y) = \omega x + (1-\omega)y + (1-\omega)\delta/\sigma_0$, *and*

$$\mathbf{\Lambda}_1 = \begin{pmatrix} 1 & \sqrt{\omega} \\ \sqrt{\omega} & 1 \end{pmatrix};$$

(c)

$$\Pr\left(\theta_{20,n} \in [q^*_{2,\alpha_1}, q^*_{2,1-\alpha_2}]\right) \quad \to \quad \iint I\{\alpha_1 \le g_2(x,y) \le 1-\alpha_2\} dF_{12}(x,y),$$

where

$$g_2(x,y) = \mathbf{\Phi}_{(\mathbf{0},\mathbf{\Lambda}_2)}\left\{-C_{21}(y), -C_{22}(x,y)\right\},$$

with $C_{21}(y) = \sqrt{1-\omega}y$, $C_{22}(x,y) = \omega x + (1-\omega)y - \omega\delta/\sigma_0$, *and*

$$\mathbf{\Lambda}_2 = \begin{pmatrix} 1 & \sqrt{1-\omega} \\ \sqrt{1-\omega} & 1 \end{pmatrix}.$$

This result generalizes the exact finite sample result in Proposition 6.2 for two-sample normal distributions to cover the general NEF of distributions. Finally, we apply the results in Theorem 6.2 to a binomial example, which may be considered as a two-sample special case of Example 1.6 given in Section 1.1.

Example 6.3 (Two-sample binomial distributions) *Suppose we have two independent samples* $X_{ij} \sim Bin(m_i, p_i)$ *for* $i = 1, 2$ *and* $j = i, \ldots, n_i$, *with known* m_i, *where* p_1, p_2 *are subject to the constraint* $p_2 \ge p_1$. *For illustration, we consider* $m_1 = m_2 = 1$. *Let* $\Delta = p_2 - p_1$. *Then the restriction becomes* $\Delta \ge 0$. *Further, we set* $\eta_0 = 0.5$ *and* $\omega = 0.25$.

Let the true values of p_1 *and* p_2 *be* $p_{10,n} = 0.5 - 0.75\Delta_{0,n}$, *and* $p_{20,n} = 0.5 + 0.25\Delta_{0,n}$, *with* $\Delta_{0,n} = \delta n^{-1/2}$, *and* δ *being a fixed nonnegative local parameter not depending on* n. *Under the current setup,* $\sigma_0^2 = \eta_0(1-\eta_0) = 0.5^2$. *In Figure 6.5, we graph the asymptotic and exact coverage probabilities of the bootstrap percentile CIs for* Δ, θ_1 *and* θ_2 *versus the true mean difference* Δ_0 *in the cases of* $(n_1, n_2) = (25, 75)$ *and* $(n_1, n_2) = (100, 300)$ *at level* $\alpha = 0.10$ *with* $\alpha_1 = \alpha_2 = 0.05$. *The asymptotic coverage probabilities are calculated by applying*

Theorem 6.2. The exact coverage probabilities are calculated by following the definition of the bootstrap percentile CIs as reviewed in Section 5.2 and discussed in Section 6.2.2. For comparison, we also include the exact coverage probabilities of the bootstrap percentile CIs of p_1, p_2, and Δ without using the constraint. We note that different choices of α_1 and α_2 can lead to different coverage behaviours of the confidence intervals, and hence give different graphs.

As we can see from Figure 6.5, the quantified local asymptotic coverage probabilities capture the general trend of the exact coverage probabilities. The two coverage probabilities become closer to each other as the sample size increases.

6.4 Proofs for Section 6.3

6.4.1 Proofs of Lemmas 6.3 and 6.4

Recall that the parameter space for θ is $\mathcal{C}_3 = \{\theta : \theta \geq d\}$, which is a closed convex set. Then by Proposition 2.4.3 in Silvapulle and Sen (2004, p. 51), $\hat{\theta}_n$ equivalently minimizes

$$(\bar{X}_n - \theta)^2$$

subject to the constraint $\theta \geq d$. That is

$$\hat{\theta}_n = \arg\min_{\theta \geq d} (\bar{X}_n - \theta)^2 = \max(\bar{X}_n, d).$$

This finishes the proof of Lemma 6.3. □

Next we come to the proof of Lemma 6.4. Recall that $\theta_{0,n} = d + n^{-1/2}\tau$ and $\sigma_0^2 = b''(\psi_0)$ with $\psi_0 = b'^{-1}(d)$. Applying the central limit theorem for a triangular array gives

$$\frac{\sum_{i=1}^{n}(X_i - \theta_{0,n})}{\sqrt{n\sigma_n^2}} \to Z,$$

in distribution as $n \to \infty$, where $\sigma_n^2 = b''(\psi_n)$ with $\psi_n = b'^{-1}(\theta_{0,n})$ and $Z \sim N(0,1)$. Further, both $b'^{-1}(x)$ and $b''(x)$ are continuous functions, and $\theta_{0,n} \to d$ as $n \to \infty$. Then

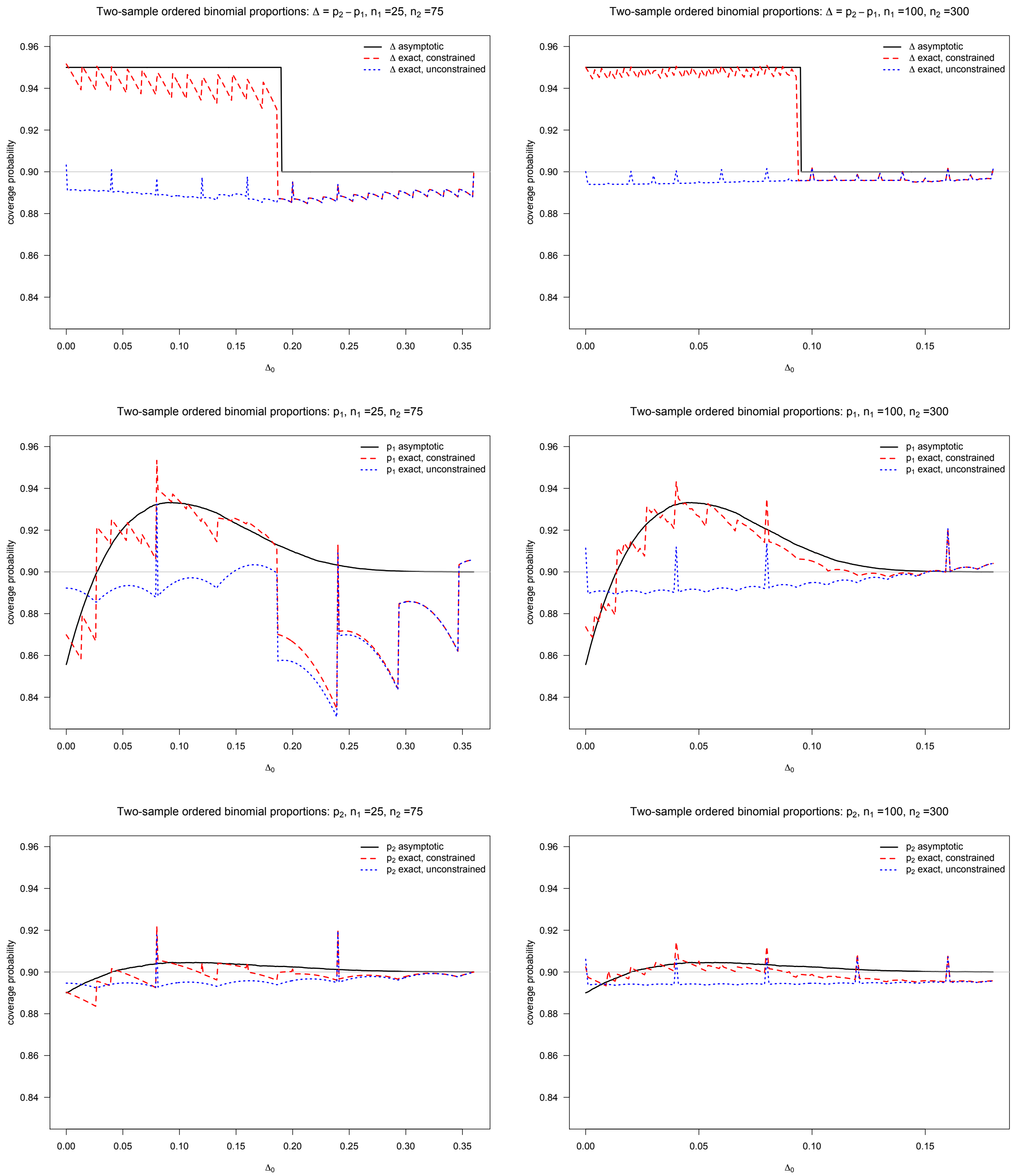

Figure 6.5: Asymptotic and exact coverage probabilities of 90% bootstrap percentile CIs for the proportions p_1 and p_2 and their difference Δ of two binomial distributions with $p_1 \leq p_2$, $\omega = 0.25$ and $\alpha_1 = \alpha_2 = 0.05$. Left panels are for $n = 100$, and right panels are for $n = 400$.

we have $\sigma_n^2 \to \sigma_0^2$ as $n \to \infty$. By Slutsky's theorem, we have

$$\frac{\sqrt{n}(\bar{X}_n - \theta_{0,n})}{\sigma_0} \to Z, \tag{6.2}$$

in distribution as $n \to \infty$. Together with the continuous mapping theorem, it follows that

$$\begin{aligned}
\frac{\sqrt{n}(\hat{\theta}_n - \theta_{0,n})}{\sigma_0} &= \frac{\sqrt{n}\{\max(\bar{X}_n, d) - \theta_{0,n}\}}{\sigma_0} \\
&= \max\left\{\frac{\sqrt{n}(\bar{X}_n - \theta_{0,n})}{\sigma_0}, \frac{\sqrt{n}(d - \theta_{0,n})}{\sigma_0}\right\} \\
&\to \max\left\{Z, -\frac{\tau}{\sigma_0}\right\},
\end{aligned}$$

in distribution as $n \to \infty$. This completes the proof of Lemma 6.4. □

6.4.2 Proof of Theorem 6.1

We first define some notation. Let $X_1^*, \ldots, X_n^*$ be the bootstrap sample from $f(x; \hat{\theta}_n)$ for the given $\hat{\theta}_n$, and $\hat{\theta}_n^* = \max(\bar{X}_n^*, d)$ be the MLE of θ based on the bootstrap sample, where $\bar{X}_n^* = \sum_{i=1}^n X_i^*/n$. Denote the bootstrap distributions of $\hat{\theta}_n$ and $\bar{X}_n$, respectively, by

$$G_n^*(x; \hat{\theta}_n) = \Pr\left(\hat{\theta}_n^* \le x | \hat{\theta}_n\right) \quad \text{and} \quad \bar{G}_n^*(x; \hat{\theta}_n) = \Pr\left(\bar{X}_n^* \le x | \hat{\theta}_n\right).$$

Then

$$G_n^*(x; \hat{\theta}_n) = \begin{cases} \bar{G}_n^*(x; \hat{\theta}_n) & x \ge d \\ 0 & x < d \end{cases}.$$

Recall that q_α^* is the α^{th} quantile of the bootstrap distribution of $\hat{\theta}_n$. Then

$$q_\alpha^* = G_n^{*-1}(\alpha; \hat{\theta}_n) = \max\{\bar{G}_n^{*-1}(\alpha; \hat{\theta}_n), d\}.$$

Therefore

$$\begin{aligned}
\Pr\left(\theta_{0,n} \in [q_{\alpha_1}^*, q_{1-\alpha_2}^*]\right) &= \Pr\left[\max\{\bar{G}_n^{*-1}(\alpha_1; \hat{\theta}_n), d\} \le \theta_{0,n} \le \max\{\bar{G}_n^{*-1}(1-\alpha_2; \hat{\theta}_n), d\}\right] \\
&= \Pr\left\{\bar{G}_n^{*-1}(\alpha_1; \hat{\theta}_n) \le \theta_{0,n} \le \bar{G}_n^{*-1}(1-\alpha_2; \hat{\theta}_n)\right\}.
\end{aligned}$$

Let

$$\bar{H}_n^*(x;\hat{\theta}_n) = \Pr\left\{\frac{\sqrt{n}(\bar{X}_n^* - \hat{\theta}_n)}{\sigma_0} \le x \Big| \hat{\theta}_n\right\},$$

which is the bootstrap distribution of the standardized $\bar{X}_n$. Then

$$\bar{G}_n^{*-1}(\alpha;\hat{\theta}_n) = n^{-1/2}\sigma_0 \bar{H}_n^{*-1}(\alpha;\hat{\theta}_n) + \hat{\theta}_n.$$

Therefore

$$\Pr\left(\theta_{0,n} \in [q_{\alpha_1}^*, q_{1-\alpha_2}^*]\right) = \Pr\left\{\bar{H}_n^{*-1}(\alpha_1;\hat{\theta}_n) \le \frac{\sqrt{n}(\theta_{0,n} - \hat{\theta}_n)}{\sigma_0} \le \bar{H}_n^{*-1}(1-\alpha_2;\hat{\theta}_n)\right\}. \quad (6.3)$$

We next study the asymptotic property of $\bar{H}_n^{*-1}(\alpha_1;\hat{\theta}_n)$, which is helpful for our proofs.

Lemma 6.7 *Assume the same setup and same assumptions as in Theorem 6.1. Then*

(a) $\hat{\theta}_n = d + o_p(1)$ *and* $\hat{\sigma}_n^2 = \sigma_0^2 + o_p(1)$, *where* $\hat{\sigma}_n^2 = b''(\hat{\psi}_n)$ *with* $\hat{\psi}_n = b'^{-1}(\hat{\theta}_n)$;

(b) $\sup_x |\bar{H}_n^*(x;\hat{\theta}_n) - \Phi(x)| = o_p(1)$;

(c) $\bar{H}_n^{*-1}(\alpha;\hat{\theta}_n) = \Phi^{-1}(\alpha) + o_p(1)$ *for any given level* $\alpha \in (0,1)$.

Proof. We first consider Part (a). Note that (6.2) implies that $\bar{X}_n - \theta_{0,n} = o_p(1)$. Recall that $\theta_{0,n} = d + n^{-1/2}\tau$. Then

$$\bar{X}_n = d + o_p(1).$$

This implies that

$$\hat{\theta}_n = \max(\bar{X}_n, d) = d + o_p(1).$$

Recall that both $b'^{-1}(x)$ and $b''(x)$ are continuous functions. By the continuous mapping theorem, we further have

$$\hat{\sigma}_n^2 = \sigma_0^2 + o_p(1).$$

This finishes the proof of Part (a).

Next we consider Part (b). We start with the limiting distribution of $\sqrt{n}(\bar{X}_n^* - \hat{\theta}_n)/\sigma_0$ for the given $\hat{\theta}_n$. Note that conditional on $\hat{\theta}_n$,

$$E(X_i^*|\hat{\theta}_n) = \hat{\theta}_n, \quad \text{and} \quad \text{Var}(X_i^*|\hat{\theta}_n) = \hat{\sigma}_n^2.$$

Let $\xrightarrow{d}$ denote "convergence in distribution". Then, by Berry-Esseén inequaltiy (Shao and Tu, 1995, Section 3.1, p. 74) or the central limit theorem (van der Vaart, 1998, Theorem 23.4), conditional on $\hat{\theta}_n$, we have

$$\frac{\sqrt{n}(\bar{X}_n^* - \hat{\theta}_n)}{\hat{\sigma}_n} \xrightarrow{d} N(0,1),$$

in probability. Recall that in Part (a), we have shown $\hat{\sigma}_n \to \sigma_0$ in probability. By conditional Slutsky's theorem (Cheng, 2015), we further have that, conditional on $\hat{\theta}_n$,

$$\frac{\sqrt{n}(\bar{X}_n^* - \hat{\theta}_n)}{\sigma_0} \xrightarrow{d} N(0,1),$$

in probability, which implies that

$$\sup_x |\bar{H}_n^*(x;\hat{\theta}_n) - \Phi(x)| = o_p(1).$$

This finishes the proof of Part (b).

With Part (b), then Part (c) is a direct application of Lemma 21.2 in van der Vaart (1998). □

We now move back to the proof of Theorem 6.1. Applying Lemma 6.7 to (6.3) gives

$$\begin{aligned}
\Pr\left(\theta_{0,n} \in [q^*_{\alpha_1}, q^*_{1-\alpha_2}]\right) &= \Pr\left\{-\Phi^{-1}(1-\alpha_2) + o_p(1) \le \frac{\sqrt{n}(\hat{\theta}_n - \theta_{0,n})}{\sigma_0} \le -\Phi^{-1}(\alpha_1) + o_p(1)\right\} \\
&= \Pr\left\{\Phi^{-1}(\alpha_2) + o_p(1) \le \frac{\sqrt{n}(\hat{\theta}_n - \theta_{0,n})}{\sigma_0} \le \Phi^{-1}(1-\alpha_1) + o_p(1)\right\} \\
&= \Pr\left\{\frac{\sqrt{n}(\hat{\theta}_n - \theta_{0,n})}{\sigma_0} \le \Phi^{-1}(1-\alpha_1) + o_p(1)\right\} \\
&\quad - \Pr\left\{\frac{\sqrt{n}(\hat{\theta}_n - \theta_{0,n})}{\sigma_0} < \Phi^{-1}(\alpha_2) + o_p(1)\right\}. \qquad (6.4)
\end{aligned}$$

Recall that in Lemma 6.4, we have shown that

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta_{0,n})}{\sigma_0} \to \max\left(Z, -\tau/\sigma_0\right).$$

That is, the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta_{0,n})/\sigma_0$ is $\Phi(x)I(x \geq -\tau/\sigma_0)$, which is continuous at $x = \Phi^{-1}(1-\alpha_1)$ and $x = \Phi^{-1}(\alpha_2)$ if $\Phi^{-1}(\alpha_2) \neq -\tau/\sigma_0$. By the definition of convergence in distribution, Slusky's theorem, and (6.4), we have that if $\Phi^{-1}(\alpha_2) \neq -\tau/\sigma_0$

$$\lim_{n\to\infty} \Pr\left(\theta_{0,n} \in [q^*_{\alpha_1}, q^*_{1-\alpha_2}]\right) = 1 - \alpha_1 - \alpha_2 I\left(\Phi^{-1}(\alpha_2) \geq -\tau/\sigma_0\right).$$

That is, for every continuous point of the limit function,

$$\begin{aligned}\lim_{n\to\infty} \Pr\left(\theta_{0,n} \in [q^*_{\alpha_1}, q^*_{1-\alpha_2}]\right) &= \begin{cases} 1-\alpha_1-\alpha_2 & \text{if } \Phi^{-1}(\alpha_2) > -\tau/\sigma_0 \\ 1-\alpha_1 & \text{if } \Phi^{-1}(\alpha_2) < -\tau/\sigma_0 \end{cases} \\ &= \begin{cases} 1-\alpha_1-\alpha_2 & \text{if } \tau/\sigma_0 > \Phi^{-1}(1-\alpha_2) \\ 1-\alpha_1 & \text{if } \tau/\sigma_0 < \Phi^{-1}(1-\alpha_2) \end{cases}.\end{aligned}$$

This finish the proof of Theorem 6.1.

6.4.3 Proofs of Lemmas 6.5 and 6.6

Note that the parameter space $\mathcal{C}_4 = \{(\theta_1, \theta_2) : \theta_2 \geq \theta_1\}$ is a closed convex set. Then by Proposition 2.4.3 in Silvapulle and Sen (2004, p. 51), $(\hat{\theta}_{n1}, \hat{\theta}_{n2})$ minimizes

$$n_1(\bar{X}_{n1} - \theta_1)^2 + n_2(\bar{X}_{n2} - \theta_2)^2.$$

That is,

$$(\hat{\theta}_{n1}, \hat{\theta}_{n2}) = \arg\min_{(\theta_1,\theta_2)\in\mathcal{C}_4} \left\{n_1(\bar{X}_{n1} - \theta_1)^2 + n_2(\bar{X}_{n2} - \theta_2)^2\right\}. \tag{6.5}$$

Following the proof of Lemma 6.2, we have

$$\hat{\theta}_{n1} = \min\left\{\bar{X}_{n1}, \omega\bar{X}_{n1} + (1-\omega)\bar{X}_{n2}\right\}, \quad \hat{\theta}_{n2} = \max\left\{\bar{X}_{n2}, \omega\bar{X}_{n1} + (1-\omega)\bar{X}_{n2}\right\},$$

and $\hat{\Delta}_n = (\bar{X}_{n2} - \bar{X}_{n1})^+$. This finishes the proof of Lemma 6.5. □

We now come to the proof of Lemma 6.6. For the convenience of presentation, we introduce some compact notation. Write $\boldsymbol{\theta}_{0,n} = (\theta_{10,n}, \theta_{20,n})^\tau$, $\bar{\mathbf{X}}_n = (\bar{X}_{n1}, \bar{X}_{n2})^\tau$ and $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{n1}, \hat{\theta}_{n2})^\tau$. Let

$$\mathbf{U}_n = (U_{n1}, U_{n2})^\tau = \frac{\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\theta}_{0,n})}{\sigma_0}.$$

Write $W_n = \omega U_{n1} + (1-\omega)U_{n2}$. Recall that $\theta_{10,n} = \eta_0 - (1-\omega)\Delta_{0,n}$ and $\theta_{20,n} = \eta_0 + \omega\Delta_{0,n}$ with $\Delta_{0,n} = \delta n^{-1/2}$, and $\sigma_0^2 = b''(\psi_0)$ with $\psi_0 = b'^{-1}(\eta_0)$.

Applying the central limit theorem for a triangular array, we have

$$\begin{pmatrix} \frac{\sum_{j=1}^{n_1}(X_{1j}-\theta_{10,n})}{\sqrt{n_1\sigma_{1,n}^2}} \\ \frac{\sum_{j=1}^{n_2}(X_{2j}-\theta_{20,n})}{\sqrt{n_2\sigma_{2,n}^2}} \end{pmatrix} \to \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$

in distribution as $n \to \infty$, where $\sigma_{i,n}^2 = b''(\psi_{i,n})$ with $\psi_{i,n} = b'^{-1}(\theta_{i0,n})$. Further, we have $\sigma_{i,n}^2 \to \sigma_0^2$ as $n \to \infty$, since $\theta_{i0,n} \to \eta_0$ as $n \to \infty$ and both $b'^{-1}(x)$ and $b''(x)$ are continuous functions. Therefore, by Slutsky's theorem, we have

$$\mathbf{U}_n \to \begin{pmatrix} \omega^{-1/2} Z_1 \\ (1-\omega)^{-1/2} Z_2 \end{pmatrix},$$

in distributions as $n \to \infty$.

We now come to the limiting distribution of $n^{1/2}(\hat{\theta}_{n1} - \theta_{10,n})/\sigma_0$. Using the form of $\hat{\theta}_{n1}$, we have

$$\begin{aligned}
\frac{n^{1/2}(\hat{\theta}_{n1} - \theta_{10,n})}{\sigma_0} &= n^{1/2}\left\{\frac{\min\left\{\bar{X}_{n1}, \omega\bar{X}_{n1} + (1-\omega)\bar{X}_{n2}\right\} - \theta_{10,n}}{\sigma_0}\right\} \\
&= \min\left\{\frac{n^{1/2}\left(\bar{X}_{n1} - \theta_{10,n}\right)}{\sigma_0}, \frac{n^{1/2}\left(\omega\bar{X}_{n1} + (1-\omega)\bar{X}_{n2} - \theta_{10,n}\right)}{\sigma_0}\right\} \\
&= \min\left\{U_{n1}, W_n + \frac{n^{1/2}\left(\eta_0 - \theta_{10,n}\right)}{\sigma_0}\right\} \\
&\to \min\left\{\sqrt{\frac{1}{\omega}}Z_1, \sqrt{\omega}Z_1 + \sqrt{1-\omega}Z_2 + \frac{(1-\omega)\delta}{\sigma_0}\right\}
\end{aligned}$$

in distribution as $n \to \infty$.

The proof for $n^{1/2}(\hat{\theta}_{n2} - \theta_{20,n})/\sigma_0$ is similar, and hence is omitted. For $\hat{\Delta}_n$, the proof is similar to that of Lemma 6.4. This finishes the proof of Lemma 6.6, and hence is also omitted. □

6.4.4 Proof of Theorem 6.2

The proof of Part (a) is similar to that of Theorem 6.1, and hence is omitted. Next we concentrate on the proof of Part (b) as the proof of Part (c) is just similar.

We first define some notation. Let $X_{11}^*, \ldots, X_{1n_1}^*$ be the bootstrap sample from $f(x; \hat{\theta}_{n1})$ for given $\hat{\theta}_{n1}$, and $X_{21}^*, \ldots, X_{2n_2}^*$ be the bootstrap sample from $f(x; \hat{\theta}_{n2})$ for given $\hat{\theta}_{n2}$. Further, let $(\hat{\theta}_{n1}^*, \hat{\theta}_{n2}^*)$ be the MLE of (θ_1, θ_2) based on X_{ij}^*'s. Then

$$\hat{\theta}_{n1}^* = \min\left\{\bar{X}_{n1}^*, \omega\bar{X}_{n1}^* + (1-\omega)\bar{X}_{n2}^*\right\}, \quad \text{and} \quad \hat{\theta}_{n2}^* = \max\left\{\bar{X}_{n2}^*, \omega\bar{X}_{n1}^* + (1-\omega)\bar{X}_{n2}^*\right\},$$

where $\bar{X}_{ni}^* = \sum_{j=1}^{n_i} X_{ij}^*/n_i$, $i = 1, 2$. Denote the bootstrap distribution of $\hat{\theta}_{n1}$ by

$$G_{n1}^*(x; \hat{\boldsymbol{\theta}}_n) = \Pr(\hat{\theta}_{n1}^* \leq x | \hat{\boldsymbol{\theta}}_n),$$

and the corresponding α^{th} quantile by $q_{1,\alpha}^*$.

Next, we mainly consider $\Pr\left(\theta_{10,n} \geq q_{\alpha_1}^*\right)$ in Part (b) of Theorem 6.2. The other part can be similarly proved. Note that

$$\Pr\left(\theta_{10,n} \geq q_{\alpha_1}^*\right) = \Pr\left\{\alpha_1 \leq G_{n1}^*(\theta_{10,n}; \hat{\boldsymbol{\theta}}_n)\right\}. \tag{6.6}$$

For $G_{n1}^*(\theta_{10,n}; \hat{\boldsymbol{\theta}}_n)$, we have

$$\begin{aligned} G_{n1}^*(\theta_{10,n}; \hat{\boldsymbol{\theta}}_n) &= \Pr\left[\min\left\{\bar{X}_{n1}^*, \omega\bar{X}_{n1}^* + (1-\omega)\bar{X}_{n2}^*\right\} \leq \theta_{10,n} | \hat{\boldsymbol{\theta}}_n\right] \\ &= \Pr\left(\bar{X}_{n1}^* \leq \theta_{10,n} | \hat{\boldsymbol{\theta}}_n\right) + \Pr\left(\omega\bar{X}_{n1}^* + (1-\omega)\bar{X}_{n2}^* \leq \theta_{10,n} | \hat{\boldsymbol{\theta}}_n\right) \\ &\quad -\Pr\left\{\bar{X}_{n1}^* \leq \theta_{10,n}, \omega\bar{X}_{n1}^* + (1-\omega)\bar{X}_{n2}^* \leq \theta_{10,n} | \hat{\boldsymbol{\theta}}_n\right\}. \end{aligned}$$

For $i = 1, 2$, let

$$U_{ni}^* = \frac{n^{1/2}\left(\bar{X}_{ni}^* - \hat{\theta}_{ni}\right)}{\sigma_0},$$

and $W_n^* = \omega U_{n1}^* + (1-\omega) U_{n2}^*$. Then after some algebra, we have

$$\begin{aligned} G_{n1}^*(\theta_{10,n}; \hat{\boldsymbol{\theta}}_n) &= \Pr\left(U_{n1}^* \le -U_{n1} | \hat{\boldsymbol{\theta}}_n\right) + \Pr\left\{W_n^* \le -W_n - (1-\omega)\delta/\sigma_0 | \hat{\boldsymbol{\theta}}_n\right\} \\ &\quad - \Pr\left\{U_{n1}^* \le -U_{n1}, W_n^* \le -W_n - (1-\omega)\delta/\sigma_0 | \hat{\boldsymbol{\theta}}_n\right\}. \end{aligned} \tag{6.7}$$

In next lemma, we study the asymptotic properties of

$$\Pr\left(U_{n1}^* \le x | \hat{\boldsymbol{\theta}}_n\right), \quad \Pr\left(W_n^* \le x | \hat{\boldsymbol{\theta}}_n\right), \quad \text{and} \quad \Pr\left(U_{n1}^* \le x, W_n^* \le y | \hat{\boldsymbol{\theta}}_n\right),$$

which are very helpful in our proofs.

Lemma 6.8 *Assume the same setup and same assumptions as in Theorem 6.2. Then*

$$\sup_x \left| \Pr\left(\sqrt{\omega} U_{n1}^* \le x | \hat{\boldsymbol{\theta}}_n\right) - \Phi(x) \right| = o_p(1), \tag{6.8}$$

$$\sup_x \left| \Pr\left(W_n^* \le x | \hat{\boldsymbol{\theta}}_n\right) - \Phi(x) \right| = o_p(1), \tag{6.9}$$

$$\sup_x \left| \Pr\left(\sqrt{\omega} U_{n1}^* \le x, W_n^* \le y | \hat{\boldsymbol{\theta}}_n\right) - \boldsymbol{\Phi}_{(\mathbf{0}, \boldsymbol{\Lambda}_1)}(x, y) \right| = o_p(1). \tag{6.10}$$

Proof. Similar to proof as Part (b) of Lemma 6.7, we can show that, conditional on $\hat{\boldsymbol{\theta}}_n$,

$$\sqrt{\omega} U_{n1}^* \xrightarrow{d} N(0,1), \quad \text{and} \quad \sqrt{1-\omega} U_{n2}^* \xrightarrow{d} N(0,1),$$

in probability. Further, conditional on $\hat{\boldsymbol{\theta}}_n$, $\sqrt{\omega} U_{n1}^*$ and $\sqrt{1-\omega} U_{n2}^*$ are independent. Hence, conditional on $\hat{\boldsymbol{\theta}}_n$, we have

$$\begin{pmatrix} \sqrt{\omega} U_{n1}^* \\ \sqrt{1-\omega} U_{n2}^* \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{2\times 2}),$$

in probability. Then by Example 3.3 of Shao and Tu (1995), we have, conditional on $\hat{\boldsymbol{\theta}}_n$, that

$$\begin{pmatrix} \sqrt{\omega} U_{n1}^* \\ W_n^* \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Lambda}_1),$$

in probability. This implies (6.8)–(6.10) and finishes the proof. ☐

Combining Lemma 6.8 and $G^*_{n1}(\theta_{10,n};\hat{\boldsymbol{\theta}}_n)$ in (6.7), we obtain

$$\begin{aligned} G^*_{n1}(\theta_{10,n};\hat{\boldsymbol{\theta}}_n) &= \Phi(-\sqrt{\omega}U_{n1}) + \Phi\left(-W_n - \frac{(1-\omega)\delta}{\sigma_0}\right) \\ &\quad -\boldsymbol{\Phi}_{(\mathbf{0},\boldsymbol{\Lambda}_1)}\left(-\sqrt{\omega}U_{n1}, -W_n - \frac{(1-\omega)\delta}{\sigma_0}\right) + o_p(1) \\ &= \Phi\{-C_{11}(U_{n1})\} + \Phi\{-C_{12}(U_{n1},U_{n2})\} \\ &\quad -\boldsymbol{\Phi}_{(\mathbf{0},\boldsymbol{\Lambda}_1)}\{-C_{11}(U_{n1}), -C_{12}(U_{n1},U_{n2})\} + o_p(1) \\ &= g_1(U_{n1},U_{n2}) + o_p(1), \qquad (6.11) \end{aligned}$$

where $g_1(x,y)$ is defined in Part (b) of Theorem 6.2.

Combining (6.6) and (6.11) and applying Slutsky's theorem, we further have

$$\begin{aligned} \lim_{n\to\infty} \Pr\left(\theta_{10,n} \geq q^*_{\alpha_1}\right) &= \lim_{n\to\infty} \Pr\{\alpha_1 \leq g_1(U_{n1},U_{n2}) + o_p(1)\} \\ &= \iint I\{\alpha_1 \leq g_1(x,y)\} dF_{12}(x,y). \qquad (6.12) \end{aligned}$$

Similarly, we find

$$\lim_{n\to\infty} \Pr\left(\theta_{10,n} \leq q^*_{1-\alpha_2}\right) = \iint I\{g_1(x,y) \leq 1-\alpha_2\} dF_{12}(x,y). \qquad (6.13)$$

Combining (6.12) and (6.13), we get

$$\lim_{n\to\infty} \Pr\left(q^*_{\alpha_1} \leq \theta_{10,n} \leq q^*_{1-\alpha_2}\right) = \iint I\{\alpha_1 \leq g_1(x,y) \leq 1-\alpha_2\} dF_{12}(x,y).$$

This finishes the proof. □

Chapter 7

Summary, discussion, and future work

7.1 Summary and discussion of the current achievements

In this thesis, we have considered, the empirical likelihood and bootstrap method, respectively, in Part I and Part II, for some constrained inference problems. Here, we first summarize what has been achieved in this thesis.

In Chapter 3, we dealt with the problem of testing homogeneity for multiple distributions with excess zero observations. By assuming the semiparametric DRM for the distributions of the positive data, we have developed an ELR test which can efficiently exploit information from the pooled data and is robust against the risk of misspecification of the underlying data distributions. Furthermore, the proposed ELR test guarantees that the asymptotic size of the test can always be controlled at the nominal level under the null hypothesis. The chi-squared limiting distribution of this ELR statistic has been derived under the homogeneous null hypothesis. We have further suggested using a bootstrap procedure to calibrate the finite sample distribution of the ELR. The consistency of the bootstrap ELR has been established under both the null and alternative hypotheses. Coupled with the bootstrap, simulation studies and a real example showed that, for practical sample sizes, the ELR test has accurate type I error, is competitive to, and sometimes more powerful than, other existing tests. Finally, the proposed ELR test can be readily implemented in practice by using logistic regression routines available in standard statistical packages.

In Chapter 4, we discussed the problem of making statistical inferences on the means of multiple distributions with excess zero observations. We followed the same modelling framework as in Chapter 3 by using the semiparametric DRM to link multiple distributions with excess zeros. Based on this semiparametric framework, we further proposed an ELR statistic for making inference on the means, which allows us to make efficient use of the entire sample information. The limiting chi-squared distribution of this ELR statistic has been established under a fairly general linear null hypothesis on the means. This result allows us to construct a test for mean equality, and confidence intervals for the mean differences and ratios, as important applications. Simulation results showed that the performance of the proposed ELR test under correctly specified basis functions in the DRM is, in general, less sensitive to unequal sample sizes, in terms of both type I error and power of the test, when compared with other popular tests for mean equality. With extensive simulation studies and a real data example, we identified a scenario in which the proposed ELR has superior performance in terms of type I error and power, and is computationally stable for testing overall mean equality when the correctly specified basis function is the logarithm function in the DRM.

As an important area of application, the ELR based semiparametric inference framework developed in Part I can also be employed to deal with zero-inflated count data. For these discrete problems we no longer have a clear distinction between the (discrete) zero counts and the (continuous) positive data in the non-standard mixture model. Hence, we adapt the idea of the hurdle model for zero-inflated count data (Min and Agresti, 2002; Bedrick and Hossain, 2013) which models the zero and positive counts separately. That is, in models (3.1) and (4.1), $G_i(x)$ is the cumulative distribution function for the positive counts. We then link G_i's through the DRM (2.1), in which $dG_i(x)$ should be understood as the probability mass function. Note that the commonly used zero-truncated Poisson and zero-truncated negative binomial distributions both satisfy the DRM condition (2.1).

As discussed in Bedrick and Hossain (2013), testing homogeneity under the two types of mixture structures of the zero-inflated Poisson and the Poisson-hurdle model are equivalent. Furthermore, it can be verified that the definition of the means under the mixture structures of the zero-inflated Poisson and the Poisson-hurdle model are also equivalent. Similar conclusions also apply to the negative binomial distribution. Therefore, for multiple groups of data containing discrete counts with zero inflation, the ELR for testing homogeneity proposed in Chapter 3, and the ELR for making inferences on the means proposed in Chapter 4, can both be directly applied.

The ELR based inference framework on the means, proposed in Chapter 4, can also be extended to more general settings. For example, given that we have prior knowledge that all, or part of, the group means are equal (Gupta and Li, 2006; Tsao and Wu, 2006; Fu et al., 2009), then the proposed framework can be used to obtain refined inference results. Further, if the number of estimating equations exceeds the number of parameters, our proposed framework can be extended to incorporate such auxiliary information to further improve the inference results (Qin and Lawless, 1994; Qin et al., 2015).

In Chapter 6, we studied the behaviour of the standard bootstrap percentile method for constructing confidence intervals when the parameters are subject to inequality constraints. We concentrated on the important cases of one- and two sample mean problems with data generated from the natural exponential family of distributions. We have quantified the local asymptotic coverage probabilities of the bootstrap percentile confidence intervals when the true constrained parameter is varying in a local neighbourhood of the boundary. Under such a setup, the asymptotic results not only provide examples that the bootstrap percentile method is not universally appropriate, but also give a quantification of its coverage behaviour in a practically meaningful way. It has also been shown, by using binomial and Poisson examples, that, the important cases that we have investigated can find rich applications in constrained inference problems.

7.2 Future work

The proposed inference methods and frameworks in this thesis are expected to be very promising for a number of research problems. In this section, we highlight some possible future research.

7.2.1 Testing homogeneity for multiple groups of zero-and-one inflated proportion data

The methodologies developed in Part I of this thesis is feasible to be generalized to data with more than one degenerate component. For example, in modelling proportion data it is common to have a zero-and-one inflated mixture structure. Multiple groups of such samples have important applications, such as in transportation safety (Ospina and Ferrari, 2012), and in marine science (Sun and Gitelman, 2016). Specifically, we consider $m+1$ independent groups of samples as follows:

$$x_{i1}, \ldots, x_{in_i} \sim F_i(x) = \nu_i I(x=0) + \omega_i I(x=1) + (1-\nu_i-\omega_i) I(0<x<1) G_i(x),$$

for $i = 0, \ldots, m$, where n_i is the ith group's sample size and the $G_i(\cdot)$'s are cumulative distribution functions with common continuous support on open interval $(0,1)$.

In the literature, research has focused on the use of Beta distributions for the G_i's. To allow for more model flexibility, and to make efficient use of all observations, the semiparametric framework under the DRM can be introduced in this context. Particularly, the Beta distribution is a special example satisfying the DRM condition with $\mathbf{q}(x) = \{\log(x), \log(1-x)\}^\top$. The scientific questions for this type of proportion data include a test for homogeneity (Sun and Gitelman, 2016). Hence, an extension of the empirical likelihood ratio test proposed in Chapter 3 would be useful and deserves further investigation.

7.2.2 Semiparametric estimation and comparison for multiple Gini indices

The empirical likelihood method plays an active role for making inference on Gini index (Qin et al., 2010; Peng, 2011). More recently, Wang and Zhao (2016) studied the inference on the difference of two Gini indices using the empirical likelihood.

Suppose we have $m+1$ populations. Let F_i be the income distribution associated with the ith population (e.g. country). For two independent random variables X_i and Y_i which share the same distribution F_i, the definition of the Gini index for the ith population is

$$\text{Gini}_i = \frac{E|X_i - Y_i|}{2\mu_i} = \frac{1}{\mu_i}\int_0^\infty \{2F_i(x) - 1\} x dF_i(x),$$

where $\mu_i = \int_0^\infty x dF_i(x)$ for $i = 0, \ldots, m$. Hence, if we have good estimates of $F_i(x)$'s, we can estimate the Gini indices and based on which inference procedures can be developed.

It may be possible to further introduce the semiparametric DRM to link the income distributions across multiple populations. In the modelling of income distributions, Davidson (2009) considered three parametric distributions for F_i in simulation studies, which are argued to be realistic in practice. Specifically, Davidson (2009) considered data generated from the exponential distribution, the Pareto distribution, and the log-normal distribution. The exponential and the log-normal distributions are clearly in the DRM family; see Table 2.1. Furthermore, the Pareto distribution with shape parameter λ_i and fixed scale parameter belongs to the DRM family with $\mathbf{q}(x) = \log(x)$. With this information, we see the possibility to make better estimation and inference on multiple Gini indices under the semiparametric setup discussed in Part I of this thesis.

7.2.3 Quantifying the coverage probabilities of bootstrap likelihood ratio confidence intervals

For the constrained inference problems consider in Part II of this thesis, the likelihood ratio is another popular way to construct confidence interval for a scalar parameter (or joint confidence region for higher dimensional parameters). In practice, the confidence interval based on bootstrapping the likelihood ratio is desirable since it does not require calculating the possibly complicated limiting distribution, and it is known for its finite sample improvement. Hence, a natural question to ask is: can we also appropriately quantify the coverage probability of bootstrap likelihood ratio function based confidence interval, in light of the framework proposed in Chapter 6?

It is still convenient to look at a simple normal example with mean $\theta \in \{\theta : \theta \geq 0\}$ and variance one. It can be checked that the likelihood ratio is no longer a pivotal quantity, and actually its distribution would depend on $\sqrt{n}\theta_0$ where θ_0 is the true value of θ. Therefore, we conjecture that the bootstrap likelihood ratio may also have some non-regular behaviour. By numerical studies, we find that the coverage behaviour of the bootstrap likelihood ratio based confidence interval for θ seems much more complicated than the bootstrap percentile confidence interval, even in this one-sample case. Hence, it becomes an interesting topic to be explored in the future.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Csaki, F., editors, *Second international symposium on information theory*, pages 267–281. Academiai Kiado, Budapest.

Anaya-Izquierdo, K., Critchley, F., and Marriott, P. (2014). When are first-order asymptotics adequate? a diagnostic. *Stat*, 3(1):17–22.

Anderson, J. A. (1979). Multivariate logistic compounds. *Biometrika*, 66(1):17–26.

Andrews, D. W. K. (1997). A simple counterexample to the bootstrap. Unpublished manuscript, Cowles Foundation for Research in Economics, Yale University.

Andrews, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68(2):399–405.

Aubin, J.-B. and Leoni-Aubin, S. (2008). Projection density estimation under a m-sample semi-parametric model. *Computational Statistics and Data Analysis*, 52(5):2451 – 2468.

Bascoul-Mollevi, C., Gourgou-Bourgade, S., and Kramar, A. (2005). Two-part statistics with paired data. *Statistics in Medicine*, 24(9):1435–1448.

Bedrick, E. J. and Hossain, A. (2013). Conditional tests for homogeneity of zero-inflated Poisson and Poisson-hurdle distributions. *Computational Statistics and Data Analysis*, 61:99–106.

Brunner, E., Dette, H., and Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association*, 92(440):1494–1502.

Cai, S. (2014). *On dual empirical likelihood inference under semiparametric density ratio models in the presence of multiple samples with applications to long term monitoring of lumber quality.* PhD thesis, University of British Columbia.

Cai, S. (2015). *drmdel: Dual Empirical Likelihood Inference under Density Ratio Models in the Presence of Multiple Samples.* R package version 1.3.1.

Cai, S., Chen, J., and Zidek, J. V. (2016). Hypothesis test in the presence of multiple samples under density ratio models. *Statistica Sinica, to appear.*

Chan, K. C. G. (2013). Survival analysis without survival data: connecting length-biased and case-control data. *Biometrika*, 100(3):764.

Chen, B., Li, P., Qin, J., and Yu, T. (2016). Using a monotonic density ratio model to find the asymptotically optimal combination of multiple diagnostic tests. *Journal of the American Statistical Association*, 111(514):861–874.

Chen, J., Chen, S.-Y., and Rao, J. N. K. (2003). Empirical likelihood confidence intervals for the mean of a population containing many zero values. *The Canadian Journal of Statistics*, 31(1):53–68.

Chen, J. and Liu, Y. (2013). Quantile and quantile-function estimations under density ratio model. *The Annals of Statistics*, 41(3):1669–1692.

Chen, S. X. and Qin, J. (2003). Empirical likelihood-based confidence intervals for data with possible zero observations. *Statistics and Probability Letters*, 65(1):29 – 37.

Chen, Y., Ning, Y., Hong, C., and Wang, S. (2014). Semiparametric tests for identifying differentially methylated loci with casecontrol designs using illumina arrays. *Genetic Epidemiology*, 38(1):42–50.

Chen, Y.-H. and Zhou, X.-H. (2006). Generalized confidence intervals for the ratio or difference of two means for lognormal populations with zeros. Working Paper 296, UW Biostatistics Working Paper Series. Unpublished manuscript.

Cheng, G. (2015). Moment consistency of the exchangeably weighted bootstrap for semiparametric m-estimation. *Scandinavian Journal of Statistics*, 42(3):665–684.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 25(3):573–578.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

Critchley, F., Marriott, P., and Salmon, M. (1996). On the differential geometry of the Wald test with nonlinear restrictions. *Econometrica*, 64(5):1213–1222.

Davidov, O., Fokianos, K., and Iliopoulos, G. (2010). Order-restricted semiparametric inference for the power bias model. *Biometrics*, 66(2):549–557.

Davidov, O., Fokianos, K., and Iliopoulos, G. (2014). Semiparametric inference for the two-way layout under order restrictions. *Scandinavian Journal of Statistics*, 41(3):622–638.

Davidson, R. (2009). Reliable inference for the Gini index. *Journal of Econometrics*, 150(1):30–40.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application.* Cambridge university press, New York.

de Carvalho, M. and Davison, A. C. (2014). Spectral density ratio models for multivariate extremes. *Journal of the American Statistical Association*, 109(506):764–776.

Delucchi, K. L. and Bostrom, A. (2004). Methods for analysis of skewed data distributions in psychiatric clinical studies: working with many zero values. *American Journal of Psychiatry*, 161(7):1159–1168.

Diao, G., Ning, J., and Qin, J. (2012). Maximum likelihood estimation for semiparametric density ratio model. *The International Journal of Biostatistics*, 8(1):370–384.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3):189–228.

Drton, M. and Williams, B. (2011). Quantifying the failure of bootstrap likelihood ratio tests. *Biometrika*, 98(4):919–934.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap.* Chapman and Hall, New York.

Feldman, G. J. and Cousins, R. D. (1998). Unified approach to the classical statistical analysis of small signals. *Physical Review D*, 57:3873–3889.

Fokianos, K. (2004). Merging information for semiparametric density estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(4):941–958.

Fokianos, K. (2007). Density ratio model selection. *Journal of Statistical Computation and Simulation*, 77(9):805–819.

Fokianos, K. and Kaimi, I. (2006). On the effect of misspecifying the density ratio model. *Annals of the Institute of Statistical Mathematics*, 58(3):475–497.

Fokianos, K., Kedem, B., Qin, J., and Short, D. A. (2001). A semiparametric approach to the one-way layout. *Technometrics*, 43(1):56–65.

Fokianos, K. and Savvides, A. (2008). On comparing several spectral densities. *Technometrics*, 50(3):317–331.

Follmann, D. (1996). A simple multivariate test for one-sided alternatives. *Journal of the American Statistical Association*, 91(434):854–861.

Follmann, D., Fay, M. P., and Proschan, M. (2009). Chop-lump tests for vaccine trials. *Biometrics*, 65(3):885–893.

Friedrich, S., Konietschke, F., and Pauly, M. (2016). *GFD: Tests for General Factorial Designs.* R package version 0.1.4.

Fu, Y., Chen, J., and Kalbfleisch, J. D. (2006). Testing for homogeneity in genetic linkage analysis. *Statistica Sinica*, 16(3):805–823.

Fu, Y., Wang, X., and Wu, C. (2009). Weighted empirical likelihood inference for multiple samples. *Journal of Statistical Planning and Inference*, 139(4):1462–1473.

Gilbert, P. B. (2000). Large sample theory of maximum likelihood estimates in semiparametric biased sampling models. *The Annals of Statistics*, 28(1):151–194.

Gleiss, A., Dakna, M., Mischak, H., and Heinze, G. (2015). Two-group comparisons of zero-inflated intensity values: the choice of test statistic matters. *Bioinformatics*, 31(14):2310–2317.

Guan, Z. (2004). A semiparametric changepoint model. *Biometrika*, 91(4):849–862.

Gupta, R. C. and Li, X. (2006). Statistical inference for the common mean of two log-normal distributions and some applications in reliability. *Computational Statistics and Data Analysis*, 50(11):3141–3164.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *The Annals of Statistics*, 16(3):927–953.

Hallstrom, A. P. (2010). A modified Wilcoxon test for non-negative distributions with a clump of zeros. *Statistics in Medicine*, 29(3):391–400.

Hu, Z. and Proschan, M. (2015). Two-part test of vaccine effect. *Statistics in Medicine*, 34(11):1904–1911.

Hu, Z., Qin, J., and Follmann, D. (2008). Semiparametric two-sample changepoint model with application to human immunodeficiency virus studies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57(5):589–607.

Huang, A. and Rathouz, P. J. (2012). Proportional likelihood ratio models for mean regression. *Biometrika*, 99(1):223–229.

Hwang, J. T. G. and Peddada, S. D. (1994). Confidence interval estimation subject to order restrictions. *The Annals of Statistics*, 22(1):67–93.

Janssen, A. and Pauls, T. (2003). How do bootstrap and permutation tests work? *The Annals of Statistics*, 31(3):768–806.

Jiang, S., Chen, B., and Tu, D. (2016). Inference on treatment-covariate interaction based on a nonparametric measure of treatment effects and censored survival data. *Statistics in Medicine*, 35(16):2715–2725.

Jiang, S. and Tu, D. (2012). Inference on the probability as a measurement of treatment effect under a density ratio model and random censoring. *Computational Statistics and Data Analysis*, 56(5):1069–1078.

Johnson, W. D., Burton, J. H., Beyl, R. A., and Romer, J. E. (2015). A simple chi-square statistic for testing homogeneity of zero-inflated distributions. *Open Journal of Statistics*, 5(6):483–493.

Kang, L., Vexler, A., Tian, L., Cooney, M., and Louis, G. M. B. (2010). Empirical and parametric likelihood interval estimation for populations with many zero values: application for assessing environmental chemical concentrations and reproductive health. *Epidemiology*, 21(4):S58–S63.

Kedem, B., Lu, G., Wei, R., and Williams, P. D. (2008). Forecasting mortality rates via density ratio modeling. *The Canadian Journal of Statistics*, 36(2):193–206.

Keziou, A. and Leoni-Aubin, S. (2008). On empirical likelihood for semiparametric two-sample density ratio models. *Journal of Statistical Planning and Inference*, 138(4):915–928.

Koopmans, L. H. (1981). *Introduction to contemporary statistical methods*. Duxbury Press, Boston.

Kosorok, M. (2008). *Introduction to empirical processes and semiparametric inference*. Springer, New York.

Kruskal, W. H. (1952). A nonparametric test for the several sample problem. *The Annals of Mathematical Statistics*, 23(4):525–540.

Lachenbruch, P. A. (1976). Analysis of data with clumping at zero. *Biometrische Zeitschrift*, 18(5):351–356.

Lachenbruch, P. A. (2001). Comparisons of two-part models with competitors. *Statistics in Medicine*, 20(8):1215–1234.

Lachenbruch, P. A. (2002). Analysis of data with excess zeros. *Statistical Methods in Medical Research*, 11(4):297–302.

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

Li, P., Liu, Y., and Qin, J. (2016). Semiparametric inference in a genetic mixture model. *Journal of the American Statistical Association, to appear.*

Li, Z., Taylor, J. M. G., and Nan, B. (2010). Construction of confidence intervals and regions for ordered binomial probabilities. *The American Statistician*, 64(4):291–298.

Liu, X., Jiang, H., and Zhou, Y. (2014). Local empirical likelihood inference for varying-coefficient density-ratio models based on case-control data. *Journal of the American Statistical Association*, 109(506):635–646.

Luo, X. and Tsai, W. Y. (2012). A proportional likelihood ratio model. *Biometrika*, 99(1):211–222.

Marazzi, A., Paccaud, F., Ruffieux, C., and Beguin, C. (1998). Fitting the distributions of length of stay by parametric models. *Medical care*, 36(6):915–927.

Min, Y. and Agresti, A. (2002). Modeling nonnegative data with clumping at zero: A survey. *Journal of Iranian Statistical Society*, 1:7–33.

Molenberghs, G. and Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician*, 61(1):22–27.

Moulton, L. H. and Halsey, N. A. (1995). A mixture model with detection limits for regression analyses of antibody response to vaccine. *Biometrics*, 51(4):1570–1578.

Muralidharan, K. and Kale, B. K. (2002). Modified gamma distribution with singularity at zero. *Communications in Statistics - Simulation and Computation*, 31(1):143–158.

Ning, Y. and Chen, Y. (2015). A class of pseudolikelihood ratio tests for homogeneity in exponential tilt mixture models. *Scandinavian Journal of Statistics*, 42(2):504–517.

Nixon, R. M. and Thompson, S. G. (2004). Parametric modelling of cost data in medical studies. *Statistics in Medicine*, 23(8):1311–1331.

Ospina, R. and Ferrari, S. L. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics and Data Analysis*, 56(6):1609–1623.

Ott, J. (1999). *Analysis of human genetic linkage*. The Johns Hopkins University Press, Baltimore.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120.

Owen, A. B. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, 19(4):1725–1747.

Owen, A. B. (2001). *Empirical likelihood*. Chapman and Hall/CRC Press, New York.

Pauly, M., Brunner, E., and Konietschke, F. (2015). Asymptotic permutation tests in general factorial designs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):461–473.

Peddada, S. D. (1997). Confidence interval estimation of population means subject to order restrictions using resampling procedures. *Statistics and Probability Letters*, 31(4):255 – 265.

Peng, L. (2011). Empirical likelihood methods for the Gini index. *Australian and New Zealand Journal of Statistics*, 53(2):131–139.

Pennington, M. (1983). Efficient estimators of abundance, for fish and plankton surveys. *Biometrics*, 39(1):281–286.

Prentice, R. L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, 66(3):403–411.

Qin, J. (1993). Empirical likelihood in biased sample problems. *The Annals of Statistics*, 21(3):1182–1196.

Qin, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika*, 85(3):619–630.

Qin, J. (1999). Empirical likelihood ratio based confidence intervals for mixture proportions. *The Annals of Statistics*, 27(4):1368–1384.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22(1):300–325.

Qin, J. and Zhang, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, 84(3):609–618.

Qin, J. and Zhang, B. (2003). Using logistic regression procedures for estimating receiver operating characteristic curves. *Biometrika*, 90(3):585–596.

Qin, J., Zhang, H., Li, P., Albanes, D., and Yu, K. (2015). Using covariate-specific disease prevalence information to increase the power of case-control studies. *Biometrika*, 102(1):169–180.

Qin, Y., Rao, J., and Wu, C. (2010). Empirical likelihood confidence intervals for the Gini measure of income inequality. *Economic Modelling*, 27(6):1429–1435.

R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Robertson, T., Wrigth, F. T., and Dykstra, R. L. (1988). *Order restricted statistical inference*. John Wiley & Sons, New York.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons, Inc., New York.

Shao, J. and Tu, D. (1995). *The jackknife and bootstrap*. Springer, Verlag.

Shen, Y., Ning, J., and Qin, J. (2012). Likelihood approaches for the invariant density ratio model with biased-sampling data. *Biometrika*, 99(2):363–378.

Silvapulle, M. J. and Sen, P. K. (2004). *Constrained statistical inference: inequality, order, and shape restrictions*, volume 912. John Wiley & Sons, New York.

Silvapulle, M. J. and Silvapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association*, 90(429):342–349.

Sun, L. and Gitelman, A. I. (2016). Simultaneous tests for homogeneity of two zero-inflated (Beta) populations. *Communications in Statistics - Theory and Methods*, 0(0):to appear.

Susko, E. (2013). Likelihood ratio tests with boundary constraints using data-dependent degrees of freedom. *Biometrika*, 100(4):1019–1023.

Taylor, D. J., Kupper, L. L., Rappaport, S. M., and Lyles, R. H. (2001). A mixture model for occupational exposure mean testing with a limit of detection. *Biometrics*, 57(3):681–688.

Taylor, S. and Pollard, K. (2009). Hypothesis tests for point-mass mixture data with application to 'omics data with many zero values. *Statistical Applications in Genetics and Molecular Biology*, 8(1):1–43.

Tsao, M. and Wu, C. (2006). Empirical likelihood inference for a common mean in the presence of heteroscedasticity. *The Canadian Journal of Statistics*, 34(1):45–59.

Tse, S. K., Chow, S. C., Lu, Q., and Cosmatos, D. (2009). Testing homogeneity of two zero-inflated Poisson populations. *Biometrical Journal*, 51(1):159–170.

Tu, W. and Zhou, X.-H. (1999). A Wald test comparing medical costs based on log-normal distributions with zero valued costs. *Statistics in Medicine*, 18(20):2749–2761.

Vallejo, G., Fernández, M. P., and Livacic-Rojas, P. E. (2010). Analysis of unbalanced factorial designs with heteroscedastic data. *Journal of Statistical Computation and Simulation*, 80(1):75–88.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press, New York.

Wagner, B. D., Robertson, C. E., and Harris, J. K. (2011). Application of two-part statistics for comparison of sequence variant counts. *PLoS ONE*, 6(5):e20296.

Wan, S. and Zhang, B. (2007). Smooth semiparametric receiver operating characteristic curves for continuous diagnostic tests. *Statistics in Medicine*, 26(12):2565–2586.

Wan, S. and Zhang, B. (2013). Semiparametric ROC surface estimation for continuous diagnostic tests via polytomous logistic regression procedures. *Journal of Statistical Computation and Simulation*, 83(12):2195–2205.

Wang, D. and Zhao, Y. (2016). Jackknife empirical likelihood for comparing two Gini indices. *The Canadian Journal of Statistics*, 44(1):102–119.

Wang, S. (2014). *Semiparametric inferences under density ratio models*. PhD thesis, University of Toledo.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29(3-4):350–362.

Woodroofe, M. and Wang, H. (2000). The problem of low counts in a signal plus noise model. *The Annals of Statistics*, 28(6):1561–1569.

Wu, C. and Yan, Y. (2012). Empirical likelihood inference for two-sample problems. *Statistics and Its Interface*, 5(3):345–354.

Zhang, B. (2002). Assessing goodness-of-fit of generalized logit models based on case-control data. *Journal of Multivariate Analysis*, 82(1):17–38.

Zhou, X.-H. and Tu, W. (1999). Comparison of several independent population means when their samples contain log-normal and possibly zero observations. *Biometrics*, 55(2):645–651.

Zhou, X.-H. and Tu, W. (2000). Interval estimation for the ratio in means of log-normally distributed medical costs with zero values. *Computational Statistics and Data Analysis*, 35(2):201 – 210.

Zou, F., Fine, J. P., and Yandell, B. S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika*, 89(1):61–75.

Appendix A

Appendix: R functions

A.1 R functions for Chapter 3

In this section, we provide `R` functions to implement the bootstrap ELR test proposed in Chapter 3:

- `elr.part0`: calculate $R_{n,1}$;
- `elr.part1`: calculate $R_{n,2}$;
- `elr`: compute R_n.
  - It is integrated with five commonly used basis functions, as describe at the beginning of Section 3.6 in Chapter 3.
- `boot.elr`: the main function.
  - Inputs: "`B`" is the number of bootstrapped samples; "`data`" input must be a data frame: 1st column is the group labels and 2nd column contains all observations (both zeros and nonzeros, and the nonzeros could be discrete or continuous).
  - Outputs: this function returns the values of the proposed ELR test statistics and corresponding p-values calibrated using the nonparametric bootstrap procedure.

As an illustration, we apply the bootstrap ELR with $B = 999$ to a data set generated from the log-normal mixture with three groups in which $(\nu_0, \nu_1, \nu_2) = (0.2, 0.3, 0.4)$, $(a_0, a_1, a_2) = (0, 0.3, 0.5)$, and $(b_0, b_1, b_2) = (1, 1, 1)$. The output looks as follows with details given in the end.

```
$ELR.pvalues
                   obs.teststat boot.pvalue
qx=(x,logx)            17.63242 0.012012012
qx=(logx,logx^2)       18.09275 0.007007007
qx=(x,logx,logx^2)     19.43278 0.028028028
qx=x                   16.76268 0.004004004
qx=logx                17.10396 0.001001001
```

The following gives the source `R` code for the above mentioned `R` functions.

```
library("nnet") # Load package for fitting multinomial logistic regression;

elr.part0 <- function(n){
n0 <- n[,1]
n1 <- n[,2]

alp <- n0/(n0+n1)
R0.alt <- sum(n0*log(alp))+sum(n1*log(1-alp))

alp0 <- sum(n0)/(sum(n0)+sum(n1))
R0.nul <- sum(n0)*log(alp0)+sum(n1)*log(1-alp0)

R0 <- 2*(R0.alt-R0.nul)
R0
}

##
elr.part1 <- function(x, qx=1){
x[,1] <- factor(x[,1])
group <- unique(x[,1])
m <- length(group)
n <- c()
for(i in 1:m){
```

```
 n <- c(n, sum(x[,1]==group[i]))
}
rho <- n/sum(n)

x1=x[,2]
x2=log(x1)
x3=x2*x2

 if(qx==1){ result <- summary(multinom(x[,1]~x1+x2, trace=F)) }  # q(x)=(1,x,logx)
 if(qx==2){ result <- summary(multinom(x[,1]~x2+x3, trace=F)) }  # q(x)=(1,logx,logx^2)
 if(qx==3){ result <- summary(multinom(x[,1]~x1+x2+x3, trace=F)) } # q(x)=(1,x,logx,logx^2)
 if(qx==4){ result <- summary(multinom(x[,1]~x1, trace=F)) }  # q(x)=(1,x)
 if(qx==5){ result <- summary(multinom(x[,1]~x2, trace=F)) }  # q(x)=(1,logx)

loglik <- -result$value-sum(n*log(rho))
## if also need AIC for selecting basis function;
# cat(paste("AIC =", result$AIC), "\n")

R1 <- 2*loglik
R1
}

##
elr <- function(data, qx=1){
data[,1] <- factor(data[,1])
group <- unique(data[,1])
m <- length(group)

ncount <- c()
  for(i in 1:m){
    xx=data[data[,1]==group[i],2]
    n00=sum(xx==0)
    n11=sum(xx>0)
    ncount=rbind(ncount, c(n00,n11))
  }
part0 <- elr.part0(ncount)
part1 <- elr.part1(data[data[,2]>0,], qx)
```

```
list(elrt=part0+part1)
}

## This is the main function;

boot.elr <- function(data, B){
data[,1] <- factor(data[,1])
group <- unique(data[,1])
    m <- length(group)
    N <- nrow(data)

boot.sample <- matrix(sample(data[,2], B*N, replace=T), N, B)

  pvalues <- NULL
  for(i in 1:5){
   test <- function(y){
      newdata <- data.frame(data[,1], y)
      elr(newdata, qx=i)$elrt
      }

  R.boot <- apply(boot.sample, 2, test)
  obs.teststat <- elr(data, qx=i)$elrt

  pvalue1 <- mean(R.boot > obs.teststat)

  res <- c(obs.teststat, pvalue1)
  pvalues <- rbind(pvalues, res)
  }

rnames <- c("qx=(x,logx)", "qx=(logx,logx^2)", "qx=(x,logx,logx^2)", "qx=x", "qx=logx")
cnames <- c("obs.teststat", "boot.pvalue")
dimnames(pvalues) <- list(rnames,cnames)

list(ELR.pvalues=pvalues)
}
```

```
## An artificial data example;

set.seed(2016)
n10 <- rbinom(1, 50, 0.2); x1=c(rep(0,n10), rlnorm(50-n10, meanlog = 0, sdlog = 1))
n20 <- rbinom(1, 50, 0.3); x2=c(rep(0,n20), rlnorm(50-n20, meanlog = 0.3, sdlog = 1))
n30 <- rbinom(1, 50, 0.4); x3=c(rep(0,n30), rlnorm(50-n30, meanlog = 0.5, sdlog = 1))
group.lab <- rep(LETTERS[1:3], rep(50,3))
group.data <- c(x1,x2,x3)
artificial.data <- data.frame(group.lab, group.data)

## Outputs for the artificial data example;

boot.elr(artificial.data, B=999)
> $ELR.pvalues
>                      obs.teststat boot.pvalue
> qx=(x,logx)              17.63242 0.012012012
> qx=(logx,logx^2)         18.09275 0.007007007
> qx=(x,logx,logx^2)       19.43278 0.028028028
> qx=x                     16.76268 0.004004004
> qx=logx                  17.10396 0.001001001
```

A.2 R functions for Chapter 4

In this section, we provide `R` functions to implement the ELR based method proposed in Chapter 4, specifically, for testing mean equality, with the basis function $\mathbf{q}(x) = \log(x)$ used in the DRM:

- `loglik.alt`: calculate the log-likelihood value under the alternative hypothesis via the logistic regression routine;
- `score`: calculate the sum of squared error loss of the score functions;
- `loglik.null`: calculate the log-likelihood value under the null hypothesis of equal means;
- `mele`: minimizing the above `score` function for a given initial point;
- `elrt`: the main function.

- The input contains four arguments, whereas it is sufficient to only supply the first argument.
- Input 1: "`dat`" input must be a data frame: 1st column is the group labels and 2nd column contains all observations (both zeros and nonzeros, and the nonzeros could be discrete or continuous);
- Input 2: "`ini`" input can be either missing or contains user supplied vector of initial values for the parameters; see the notes below for details.
- Input 3: "`tol`" input sets the tolerance level that controls the minimized value of the objective function for each given initial point, with default `1e-8`.
- Input 4: "`max.ini`" input controls the maximum number of different initial points to use before convergence, with default `200`; see the notes below for details.

- The output is a list object with following four components.
- Output 1: "`$score`" is the minimized value of the objective function.
- Output 2: "`$elrt`" returns the value of the proposed ELR test statistic R_n.
- Output 3: "`$p.value`" returns the corresponding p-value of the observed ELR test statistic R_n based on the chi-squared null limiting distribution.
- Output 4: "`$mele`" returns the parameter estimates and the corresponding values of the Lagrange multipliers under the null hypothesis; see the notes below for details.

As an illustration, we apply the proposed ELR to test the equality for two-sample means of the mice data example in Section 4.5 with two groups, Spring vs. Winter. The output looks typically as follows with details given in the end.

```
> elrt(artificial.data)
> $obj
> [1] 1.031804e-16
> $elrt
> [1] 1.171699
> $p.value
> [1] 0.5566329
> $mele
> [1] 0.128085525 0.327879781 0.429639613 -0.009463512 0.277062633 -0.053600908 0.453009705
> [8] 0.306526854 0.248136846 0.008831648  0.025351896
```

Here we will refer to some notation as defined in Chapter 4. We provide some notes on details of the above functions, and also give some tips for users:

- Note 1: the input "`ini`" is in the form of a vector containing initial of parameters ordering as $(\nu_0, \ldots, \nu_m, \alpha_1, \beta_1, \ldots, \alpha_m, \beta_m, t_1, \ldots, t_m)$.

- Note 2: if the input "`ini`" is missing, as default, we use the sample estimates of $\hat{\boldsymbol{\nu}}$ as initial value for $\boldsymbol{\nu}$, use the estimated parameters $(\hat{\alpha}_1, \hat{\beta}_1, \ldots, \hat{\alpha}_m, \hat{\beta}_m)$ obtained from logistic regression as initial value for $\boldsymbol{\theta}$, and set the initial value for $\mathbf{t}$ at its true value zero.

- Note 3: for the input "`ini`", we do not have to worry about $\boldsymbol{\lambda}$ by noting a fact as follows. By setting the $\partial\ell(\boldsymbol{\eta}, \boldsymbol{\lambda}, \mathbf{t})/\partial\boldsymbol{\psi}$ to $\mathbf{0}$, we get

$$\tilde{\lambda}_r = \frac{(1-\tilde{\nu}_r)n_{r0}}{\tilde{\nu}_r n_{\cdot 1}}, \qquad r = 1, \ldots, m.$$

  Hence, by incorporating this relationship, we can express $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\nu})$ as function of $\boldsymbol{\nu}$ in order to reduce the dimension of parameters in optimization.

- Note 4: the input "`max.ini`" specifies the maximum number of times that we repeat the following procedure to generate a new initial point or until a convergence is reached. Every time when the default or user supplied "`ini`" does not result in a successful convergence, we add a small random noise to the $\boldsymbol{\theta}$ components in "`ini`" and create a new "`ini`" for the next iteration. Note that we always suggest the initial values of $(\boldsymbol{\nu}, \mathbf{t})$ at $(\hat{\boldsymbol{\nu}}, \mathbf{0})$.

- Note 5: "`$mele`" returns the parameter estimates of $\tilde{\boldsymbol{\nu}}$ and $\tilde{\boldsymbol{\theta}}$, and the corresponding values of Lagrange multipliers $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\mathbf{t}}$, under the null hypothesis. The output is ordered as $(\tilde{\nu}_0, \ldots, \tilde{\nu}_m, \tilde{\alpha}_1, \tilde{\beta}_1, \ldots, \tilde{\alpha}_m, \tilde{\beta}_m, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_m, \tilde{t}_1, \ldots, \tilde{t}_m)$. Here $\tilde{\boldsymbol{\lambda}}$ (and each $\boldsymbol{\lambda}$ value in iterations) is determined by the above relationship $\tilde{\lambda}_r = \lambda_r(\tilde{\nu}_r)$.

- Note 6: for understanding the programming of the below attached source functions, we used `lambda1` and `lambda2` to stand for $\boldsymbol{\lambda}$ and $\mathbf{t}$ as used in Chapter 4.

- Tip 1: in some extreme cases that the positive observations of each group are well separated, or nearly well separated, the algorithm can fail to converge, especially when some sample sizes are too small or zero proportions are too large.

- Tip 2: our experiences suggest that these functions perform stable for up to four groups comparison, with moderate sample sizes.

The following gives the source R code for the above mentioned R functions.

```
library("nnet")   # load package for fitting multinomial logistic regression;
##
loglik.alt <- function(dat){

dat[,1] <- factor(dat[,1])
group <- unique(dat[,1])
m <- length(group)
ncount <- c()
  for(i in 1:m){
    xx=dat[dat[,1]==group[i],2]
    n00=sum(xx==0)
    n11=sum(xx>0)
    ncount=rbind(ncount, c(n00,n11))
  }
   n0=ncount[,1]
   n1=ncount[,2]
   rho=n1/sum(n1)
   nu=n0/(n0+n1)
    if(any(nu==0)){nu=nu+1e-10; warning("'some group observations are strictly positive'")}
   lambda1=(1-nu[-1])*ncount[-1,1]/(nu[-1]*sum(ncount[,2]))

   z1=dat[dat[,2]>0,2]
   z2=log(z1)

   result <- summary(multinom(factor(dat[dat[,2]>0,1])~z2, trace=F))

loglik <- sum(n0*log(nu))+sum(n1*log(1-nu)) - result$value-sum(n1*log(rho))

out <- matrix(result$coefficients, ncol=2)
out[,1] <- out[,1]-log(rho[-1]/rho[1]); theta <- as.vector(t(out))

return(list(value=as.numeric(loglik),m=m,ncount=ncount,nu=nu,theta=theta,lambda1=lambda1))
}

##
```

```
score <- function(dat, par){

dat[,1] <- factor(dat[,1])
group <- unique(dat[,1])
m <- length(group)
z <- dat[dat[,2]>0,2]

nu <- par[1:m]
theta <- par[(m+1):(3*m-2)]
lambda2 <- par[(3*m-1):(4*m-3)]

w.matrix <- NULL; ee.matrix <- NULL; ncount <- NULL; log.nonzero.sum <- NULL
  for(i in 1:m){
    xi <- dat[dat[,1]==group[i],2]
    n00=sum(xi==0)
    n11=sum(xi>0)
    ncount=rbind(ncount, c(n00,n11))

    if(i>1){
       wi <- exp(theta[2*i-3]+theta[2*i-2]*log(z))
       w.matrix <- rbind(w.matrix, wi)  # for q(x)=(1,log(x));
       eei <- (1-nu[i])*z*wi - (1-nu[1])*z
       ee.matrix <- rbind(ee.matrix, eei)
       xi.nonzero <- xi[xi>0]
       log.nonzero.sum <- c(log.nonzero.sum, sum(log(xi.nonzero)))
    }
   }

lambda1 <- (1-nu[-1])*ncount[-1,1]/(nu[-1]*sum(ncount[,2]))
deno <- 1+ as.vector(lambda1)%*%(w.matrix-1)+as.vector(lambda2)%*%ee.matrix

s1 <- ncount[1,1]/nu[1]-ncount[1,2]/(1-nu[1]) - sum(lambda2)*sum(z/deno)
ss.score <- s1^2
for(i in 2:m){
  s2 <- ncount[i,1]/nu[i]-ncount[i,2]/(1-nu[i]) + sum(lambda2[i-1]*z*w.matrix[i-1,]/deno)
  nume3 <- lambda1[i-1]*w.matrix[i-1,] + lambda2[i-1]*(1-nu[i])*z*w.matrix[i-1,]
  s3 <- ncount[i,2] - sum(nume3/deno)
```

```
  nume4<- lambda1[i-1]*log(z)*w.matrix[i-1,]+lambda2[i-1]*(1-nu[i])*z*log(z)*w.matrix[i-1,]
  s4 <- log.nonzero.sum[i-1] - sum(nume4/deno)
  s6 <- sum(ee.matrix[i-1,]/deno)

  ss.score <- ss.score + sum(s2^2,s3^2,s4^2,s6^2)
  }

if(is.na(ss.score)| ss.score>1e10|is.nan(ss.score)){ val=1e10 }

return(ss.score)
}

##
loglik.null <- function(dat, par){

dat[,1] <- factor(dat[,1])
group <- unique(dat[,1])
m <- length(group)
z <- dat[dat[,2]>0,2]

nu <- par[1:m]
theta <- par[(m+1):(3*m-2)]
lambda1 <- par[(3*m-1):(4*m-3)]
lambda2 <- par[(4*m-2):(5*m-4)]

w.matrix <- NULL; ee.matrix <- NULL; ncount <- NULL; log.nonzero.sum <- NULL
for(i in 1:m){
    xi <- dat[dat[,1]==group[i],2]
    n00=sum(xi==0)
    n11=sum(xi>0)
    ncount=rbind(ncount, c(n00,n11))

    if(i>1){
       wi <- exp(theta[2*i-3]+theta[2*i-2]*log(z))
       w.matrix <- rbind(w.matrix, wi)
       eei <- (1-nu[i])*z*wi - (1-nu[1])*z
       ee.matrix <- rbind(ee.matrix, eei)
```

```
        xi.nonzero <- xi[xi>0]
        log.nonzero.sum <- c(log.nonzero.sum,sum(theta[2*i-3]+theta[2*i-2]*log(xi.nonzero)))
    }
}

pkj <- log(1+ as.vector(lambda1)%*%(w.matrix-1)+as.vector(lambda2)%*%ee.matrix + 1e-20)
value <- sum(ncount[,1]*log(nu) + ncount[,2]*log(1-nu)) + sum(log.nonzero.sum) - sum(pkj)

return(loglik.null=value)
}

##
mele <- function(dat, par0){

res.alt <- loglik.alt(dat)
m <- res.alt$m

if(length(par0)!=3*(m-1)+m){
   stop(" 'parameter dimension does not match the no. of groups and basis function!' ")}

res <- nlminb(start=par0, score, lower=c(rep(0,m),rep(-20,(2*m-2)),rep(-2,m-1)),
              upper=c(rep(1,m),rep(20,(2*m-2)),rep(2,m-1)), dat=dat)
obj <- res$objective

par.temp <- as.vector(res$par)
nu=par.temp[1:m]
ncount <- res.alt$ncount
lambda1 <- (1-nu[-1])*ncount[-1,1]/(nu[-1]*sum(ncount[,2]))

mele <- c(par.temp[1:(3*m-2)], lambda1, par.temp[(3*m-1):(4*m-3)])
loglik0 <- loglik.null(dat, mele)
elrt.temp = res.alt$value - loglik0
elrt = as.numeric(2*elrt.temp)
pvalue = 1-pchisq(elrt, df=(m-1))

return(list(score=obj, elrt=elrt, p.value=pvalue, mele=mele))
}
```

```
## This is the main function to use;

elrt <- function(dat, ini, tol=1e-8, max.ini=200){

res.alt <- loglik.alt(dat)
m <- res.alt$m
alt <- c(res.alt$nu, res.alt$theta, rep(0,m-1))
lam1=res.alt$lambda1
if(missing(ini)==TRUE) ini=alt
res <- mele(dat, ini)
obj <- res$score
elrt <- res$elrt
j=0
while( (obj>tol | elrt<0 | is.nan(elrt) | is.na(elrt)) & j< max.ini ){

if(obj<10*m){
 ini<-c(res$mele[c(1:(3*m-2),(4*m-2):(5*m-4))]+c(rep(0,m),rnorm((2*m-2),0,0.01),rep(0,m-1)))}
 else{ ini <- alt + c(rep(0,m), rnorm((2*m-2),0, 0.1), rep(0,m-1)) }

val.null <- loglik.null(dat, c(ini[1:(3*m-2)],lam1,ini[(3*m-1):(4*m-3)]))
   while(is.nan(val.null) | is.na(val.null)){
    ini <- alt + c(rep(0,m), rnorm((2*m-2),0,sqrt(obj)/nrow(dat)), rep(0,m-1))
    val.null <- loglik.null(dat, c(ini[1:(3*m-2)],lam1,ini[(3*m-1):(4*m-3)]))
    }
res <- mele(dat, ini)
obj <- res$score
elrt <- res$elrt
j=j+1
}

if(j==max.ini){
cat(paste("\n", " !NOTE: the maximum number of different initial points reached:", j,
    "; may consider increasing max.ini and/or tol to converge! ", "\n", "\n")) }

return(list(obj=obj, elrt=elrt, p.value=res$p.value, mele=res$mele))
}
```

```
## Same illustrative artificial data example as used in last section;

set.seed(2016)
n10 <- rbinom(1, 50, 0.2); x1=c(rep(0,n10), rlnorm(50-n10, meanlog = 0, sdlog = 1))
n20 <- rbinom(1, 50, 0.3); x2=c(rep(0,n20), rlnorm(50-n20, meanlog = 0.3, sdlog = 1))
n30 <- rbinom(1, 50, 0.4); x3=c(rep(0,n30), rlnorm(50-n30, meanlog = 0.5, sdlog = 1))
group.lab <- rep(LETTERS[1:3], rep(50,3))
group.data <- c(x1,x2,x3)
artificial.data <- data.frame(group.lab, group.data)

## Outputs for this artificial data example;

elrt(artificial.data)
> $obj
> [1] 1.031804e-16
> $elrt
> [1] 1.171699
> $p.value
> [1] 0.5566329
> $mele
> [1] 0.128085525 0.327879781 0.429639613 -0.009463512 0.277062633 -0.053600908 0.453009705
> [8] 0.306526854 0.248136846 0.008831648  0.025351896
```