A Dissertation entitled Statistical Inferences under a Semiparametric Finite Mixture Model

> By Shiju Zhang

As partial fulfillment of the requirements for the Doctor of Philosophy Degree in Mathematics

Advisor: Prof. Biao Zhang

Graduate School

The University of Toledo December 2005

An Abstract of

Statistical Inferences under a Semiparametric Finite Mixture Model

Shiju Zhang

Submitted in partial fulfillment of the requirements for the Doctor of Philosophy Degree in Mathematics

> The University of Toledo December 2005

We consider the inference problem of a finite mixture model based on data from multiple samples, each of which is from a mixture of two common components. Under the assumption that the ratio of the two component densities takes a known parametric form, we obtain maximum semiparametric likelihood estimates of the parameters via EM or MM algorithms, and establish the large sample results for those estimators. We then develop empirical likelihood ratio-based statistics for constructing confidence intervals for and testing statistical hypotheses on mixing proportions. We show that the statistics are asymptotically chi-square distributed. In addition, a goodness-offit test is proposed for testing the density ratio assumption. Simulation studies are carried out to evaluate the performances of the proposed statistics and tests.

Acknowledgments

I would like to thank my advisor Prof. Biao Zhang for his creative ideas, numerous suggestions, patience, and so much help. My thanks also go to Prof. Donald White for the knowledge I have learnt from him. I would also take this opportunity to thank Dr. Yulin Li for his helpful seminar course.

I am grateful to my wife for her understanding and support.

To Saihua, Cindy, and David.

Contents

\mathbf{A}	bstra	ct	ii
A	cknov	wledgments	iii
D	edica	tion	iv
\mathbf{C}	onter	nts	v
\mathbf{Li}	st of	Figures	viii
Li	st of	Tables	ix
1	The	Problem of Interest, Literature Review, and Preliminaries	1
	1.1	The Problem of Interest	2
	1.2	Finite Mixture Models	4
	1.3	Density Ratio Model	11
	1.4	Empirical Likelihood	12
	1.5	EM and MM Algorithms	14
	1.6	Motivations and an Outline of This Research	16
	1.7	Some Notations	18

2	Sen	niparan	netric Estimation of the $K2M$ Model and Asymptotics	20
	2.1	Empir	ical Likelihood-based Semiparametric Estimators	21
	2.2	Strong	Consistency and Asymptotic Normality of $(\widetilde{\lambda}, \widetilde{\beta}, \widetilde{\alpha})$	25
	2.3	CDF	Estimators and a Graphical Diagnostic	28
	2.4	Proofs		30
		2.4.1	Proof of Lemmas $(2.2.1)$ and $(2.2.2)$	30
		2.4.2	Proof of Theorem $(2.2.3)$	32
		2.4.3	Proof of Theorem $(2.3.1)$	44
2	БЛ/	and M	IM Algorithms for the K2M Model and Its Special Cases	59
3	EIVI		IN Algorithms for the K2M Model and its Special Cases	54
	3.1	MM A	lgorithms for Model 2 and Model 3	52
	3.2	EM ar	nd MM Algorithms for the $K2M$ Model $\ldots \ldots \ldots \ldots \ldots$	55
		3.2.1	An MM Algorithm for K-Sample Normal Mixtures \ldots	55
		3.2.2	An EM Algorithm for the $K2M$ Model	57
		3.2.3	An MM Algorithm for the $K2M$ Model	62
	3.3	Simula	ation Studies	64
	3.4	Real E	Examples	65
1	Stat	tistical	Inference on Parameters	72
т	Dia	usuca		12
	4.1	Confid	ence Intervals	72
		4.1.1	A Normal-theory-based Confidence Region for λ	73
		4.1.2	A Likelihood Ratio-based Confidence Region for λ	73
		4.1.3	A Normal-theory-based Confidence Interval for β_2	74
		4.1.4	A Likelihood Ratio-based Confidence Interval for β_2	74

	4.2	Empirical Likelihood Ratio Tests	75	
	4.3	Simulation Studies	75	
		4.3.1 Q-Q plots for Empirical Likelihood Ratio Statistics	75	
		4.3.2 Coverage of Confidence Intervals	76	
	4.4	Proofs	79	
		4.4.1 Proof of Theorem (4.1.2)	79	
		4.4.2 Proof of Theorem $(4.1.1)$	82	
5	AK	Colmogrov-Smirnov-type GOF Test for the Density Ratio Model &	85	
	5.1	Introduction	85	
	5.2	Weak Convergence of $D_n(t)$	86	
	5.3	A Kolmogorov-Smirnov Goodness-of-Fit Test	86	
		5.3.1 The K-S Test Statistic: $\Delta_n \dots \dots \dots \dots \dots \dots \dots \dots$	87	
		5.3.2 A Bootstrap Procedure for Approximating the P -value	88	
		5.3.3 The Power under Local Alternatives	90	
		5.3.4 A Comparison of Powers: K-S Test against t Test	91	
	5.4	Examples	92	
	5.5	Proofs	93	
6	Cor	clusions and Future Research	98	
R	References 100			

List of Figures

3-1	Histograms of the six tuberculin data for 1990	70
3-2	Fitted CDF's and the non-parametric CDF's for 1990 tuberculin data.	70
3-3	Histograms of the 2-h plasma glucose data	71
4-1	Q-Q plot for λ	77
4-2	Q-Q plot for β_2	79

List of Tables

3.1	A simulation: estimation of mixture proportions from 1000 samples	66
3.2	Tolerance $\epsilon = 10^{-10};$ Pentium 4 CPU 2.53GHz 504 Mb of RAM $~$	67
4.1	Average length, midpoint and coverage for λ_1 from 1000 samples	78
4.2	Average length, midpoint and coverage for β_2 from 1000 samples	78
5.1	Powers for testing $\gamma = 0$ when $\psi(x; \beta_2) = \beta_2 x + \gamma x^2$ with $\beta_2 = 0.5$.	92

Chapter 1

The Problem of Interest, Literature Review, and Preliminaries

We first introduce the problem of interest, then briefly review the important research area of statistics - finite mixture models. We have two focuses. One is data, that is, whether the data consist of a single sample or multiple samples, and if the latter, whether training samples are available. The other is the method adopted to attack the problem, that is, whether the method is parametric, nonparametric or semiparametric. In the third section, we introduce the so-called density ratio model that naturally arises in the context of logistic regression. Next, we introduce the tool of (semiparametric) empirical likelihood which is often used to summarize information contained in given samples. To solve the resulting optimization problem in this dissertation work, we next in section five introduce the EM and MM algorithms. Finally, we raise problems that motivate the research and outline the main work completed in this research.

1.1 The Problem of Interest

Assume that random variables X_1, X_2, \dots, X_K have distributions mixed by two common components in different proportions; that is, X_k has a cumulative distribution function $H_k(x) = \lambda_k F(x) + (1 - \lambda_k)G(x)$, $k = 1, \dots, K$. Denote the corresponding densities of F(x), G(x) and $H_k(x)$ by f(x), g(x) and $h_k(x)$, respectively.

We assume that a sample is available from each mixture distribution and the K samples are jointly independent. Specifically, we make the following assumptions:

- (A₁) for each $k, x_{k1}, x_{k2}, \dots, x_{kn_k}$ are *iid* with CDF $H_k(x)$, and
- (A_2) the K samples are jointly independent.

For ease of reference, this model is referred to as a K2M model.

In order to identify the component distributions, further assumptions must be made. To avoid a distributional assumption about the two components, we further assume that

$$\log\left\{\frac{g(x)}{f(x)}\right\} = \beta_1 + \varphi(x;\beta_2), \qquad (1.1.1)$$

where the function $\varphi(x; \beta_2)$ is known, arbitrary up to a parameter β_2 , and differentiable w.r.t β_2 . To avoid the problem of irregularity as noted by Zou, Fine and Yandell (2002), we assume that, there exists at least one value of x such that $\varphi(x; \beta_2) \neq 0$. For example, in the case of $\varphi(x; \beta_2) = \beta_2^T x$, β_2 can not be zero. (1.1.1) is termed the density ratio model (Anderson, 1979; Qin, 1998). The K2M model is generally not identifiable when K = 1. In the sequel, we assume that K > 1. Identifiability of the parameters in the K2M model is then guaranteed by the theorem three of Gilbert, et al (1999). Under the assumption of the K2M model, we may write

$$dH_k(x) = [\lambda_k + (1 - \lambda_k)\omega(x;\beta)]dF(x), \quad k = 1, \cdots, K,$$
(1.1.2)

where $\omega(x;\beta) = \exp[\beta_1 + \varphi(x;\beta_2)]$ and $\beta = (\beta_1,\beta_2^T)^T$. (1.1.2) can be treated as a biased sampling problem (Vardi, 1982) with weights

$$\omega_k(x;\lambda,\beta) = \lambda_k + (1-\lambda_k)\omega(x;\beta), \quad k = 1, \cdots, K.$$

Special cases of the K2M model that were studied in the literature are the following.

Model 2: $K = 3, \lambda_1 = 1, \lambda_2 = 0, \lambda_3 = \lambda,$

Model 3: $K = 2, \lambda_1 = 1, \lambda_2 = \lambda,$

Model 4: $\lambda_1, \lambda_2, \cdots, \lambda_K$ are all known.

Model 2 was first proposed by Hosmer (1973). Under density ratio assumption (1.1.1), it was again studied by Anderson (1979), Qin (1999) and Zhang (2002). Model 3 is also a special case of Model 2. Without the density ratio assumption (1.1.1), Model 3 was considered by Smith and Vounatsou (1997). Zou, Fine and

Yandell (2002) and Fine et al (2004) studied model 4 in the genetics context.

The primary interest is, under the K2M model with K samples, to estimate the mixing proportions λ_k 's and to make inference about them. Since we have imposed the density ratio assumption, the assumption in turn calls for an evaluation of its goodness of fit.

1.2 Finite Mixture Models

Suppose that there are g populations G_1, G_2, \dots, G_g mixed up in some manner so that a randomly selected item from the superpopulation G has a distribution showing multimodality. Specifically, suppose that the g populations have density functions of $f_i(x; \theta_i), i = 1, 2, \dots, g$, respectively. Let D denote the indicator function taking value i if a randomly selected item from the superpopulation is in fact from subpopulation i. Then $\pi_i = p(D = i)$ is the proportion of G_i in the superpopulation. Consequently, the density of an observation in the superpopulation, $f(x; \theta, \pi)$, is the convex combination of the individual densities; that is,

$$f(x;\theta,\pi) = \sum_{i=1}^{g} \pi_i f_i(x;\theta_i),$$

where θ denotes the vector of unknown parameters of $\theta_1, \theta_2, \cdots, \theta_g$ and π is the vector of π_i 's, and

$$\sum_{i=1}^{g} \pi_i = 1 \quad \text{and} \quad \pi_i > 0 \quad (i = 1, 2, \cdots, g)$$

If a sample from the superpopulation G is available and parametric assumptions about the individual densities are made, one can estimate the unknown parameters π_i 's and θ_i 's. Inference about the mixing proportions π_i 's and inference about the number of components are often of interest in practice. For comprehensive review, see, for example, the monograph of McLachlan and Peel (1999).

In some classification problems, data from one or more subpopulations may be available. Such data are usually referred to as training data. Hosmer (1973) studied a two-component mixture model with data from the superpopulation and data from each of the two subpopulations. The data structure essentially appears as the following:

$$x_1, x_2, \cdots, x_{n_0}, \quad iid \sim F(x),$$

$$y_1, y_2, \cdots, y_{n_1}, \quad iid \sim G(y),$$

$$z_1, z_2, \cdots, z_{n_2}, \quad iid \sim H(z) = \lambda F(z) + (1 - \lambda)G(z).$$

Assuming normality for both component densities, Hosmer (1973) estimated the proportion of male and the proportion of female fish in a population of halibut from univariate data. The model was later studied by Murry and Titterington (1978) using bayesian and kernel methods, and Hall and Titterington (1984) by grouping data. Anderson (1979) studied the same problem with the density ratio model postulated to connect the two underlying densities. Recently, Qin (1999) revisited this problem under Anderson's assumption by making use of empirical likelihood. The computational issue of Hosmer's model under Anderson's assumption was investigated by Zhang (2002), who proposed a semiparametric EM algorithm, along with the logistic regression for estimating unknown parameters. In a different context, Heller and Qin (2001) studied this model using a nonparametric method. They made no assumption at all about the component densities or their relationship. Instead, they introduced a parameter to connect the two component densities. Specifically, let $\theta = p(Y_1 > X_1)$. One can show, under independence of X and Y, that $\theta = \int \bar{G}(x)dF(x)$, where $\bar{G}(x) = 1 - G(x)$. Since no distributional assumption was made, Heller and Qin used a pairwise rank-based likelihood method to perform inference on λ and θ .

Training samples may not be available from both subpopulations, as in Smith and Vounatsou (1997). Their work was motivated by the fact that many biomedical assays involve classifying samples into two groups according as whether some output (e.g., parasite density) exceeds a given cutoff. Many such assays do not classify all samples correctly because there is an overlap between the distributions of the output for the two groups. Often, however, a sample from the distribution of true negative (that is, the control) is available but there is no gold standard - a perfect classifier - for the true positive, which can only be identified with by using the assay itself, and then with uncertainty. This problem arises in parasite epidemiology (Smith et al, 1994). Clinical malaria can be diagnosed by the presence of parasites and fever. However, in endemic area, children can tolerate malaria parasites without the development of any signs of disease and they may have fever for other causes. In their analysis, Smith and Vounatsou considered a mixture of two components with an uncategorized sample from the mixture and a categorized sample from one of the two components. They assumed that the risk of positivity is known to be a monotonic function of the output, the larger, the more indicative of a positive. They proposed four different methods to estimate the mixing proportion and their methods were used to analyze data from a study that quantifies the capacity of mouse cells in culture to transfer small molecules between one another. The same model was discussed by Vounatsou et al (1998) in the bayesian framework using the Gibbs sampler. The first step of their method is similar to that of Hall (1981); that is, both samples are grouped subjectively into 6 to 10 categories.

In the situation that there is no training sample available, Choi (1979) considered a 2-sample two-component mixture model. Two samples are each from one of the two mixtures so that there is no membership information available, which is different from the above situations where membership information is available and information about at least one subpopulation is available. As explained by Choi (1979), mixture models are of relevance in a life testing situation, where there are two causes for failure which act in a mutually exclusive manner. The density of the random variable representing failure or survival time can be modelled as $h(x; \phi) = \pi f_1(x; \theta_1) + (1 - 1) f_2(x; \phi_2)$ π) $f_2(x;\theta_2)$, where $f_i(x;\theta)$ is the density for failure time solely due to cause i (i = 1, 2). Choi (1979) used the above model to compare the toxicity of two chemical agents used in chemotherapy. It was assumed that death was attributable either to toxicity of the agent or to regrowth of the tumor; the toxicity death usually precedes the passage due to regrowth. Following Choi (1979), the density of the time to death under chemical agent k is $h_k(x; \phi_k) = \pi_k f_1(x; \theta_1) + (1 - \pi_k) f_2(x; \theta_2), \quad (k = 1, 2), \text{ where}$ $\phi_k = (\pi_k, \theta_1^T, \theta_2^T)^T$. Testing whether the two mixtures are identical is equivalent to testing whether $H_0: \pi_1 = \pi_2$ holds. For univariate normal and exponential densities, Choi presented four methods for testing H_0 , the homogeneity of the two mixing proportions. Two were parametric tests based on the asymptotic normality of two unbiased estimators of $\pi_1 - \pi_2$, while the other two were nonparametric based on the U and Kolmogorov-Smirnov statistics respectively. The same model was studied by McLachlan et al (1982). In order to test the equality of the two mixing proportions, with normal assumptions about both component densities, a likelihood ratio test was put forward. Their test has more power, at least asymptotically, and also it can handle multivariate data. Since the null hypothesis H_0 is specified in the interior of the parameter space, regularity conditions do not break down as with null hypotheses concerning the number of distinct components in the mixture. McLachlan et al (1982) analyzed data representing survival times in weeks for two sets of rats which were given dosages of cytoxan at a concentration of 60 mg/kg. The second set was given the full dosage once weekly, while the first received half the dosage twice weekly. They compared the toxicity of the chemical agent at the two dosage levels, assuming that f_1 and f_2 are both normal with different means and different variances.

In the context of tuberculous infection, Nagelkerke et al (2001) studied mixtures of M. tuberculosis (TB) and non-specific tuberculin reactions – environmental mycobacteria (EM). Their primary purpose is to estimate tuberculosis infection prevalence. To meet this end, the marginal distribution of indurations needs to be separated into its component distributions. Here induration is a skin reaction of variable size caused by allergy to some kind of chemical agent. When EM infections are uncommon and the prevalence of TB infection high, the distribution of indurations in response to tuberculin often has two clear modes, one near 0 mm, and one somewhere between 14 and 20 mm of induration size. The author assumed that all non-infected individuals

have inducations $\leq 1 \text{ mm}$ and consequently that inducations greater than 1 mm are caused either by infection with EM or by infection with TB. The advantage of this is that only mixtures of two distribution need to be considered.

For separating a mixture into its component distributions, some knowledge of, or assumptions about, the two distributions is required. Without any restrictions on the distributions, the problem may not have a unique solution. Thus, some prior knowledge on the two distributions has to be obtained or assumed. Based on these considerations, Nagelkerke et al (2001) postulated a logistic model for the probability of a TB infection given an observed induration x

$$\frac{p(\mathrm{TB}|x)}{p(\mathrm{EM}|x)} = \exp(\alpha_k + \beta x), \quad k = 1, 2, \cdots, K,$$

where α_i 's are population-specific "prevalence" parameters and β is the common odds ratio across all K populations. By Bayes' rule, this is in fact equivalent to the following semiparametric model:

$$g_k(x) = f_k(x) \exp(\alpha_k^* + \beta x), \quad k = 1, 2, \cdots, K,$$

where $\alpha_k^* = \alpha_k + \log \frac{\pi_k}{1-\pi_k}$ and $\pi_k = p(\text{TB in population } k)$, the prevalence of TB in population k, and $f_k(x) = f_k(x|\text{EM})$ and $g_k(x) = f_k(x|\text{TB})$ are the underlying densities of EM and TB in population k, respectively.

Nagelkerke et al (2001) analyzed a data set consisting of 42 samples, each of which is from one of 42 mixtures. They estimated the parameters in the framework of contingency tables using Poisson log-likelihood. Neuenschwander et al (2000) used a Bayesian approach to finite mixture analysis of the same data. They made varying assumptions about the two component densities. The assumptions range from the assumption that both belong to a fully parameterized class of distributions (e.g., Weibull, gamma or log-normal) to the "vague" assumption that the distribution of TB infections is log-concave, and no assumptions regarding the EM distribution are made. Their most general assumption is that both distributions are log-concave. MCMC methods using the Metropolis algorithm were used for estimating all parameters.

Recently, Zou, Fine and Yandell (2002) and Fine, et al (2004) considered a rat study of the genetics of tumor development. The observations arose from $K (\geq 2)$ discrete mixtures consisting of $L (\geq 2)$ components. In the genetics set-up, K is the number of genotypes at the flanking markers and L is the number of possible genotypes at the locus between the markers. Each observation originates in one of the K mixtures and the label of the mixture generating the datum is known. The mixing probabilities are also known and may vary among the K mixtures, while the L component distributions are common to all mixtures. The standard analyses for such problems posit parametric models for the component distributions (Doerge et al, 1997). Since standard approaches to estimation involve parametric assumptions for the component distributions and may be sensitive to model misspecification, Zou, Fine and Yandell (2001), in the case of L = 2, used a semiparametric method by assuming a density ratio model for the two component densities, while Fine, et al (2004) took a fully nonparametric arguments for general L in their study.

1.3 Density Ratio Model

If two densities are related in such a way that the ratio of them takes a known and arbitrary form, then they are said to follow a density ratio model. For two densities f(x) and g(x), the density ratio model is represented by

$$\frac{g(x)}{f(x)} = \exp(\beta_1 + \varphi(x; \beta_2)), \qquad (1.3.1)$$

where $(\beta_1, \beta_2^T)^T$ is the parameter and φ is a smooth function taking a known form.

The density ratio model (1.3.1) arises naturally from the logistic regression model. Following Kay and Little (1987) and Qin and Zhang (1997), suppose we have data concerning an explanatory variable X and a binary response variable Y taking values 0 and 1. Denote the conditional density functions of X given Y = j (j = 0, 1) by f(x) and g(x), respectively. Consider the assumption that the log-ratio of the two densities is a linear function of x; that is, $\log[\frac{g(x)}{f(x)}] = \beta_1^* + \beta_2^T x$. Since, by Bayes' rule,

$$\log \frac{p(Y=1|x)}{p(Y=0|x)} = \log \frac{g(x)}{f(x)} + \log \frac{\pi}{1-\pi},$$

where $\pi = p(Y = 1)$, we have that

$$\log \frac{p(Y=1|x)}{p(Y=0|x)} = \beta_1^* + \beta_2^T x + \log \frac{\pi}{1-\pi} = \beta_1 + \beta_2^T x,$$

where $\beta_1 = \beta_1^* + \log \frac{\pi}{1-\pi}$. Thus, if the log-ratio of the conditional densities of X given Y is linear in x, the logistic model is correct for the conditional distribution of Y given

X = x. The same argument extends to the general case of (1.3.1), which includes, as special cases, the biased sampling problem, the multiplicative-intercept risk model (Hsieh, et al, 1985; Weinberg, et al, 1993; Zhang, 2000) which is an extension of the odds-linear model (Weinberg and Sandler, 1991; Wacholder and Weinberg, 1994), and some familiar generalized linear models such as the logistic, the probit, and the c-loglog models with case-control data. In application, the function $\varphi(x;\beta)$ can take a form as simple as βx (Qin and Zhang, 1997), which includes many familiar densities, such as two normal densities with the same variance and two exponential densities, as special cases, or a slightly complicated form as $\beta_2 \phi(x)$, which includes the exponential family as a special case.

1.4 Empirical Likelihood

Without assumptions about the distribution underlying a data set, Art Owen, in a series of papers (Owen 1988, 1990, 1991), has proposed a strategy based on the multinomial likelihood function. Assume we have a data set of *iid* observations X_1, X_2, \dots, X_n . The data points may be, essentially, of any Euclidean dimension, say d. Then the *empirical likelihood* of the data $\{X_1, X_2, \dots, X_n\}$ is defined as L(F) = $\prod_{i=1}^n p_i$, where $p_i = dF(X_i), i = 1, 2, \dots, n$, are the jumps of the distribution of F at data points, summing to unity. Inference may be based on

$$\Re(\theta) = \sup\{\prod_{i=1}^{n} np_i \; ; \; T(F) = \theta, \; \sum_{i=1}^{n} p_i = 1, p_i \ge 0, i = 1, 2, \cdots, n.\}$$

the so called *profile empirical likelihood ratio function*. The constraint $T(F) = \theta$ may

take the form of $E[g(X;\theta)] = 0$; that is, $g(x;\theta)$ is an *M*-functional. In this case,

$$\Re(\theta) = \sup\{\prod_{i=1}^{n} np_i \mid \sum_{i=1}^{n} p_i g(X_i; \theta) = 0, p_i \ge 0, \sum_{i=1}^{n} p_i = 1\}$$

Assume that $E[g(X;\theta_0)] = 0$, and $Var[g(X;\theta_0)]$ is finite and has rank q > 0. Then as shown by Qin and Lawless (1994), $-2\log \Re(\theta) \to \chi_q^2$ in distribution as $n \to \infty$. Using this fact, we can test the hypothesis about θ or construct a confidence region for θ . For example, the empirical likelihood ratio-based $1 - \alpha$ confidence region is of the form $\{\theta| - 2\log \Re(\theta) \le \chi_q^2(1 - \alpha)\}$. The empirical likelihood ratio-based test of the null hypothesis that $\theta = \theta_0$ rejects when $-2\log \Re(\theta_0)$ is greater than an appropriate critical χ^2 value. Qin and Lawless (1994) discussed how to link estimating equations and the empirical likelihood, and developed methods of combining auxiliary information about parameters.

An essential advantage of the empirical likelihood is that any side information can be taken into account in terms of constraints, subject to which the empirical likelihood is maximized. For a comprehensive review of empirical likelihood, see Hall and La Scala (1990) and Owen's recent monograph (Owen 2001).

Owen's empirical likelihood has been extended to the semiparametric setting by Qin (1993, 1998, 1999, 2000), Qin and Zhang (1997), Qin, Leung and Shao (2002), Qin and Leung (2005), and Qin and Zhang (2005), among others. Semiparametric empirical likelihood has been used in the context of goodness-of-fit tests (Qin and Zhang 1997; Zhang 1999, 2001), mixture models (Qin 1999; Zhang 2002; Qin and Leung 2005), complex survey (Qin et al, 2002), truncated data (Li, et al, 1997), and other contexts. Qin (1993) established the connection of semiparametric empirical likelihood to the biased sampling problem. It's well known that parametric likelihood of high dimensional data can be written as the product of marginal likelihood and conditional likelihood. A similar result about semiparametric empirical likelihood is established by Qin and Zhang (2005) in some useful contexts.

1.5 EM and MM Algorithms

Maximum likelihood is the dominant form of estimation in applied statistics. Because closed-form solutions to likelihood equations are the exception rather than the rule, numerical methods for finding maximum likelihood estimates are of paramount importance. At the heart of the EM algorithm is some notion of missing data. Data can be missing in the ordinary sense of a failure to record certain observations on certain cases. Data can also be missing in a theoretical sense. An EM starts from constructing so-called the complete likelihood. That is, one pretends that there is no missing data and construct a likelihood based on both the observed data and the missing data. When the complete data likelihood gets maximized, any missing value will be filled in with some appropriate procedure. Specifically, an EM algorithm consists of two steps. The E step finds a way to impute the missing values, usually by taking expectation. After filling in the missing values in the E step, a usually simpler surrogate function to the original likelihood is maximized in the M step. The price we pay for this simplification is that the EM algorithm is iterative and usually slow.

One of the notable advantage of the EM algorithm is its numerical stability. Any

EM algorithm leads to a steady increase in the likelihood of the observed data. In addition, EM algorithms handle parameter constraints gracefully. Any constraints are handled in M step. For a comprehensive review, see "The EM algorithm and extensions" by McLachlan (1997).

The MM algorithm is another useful optimization tool in mathematics and statistics, which has not been widely recognized by statisticians. As stated in Lange's new monograph (Lange, 2004), the MM algorithm is an extension of the EM algorithm. In maximization problems, the first M stands for "minorize" and the second for "maximize". When it is successful, the MM algorithm replace a difficult optimization problem by a simpler one. Simplicity can be attained by avoiding large matrix inversions, linearizing a optimization problem, separating the variables and dealing with constraint optimization problem nicely, and so on. As for the EM algorithm, simplification is paid with iteration and slowness of convergence. On the other hand, the MM algorithm shares the similar properties as with the EM algorithm, for example, numerical stability.

Here is an introduction of the MM algorithm. When an objective function, say f(x), is hard to maximize, it can be replaced by a carefully chosen surrogate function, denoted by $g(x|x^{(m)})$ satisfying

1.
$$f(x) \ge g(x|x^{(m)})$$
 for all $x \in D$, the domain of $f(x)$,

2.
$$f(x^{(m)}) = g(x^{(m)}|x^{(m)}),$$

where $x^{(m)}$ is the current iterate and g is called the minorizing function of f. This method is called MM algorithm which is useful especially in solving high dimensional problems. The construction of the surrogate function is toward the goal of making the parameters as isolated to each other as possible.

The MM algorithm for maximization problems has a nice ascent property; that is, $f(x^{(m)} < f(x^{(m+1)})$, which makes the algorithm numerically stable though it is usually slow. Minorization has many properties, one of which is the additivity. That is, if g_1 and g_2 are minorizing functions of f_1 and f_2 , respectively, then $g_1 + g_2$ is a minorizing function of $f_1 + f_2$.

In statistics, one often deals with the problem of maximizing a log-likelihood function. The log-likelihood can be written as the sum of usually nonlinear terms. Each nonlinear term can be minorized by a surrogate function. By the additivity property of MM algorithm, the log-likelihood can be replaced by a simpler function.

1.6 Motivations and an Outline of This Research

K-sample mixtures with training sample(s) have been widely considered by many researchers using alternative methods. K-sample mixtures without training samples arise in many contexts such as life testing, epidemiology and genetics, yet, to my knowledge, have not received much attention, which is the fact that motivates me to study mixtures. Although nonparametric methods, such as grouping (Hall, 1981) and pairwise rank based methods (Heller and Qin, 2002), may be possible, I go with semiparametric methods which basically assume the validity of the density ratio model. This topic is different from Zou, Fine and Yandell (2002) in that the mixing proportions in my setting are unknown. Zou's known proportions were derived from well-known genetics assumptions. By proposing mixture models with unknown mixing proportions, we provide a way to test the proportions, thus possibly some genetics assumptions. This is a second motivation.

Goodness of fit of models should generally be assessed after model fitting. To make valid inferences, the density ratio model should pass a goodness-of-fit test. There has been no formal goodness-of-fit test proposed for K-mixtures in the literature, therefore, it is a third motivation for me to choose this research topic.

A fourth motivation is to make inference about the difference of two proportions. Suppose two groups of subjects are administered with different treatments. A gene of interest may be normally or abnormally expressed. This can be viewed as a 2-sample two-component mixture model problem. The effects of treatments can be compared in terms of the difference of the two proportions that the gene is abnormally expressed in the two groups. A comparison of death rates due to toxicity for two chemical agents is another example. Note that the comparison of proportions here is different from the comparison of usual population proportions based on two samples in that the group identifier of each observation taken from each mixture is missing, which clearly complicates the comparison.

Following an empirical likelihood-based procedure, in Chapter two, we obtain the maximum semiparametric likelihood estimators for the parameters. The large sample results will be given. Since the empirical likelihood method is computer-intensive, the computational issue is discussed in Chapter three. Instead of the usual choice of Newton-Raphson method, we consider two alternatives – the EM method and the MM method. Problems of inference on interest parameters are studied in Chapter

four. Since our methodology is based on the density ratio model, inference should be drawn only when this assumption is valid. Chapter five considers the goodnessof-fit problem of our model. We follow the idea of Qin and Zhang (1997), Zhang (2001, 2002b) to develop a Kolmogorov-Smirnov-type statistic for testing the fit of our model. Chapter six concludes the research and points out possible future work.

1.7 Some Notations

The following equations or notation will be used throughout.

$$\begin{split} \theta &= (\lambda, \beta, \alpha) \qquad \theta_0 = (\lambda_0, \beta_0, 0) \qquad \lambda_0 = (\lambda_{10}, \cdots, \lambda_{K0}) \qquad \beta_0 = (\beta_{10}, \beta_{20}) \\ l(\lambda, \beta) &= \text{the profile log-likelihood} \qquad Q_n = \frac{1}{n} \frac{\partial l}{\partial \theta} \qquad S = \frac{\partial^2 l}{\partial \theta \theta^T} \\ \rho_{nk} &= \frac{n_k}{n} \qquad \rho_k = \lim_{n \to \infty} \rho_{nk}, \qquad \text{for } k = 1, 2, \cdots, K \\ \rho_n &= (\rho_{n1}, \cdots, \rho_{nK})^T \qquad \rho = (\rho_1, \rho_2, \cdots, \rho_K)^T \\ e_k &= (0, \cdots, 0, 1, 0, \cdots, 0)^T, \quad \text{the } k\text{th element being } 1, \text{ for } k = 1, 2, \cdots, K \\ \xi &= \frac{1}{n} \sum_{k=1}^K n_k (1 - \lambda_{k0}) \qquad \phi = \lim_{n \to \infty} \xi = \sum_{k=1}^K \rho_k (1 - \lambda_{k0}) \\ \widehat{H}_k(x) \text{is the ECDF of } x_{k1}, \cdots, x_{kn_K} \end{split}$$

$$\begin{split} \omega_{k}(t;\lambda_{k0},\beta_{0}) &= \lambda_{k0} + (1-\lambda_{k0})\omega(t;\beta_{0}) \\ r(t;\lambda_{0},\beta_{0}) &= 1 + \xi[\omega(t;\beta_{0})-1] = \sum_{k=1}^{K} \rho_{nk}\omega_{k}(t;\lambda_{k0},\beta_{0}) \\ R(t;\lambda_{0},\beta_{0}) &= \lim_{n\to\infty} r(t;\lambda_{0},\beta_{0}) = 1 + \phi[\omega(t;\beta_{0})-1] \\ g(t;\lambda_{0},\beta_{0}) &= \frac{\omega(t;\beta_{0})-1}{r(t;\lambda_{0},\beta_{0})} \\ G(t;\lambda_{0},\beta_{0}) &= \lim_{n\to\infty} g(t;\lambda_{0},\beta_{0}) = \frac{\omega(t;\beta_{0})-1}{R(t;\lambda_{0},\beta_{0})} \\ l_{1}(\lambda,\beta,\alpha) &= -\sum_{i=1}^{n} \log[1+\alpha g(t_{i};\lambda,\beta)]] \\ l_{2}(\lambda,\beta,\alpha) &= -\sum_{i=1}^{n} \log[r(t_{i};\lambda,\beta)] + \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} \log[\omega_{k}(x_{kj};\lambda,\beta)] \\ l(\lambda,\beta,\alpha) &= l_{1}(\lambda,\beta,\alpha) + l_{2}(\lambda,\beta,\alpha) \end{split}$$

$$\begin{aligned} \tau(t) &= \int_{-\infty}^{t} \frac{[1 - \omega(x; \beta_0)]^2}{R(x; \lambda_0, \beta_0)} dF \\ \eta_k(t) &= \int_{-\infty}^{t} \frac{[1 - \omega(x; \beta_0)]^2}{\omega_k(x; \lambda_{k0}, \beta_0)} dF \\ \psi_{1k}(t) &= \int_{-\infty}^{t} \frac{\partial \omega(x; \beta_0)}{\partial \beta} \frac{1}{\omega_k(x; \lambda_{k0}, \beta_0)} dF \\ \psi_{2k}(t) &= \int_{-\infty}^{t} \frac{\partial \omega(x; \beta_0)}{\partial \beta} \frac{\partial \omega(x; \beta_0)}{\partial \beta^T} \frac{1}{\omega_k(x; \lambda_{k0}, \beta_0)} dF \\ \varsigma_1(t) &= \int_{-\infty}^{t} \frac{\partial \omega(x; \beta_0)}{\partial \beta} \frac{1}{R(x; \lambda_0, \beta_0)} dF \\ \varsigma_2(t) &= \int_{-\infty}^{t} \frac{\partial \omega(x; \beta_0)}{\partial \beta} \frac{\partial \omega(x; \beta_0)}{\partial \beta^T} \frac{1}{R(x; \lambda_0, \beta_0)} dF \\ \tau &= \tau(\infty) \quad \eta_k = \eta_k(\infty) \quad \psi_{1k} = \psi_{1k}(\infty) \quad \psi_{2k} = \psi_{2k}(\infty) \quad \varsigma_1 = \varsigma_1(\infty) \quad \varsigma_2 = \varsigma_2(\infty) \end{aligned}$$

Chapter 2

Semiparametric Estimation of the K2M Model and Asymptotics

We start working on the problem of interest presented in the previous chapter. We first construct the nonparametric likelihood of our data. By separating the likelihood into a first part containing n nuisance parameters and a second part containing finite parameters, we form the so-called empirical likelihood that corresponds to the first part of the partition. The empirical likelihood is first maximized subject to some constraints that reflect our knowledge about the data. Then a profile likelihood is formed by plugging the maximizer just obtained back in the original nonparametric likelihood. Computation and inference will essentially be based on this profile likelihood.

2.1 Empirical Likelihood-based Semiparametric Estimators

Following the problem setting in Section 1.1, let $t = (x_{k1}, \dots, x_{kn_k}; \dots; x_{K1}, \dots, x_{Kn_K})$ denote the combined data, $\lambda = (\lambda_1, \dots, \lambda_K)^T$, and $\beta = (\beta_1, \beta_2)^T$. Based on the observed data t and the density ratio assumption (1.1.1), the likelihood can be written as

$$L(\lambda,\beta,F) = \prod_{k=1}^{K} \prod_{j=1}^{n_k} \omega_k(x_{kj};\lambda,\beta) dF(x_{kj})$$
$$= \prod_{i=1}^{n} p_i \prod_{k=1}^{K} \prod_{j=1}^{n_k} \omega_k(x_{kj};\lambda,\beta),$$

where $p_i = dF(t_i), i = 1, \dots, n$ with $p_i > 0$ and $\sum p_i = 1$. The log-likelihood is

$$l(\lambda, \beta, F) = \sum_{i=1}^{n} log(p_i) + \sum_{k=1}^{K} \sum_{j=1}^{n_k} log[\omega_k(x_{kj}; \lambda, \beta)]$$
(2.1.1)

which can be maximized in two steps as follows.

Step 1. For fixed (λ, β) , one maximizes the empirical log-likelihood $\sum_{i=1}^{n} log(p_i)$, subject to the constraints

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i[\omega(t_i, \beta) - 1] = 0, \quad p_i > 0,$$
(2.1.2)

where $\omega(t;\beta)$ is given in (1.1.2). This gives (Qin and Lawless, 1994)

$$p_i = \frac{1}{n} \frac{1}{1 + \eta[\omega(t_i; \beta) - 1]}, \quad i = 1, \cdots, n,$$
(2.1.3)

where η is the Lagrange multiplier, which is determined by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\omega(t_i;\beta)-1}{1+\eta[\omega(t_i;\beta)-1]}=0, \quad i=1,\cdots,n.$$
(2.1.4)

We change the variables (λ, β, η) to (λ, β, α) , where

$$\alpha = \eta - \sum_{k=1}^{K} \rho_k (1 - \lambda_k).$$
(2.1.5)

Then p_i can be written as

$$p_{i} = \frac{1}{n} \frac{1}{r(t_{i}; \lambda, \beta)} \frac{1}{1 + \alpha[\omega(t_{i}; \beta) - 1]/r(t_{i}; \lambda, \beta)},$$
(2.1.6)

where $r(t; \lambda, \beta) = 1 + \xi[\omega(t; \beta) - 1], \quad i = 1, \dots, n \text{ and } \xi = \frac{1}{n} \sum_{k=1}^{K} n_k (1 - \lambda_k).$ The constrained equation (2.1.4) now becomes

$$\frac{1}{n}\sum_{i=1}^{n}\frac{g(t_i;\lambda,\beta)}{1+\alpha g(t_i;\lambda,\beta)} = 0,$$
(2.1.7)

where

$$g(t;\lambda,\beta) = \frac{\omega(t;\beta) - 1}{r(t;\lambda,\beta)}.$$
(2.1.8)

The major advantage of changing variable is that, the log-likelihood (2.1.1) can be

written in a form from which a useful partial likelihood is naturally generated. This will be remarked in Section (2.2). Another advantage of changing variable is the fact that

$$\begin{split} E\left[\frac{1}{n}\sum_{i=1}^{n}g(t_{i};\lambda_{0},\beta_{0})\right] &= \frac{1}{n}E\left[\sum_{k=1}^{K}\sum_{j=1}^{n_{k}}g(x_{kj};\lambda_{0},\beta_{0})\right] \\ &= \sum_{k=1}^{K}\frac{n_{k}}{n}Eg(x_{kj};\lambda_{0},\beta_{0}) \\ &= \frac{1}{n}\sum_{k=1}^{K}n_{k}\int g(t;\lambda_{0},\beta_{0})\omega_{k}(t;\lambda_{0},\beta_{0})dF(t) \\ &= \int\left\{\frac{1}{n}\sum_{k=1}^{K}n_{k}[\lambda_{k0} + (1-\lambda_{k0})\omega(t;\beta_{0})]\right\}g(t;\lambda_{0},\beta_{0})dF(t) \\ &= \int [1+\xi(\omega(t;\beta_{0})-1)]g(t;\lambda_{0},\beta_{0})dF(t) \\ &= \int r(t;\lambda_{0},\beta_{0})\frac{\omega(t;\beta_{0})-1}{r(t;\lambda_{0},\beta_{0})}dF(t) \\ &= 0, \end{split}$$

and thus, by the central limit theorem,

$$\frac{1}{n} \sum_{i=1}^{n} g(t_i; \lambda_0, \beta_0) = \sum_{k=1}^{K} \frac{n_k}{n} \frac{1}{n_k} \sum_{j=1}^{n_k} g(x_{kj}; \lambda_0, \beta_0)$$
$$= \sum_{k=1}^{K} \frac{n_k}{n} \left[\mu_k + O_p(n_k^{-1/2}) \right]$$
$$= O_p(n^{-1/2}),$$
(2.1.9)

where $\mu_k = E(g(X_k; \lambda_0, \beta_0))$. A result of almost sure convergence can be obtained by

using the law of the iterated logarithm:

$$\frac{1}{n} \sum_{i=1}^{n} g(t_i; \lambda_0, \beta_0) = \sum_{k=1}^{K} \frac{n_k}{n} \frac{1}{n_k} \sum_{j=1}^{n_k} g(x_{kj}; \lambda_0, \beta_0)$$
$$= \sum_{k=1}^{K} \frac{n_k}{n} \left[\mu_k + O(n_k^{-1/2} (\log \log n_k)^{1/2}) \right]$$
$$= \sum_{k=1}^{K} \frac{n_k}{n} O(n_k^{-1/2} (\log \log n_k)^{1/2}) \quad (a.s.).$$

By similar arguments as in Qin and Lawless (1994), we can show that, under certain conditions, the constraint equation (2.1.7) uniquely determines an differentiable implicit function $\alpha = \alpha(\lambda, \beta)$ in an neighborhood of (λ_0, β_0) , the true value of (λ, β) . Plugging (2.1.6) in (2.1.1) gives the profiled log-likelihood

$$l(\lambda, \beta, \alpha(\lambda, \beta)) = l_1(\lambda, \beta, \alpha(\lambda, \beta)) + l_2(\lambda, \beta),$$

where

$$l_1(\lambda,\beta,\alpha(\lambda,\beta)) = -\sum_{i=1}^n \log[1+\alpha(\lambda,\beta)g(t_i;\lambda,\beta)],$$

$$l_2(\lambda,\beta) = -\sum_{i=1}^n \log[r(t_i;\lambda,\beta)] + \sum_{k=1}^K \sum_{j=1}^{n_K} \log[\omega_k(x_{kj};\lambda,\beta)].$$
(2.1.11)

Step 2. Maximize $l(\lambda, \beta, \alpha(\lambda, \beta))$ with respect to (λ, β) . We have the following score

equations

$$\frac{\partial l}{\partial \lambda} = \frac{\partial l_2}{\partial \lambda} - \sum_{i=1}^n \frac{\alpha(\lambda, \beta) \partial g(t_i; \lambda) / \partial \lambda + \partial \alpha(\lambda, \beta) / \partial \lambda g(t_i; \lambda, \beta)}{1 + \alpha(\lambda, \beta) g(t_i; \lambda, \beta)} = 0,$$
$$\frac{\partial l}{\partial \beta} = \frac{\partial l_2}{\partial \beta} - \sum_{i=1}^n \frac{\alpha(\lambda, \beta) \partial g(t_i; \lambda) / \partial \beta + \partial \alpha(\lambda, \beta) / \partial \beta g(t_i; \lambda, \beta)}{1 + \alpha(\lambda, \beta) g(t_i; \lambda, \beta)} = 0,$$

which are equivalent to

$$\frac{\partial l}{\partial \lambda} = \frac{\partial l_2}{\partial \lambda} - \sum_{i=1}^n \frac{\alpha(\lambda, \beta) \partial g(t_i; \lambda) / \partial \lambda}{1 + \alpha(\lambda, \beta) g(t_i; \lambda, \beta)} = 0,$$

$$\frac{\partial l}{\partial \beta} = \frac{\partial l_2}{\partial \beta} - \sum_{i=1}^n \frac{\alpha(\lambda, \beta) \partial g(t_i; \lambda) / \partial \beta}{1 + \alpha(\lambda, \beta) g(t_i; \lambda, \beta)} = 0.$$
(2.1.12)

Let $(\tilde{\lambda}, \tilde{\beta})$ be a solution to the above system of equations in the neighborhood of the true value of (λ_0, β_0) . We call $(\tilde{\lambda}, \tilde{\beta})$ a maximum semiparametric likelihood estimator (MSLE) of (λ, β) . The main results are presented in the following series of theorems.

2.2 Strong Consistency and Asymptotic Normality of $(\widetilde{\lambda}, \widetilde{\beta}, \widetilde{\alpha})$

The main theorem in this section establishes the root-*n* consistency of the semiparametric estimators. This result will be further used to prove the root-*n* consistency of the semiparametric estimators of the distributions F(x), G(x), and $H_k(x)$, $k = 1, 2, \dots, K$. To show the root-*n* consistency, we first give two lemmas. The first lemma shows that the Lagrange multiplier is close to zero when the parameter values are close to their true values at an appropriate rate. All proofs will be given in the section of proofs.

Lemma 2.2.1. Assume that $g(x;\theta)$ is first-order continuously differentiable in a neighborhood of the true value θ_0 , and g^3 and $||\partial g/\partial \theta||$ are bounded by some integrable function in a neighborhood of θ , where ||.|| is the Euclidean norm. Then, $\alpha = \alpha(\lambda, \beta) = O(n^{-1/3})$ (a.s.), uniformly about $\theta = (\lambda, \beta)$ in $B(\theta_0) = \{\theta; ||\theta - \theta_0|| \le n^{-1/3}\}$.

The following lemma will be used to prove the existence of an maximum semiparametric likelihood estimator.

Lemma 2.2.2. Under the assumptions in Lemma (2.2.1), on the surface of the ball $B(\theta_0) = \{\theta; ||\theta - \theta_0|| \le n^{-1/3}\}, \text{ with probability 1},$

1.
$$l_1(\lambda, \beta, \alpha(\lambda, \beta)) < l_1(\lambda_0, \beta_0, \alpha(\lambda_0, \beta_0)),$$

- 2. $l_2(\lambda,\beta) < l_2(\lambda_0,\beta_0)$, and thus
- 3. $l(\lambda, \beta, \alpha(\lambda, \beta)) < l(\lambda_0, \beta_0, \alpha(\lambda_0, \beta_0)),$

when n is large enough.

Remark 2.2.1. If we enlarge the parameter space and treat λ , β , and α jointly as independent parameters, a Taylor expansion of $l(\lambda, \beta, \alpha)$ on the surface of the ball $B(\theta_0, 0) = \{(\theta, \alpha); ||(\theta, \alpha) - (\theta_0, 0)||\}$ is

$$l(\lambda,\beta,\alpha) = l(\lambda_0,\beta_0,\alpha_0) + \sum_{k=1}^{K} O((n_k \log \log n_k)^{1/2}) u n^{-1/3} + \frac{1}{2} u^T e_{11} u n^{1/3} + o(n^{1/3}) \quad (a.s.),$$

where ||u|| = 1. Then, similar arguments can be made since e_{11} is negative definite.

Remark 2.2.2. The second part of Lemma (2.2.2) suggests that inference may be based on l_2 , a partial likelihood as studied in Zou, Fine and Yandell (2002), Zou and Fine (2002), and Qin and Zhang (2005). When this partial likelihood is used in estimation and inference, efficiency loss needs to be assessed via, for example, checking the relative efficiency of the partial likelihood-based estimator to the estimator under the full likelihood.

Now, we state the main results: the existence and strongly consistency of a local maximum likelihood estimator of $\theta_0 = (\lambda_0, \beta_0)$.

Theorem 2.2.3. In addition to the conditions of Lemma (2.2.1), we assume that

- 1. $g(x;\theta)$ is second-order continuously differentiable in an $O(n^{-1/3})$ neighborhood of θ_0 ,
- 2. as $n = \sum_{k=1}^{K} n_k \to \infty$, $\frac{n_k}{n} \to \rho_k > 0$, $k = 1, \cdots, K$, and
- 3. $0 < \lambda_k < 1$, $k = 1, \dots, K$, and at least two λ_k 's are not same.

Under these assumptions, with probability 1, $l(\lambda, \beta, \alpha(\lambda, \beta))$ has a local maximum in an $O(n^{-1/3})$ neighborhood of (α_0, β_0) . Moreover, the maximizer $(\widetilde{\lambda}, \widetilde{\beta})$ and $\widetilde{\alpha} = \alpha(\widetilde{\lambda}, \widetilde{\beta})$ satisfy the score equations (2.1.12) and the constraint equation (2.1.7) respectively, and as $n \to \infty$,

$$\sqrt{n} \begin{pmatrix} \widetilde{\lambda} - \lambda_0 \\ \widetilde{\beta} - \beta_0 \\ \widetilde{\alpha} - 0 \end{pmatrix} \to N(0, U), \qquad (2.2.1)$$
where U is given in the section of proofs.

Remark 2.2.3. The first part of the theorem (2.2.3) implies that the maximum semiparametric likelihood estimator is strongly consistent.

2.3 CDF Estimators and a Graphical Diagnostic

Based on the observed data, F, G, and $H_k(x)$ can be estimated by

$$\widetilde{F}(x) = \sum_{i=1}^{n} \widetilde{p}_{i} I(t_{i} \leq x), \quad \widetilde{G}(x) = \sum_{i=1}^{n} \widetilde{p}_{i} \omega(t_{i}; \widetilde{\beta}) I(t_{i} \leq x), \text{ and}$$
$$\widetilde{H}_{k}(x) = \sum_{i=1}^{n} \widetilde{p}_{i} [\widetilde{\lambda}_{k} + (1 - \widetilde{\lambda}_{k}) \omega(t_{i}; \widetilde{\beta})] I(t_{i} \leq x), \quad k = 1, \cdots, K,$$

respectively. We can prove that

$$\sum_{k=1}^{K} n_k [\tilde{H}_k(x) - \hat{H}_k(x)] = 0, \qquad (2.3.1)$$

where $\widehat{H}_k(x)$ is the empirical CDF of $H_k(x)$.

To seek asymptotic expressions for $\widetilde{F}(t)$, $\widetilde{G}(t)$, and $\widetilde{H}_k(t)$, first denote

$$F_{1}(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{r(t_{i};\lambda_{0},\beta_{0})} I(t_{i} \leq t),$$

$$F_{2}(t) = \int_{-\infty}^{t} \left(G(x;\lambda_{0},\beta_{0})\rho^{T}, -\frac{\phi}{R(x;\lambda_{0},\beta_{0})} \frac{\partial\omega(x;\beta)}{\partial\beta^{T}}, -G(x;\lambda_{0},\beta_{0}) \right) dF(x) \ S^{-1}Q_{n},$$

$$G_{1}(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{\omega(t_{i};\beta_{0})}{r(t_{i};\lambda_{0},\beta_{0})} I(t_{i} \leq t),$$

$$G_{2}(t) = \int_{-\infty}^{t} \left(G(x;\lambda_{0},\beta_{0})\rho^{T}, -\frac{\phi}{R(x;\lambda_{0},\beta_{0})} \frac{\partial\omega(x;\beta)}{\partial\beta^{T}}, -G(x;\lambda_{0},\beta_{0}) \right) dG(x) \ S^{-1}Q_{n},$$

$$H_{k1}(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{\omega_k(t_i; \lambda_0, \beta_0)}{r(t_i; \lambda_0, \beta_0)} I(t_i \le t),$$
$$H_{k2}(t) = \int_{-\infty}^{t} \left(G(x; \lambda_0, \beta_0) \rho^T, -\frac{\phi}{R(x; \lambda_0, \beta_0)} \frac{\partial \omega(x; \beta)}{\partial \beta^T}, -G(x; \lambda_0, \beta_0) \right) dH_k(x) \ S^{-1}Q_n,$$

where ϕ and ρ are defined in (1.7.1), and G and R in (1.7.2).

Theorem 2.3.1. Under regularity conditions,

$$\widetilde{F}(t) = F_1(t) - F_2(t) + o_p(n^{-1/2}),$$

$$\widetilde{G}(t) = G_1(t) - G_2(t) + o_p(n^{-1/2}),$$

$$\widetilde{H}_k(t) = H_{k1} - H_{k2} + o_p(n^{-1/2}), \text{ for } k = 1, 2, \cdots, K$$

Theorem 2.3.2. Under the condition as in Theorem (2.2.3), as $n \to \infty$, we have

1. $\sqrt{n} [\widetilde{F}(x) - F(x)] \to B_F(x),$ 2. $\sqrt{n} [\widetilde{G}(x) - G(x)] \to B_G(x), and$ 3. $\sqrt{n} [\widetilde{H}_k(x) - H_k(x)] \to B_{H_k}(x), for each k,$

where $B_F(x)$, $B_G(x)$, and $B_{H_k}(x)$ are mean zero Gaussian processes with continuous paths and respective covariance structures Σ_F , Σ_G , and Σ_{H_k} specified later in the proof section.

The covariance function, Σ_F , can be consistently estimated by $\widetilde{\Sigma}_F$ with consistent estimates in place of theoretical quantities in Σ_F . The resulting plug-in formulas are tedious and are omitted. A 95% confidence interval for F(x), based on normal theory, is $\widetilde{F}(x) \pm 1.96 n^{-1/2} \widetilde{\Sigma}_F$, with any upper limit exceeding 1 replaced by 1. Similar statements can be made for Σ_G and Σ_{H_k} .

To see whether or not $\widetilde{H}_k(x)$ is a good estimate for $H_k(x)$, a natural diagnostic is to plot $(\widehat{H}_k(x), \widetilde{H}_k(x))$, $k = 1, \dots, K$. Substantial difference in any one of the Kplots would indicate inadequacy of the density ratio model (1.1.1).

2.4 Proofs

In this section, we give detailed proofs of our results.

2.4.1 Proof of Lemmas (2.2.1) and (2.2.2)

The proof of Lemmas (2.2.1) is similar to that of Owen (1990). Suppose that θ satisfies $||\theta - \theta_0|| \leq n^{-1/3}$. Let $Z_n = \max_i |g(t_i; \theta)|$. Following Owen's method, we can show that, with probability 1, $Z_n = o(n^{1/3})$ under our conditions. Now

$$0 = \left| \frac{1}{n} \sum_{i=1}^{n} \frac{g(t_i; \theta)}{1 + \alpha g(t_i; \theta)} \right|$$

$$= \frac{1}{n} \left| \sum_{i=1}^{n} g(t_i; \theta) - \alpha \sum_{i=1}^{n} \frac{g^2(t_i; \theta)}{1 + \alpha g(t_i; \theta)} \right|$$

$$\geq \frac{|\alpha|}{n} \sum_{i=1}^{n} \frac{g^2(t_i; \theta)}{1 + \alpha g(t_i; \theta)} - \frac{1}{n} \left| \sum_{i=1}^{n} g(t_i; \theta) \right|$$

$$\geq \frac{|\alpha|S}{1 + |\alpha|Z_n} - \frac{1}{n} \left| \sum_{i=1}^{n} g(t_i; \theta) \right|,$$
(2.4.1)

where $S = \frac{1}{n} \sum_{i=1}^{n} g^2(t_i; \theta)$. By Taylor expansion,

$$\frac{1}{n}\sum_{i=1}^{n}g(t_{i};\theta) = \frac{1}{n}\sum_{i=1}^{n}g(t_{i};\theta_{0}) + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial g(t_{i};\theta_{*})}{\partial \theta}(\theta - \theta_{0}),$$
(2.4.2)

the last term being $O(n^{-1/3})$ by our conditions and the strong law of large numbers, the first term by (2.1.10) being negligible, and θ_* satisfying $\theta_* = c\theta_0 + (1-c)\theta$, for some $c \in (0, 1)$. It follows from (2.4.1) and (2.4.2) that

$$\frac{|\alpha|S}{1+|\alpha|Z_n} = O(n^{-1/3}).$$

Therefore, $|\alpha| = O(n^{-1/3})$ uniformly about θ in the $O(n^{-1/3})$ neighborhood of θ_0 , since $Z_n = o(n^{1/3})$ and S converges to a finite quantity almost surely. The proof of Lemma (2.2.1) is completed.

Now, let's prove Lemma (2.2.2). Proof of part 1: Following the lines as in Qin (1999), we can show, using Taylor expansion, the law of the iterated logarithm, and the result of $|\alpha| = O(n^{-1/3})$ in the ball $B(\theta_0)$, that $l_1(\lambda, \beta, \alpha(\lambda, \beta)) < l_1(\lambda_0, \beta_0, \alpha(\lambda_0, \beta_0))$, (a.s.) on the surface of $B(\theta_0)$. Proof of part 2: Following Qin (1993) and using the law of the iterated logarithm and the strong law of large numbers, on the surface of the ball $B(\theta_0) = \{\theta; ||\theta - \theta_0||\}$, we have, almost surely, that $l_2(\lambda,\beta)$ equals

$$\begin{split} l_{1}(\lambda_{0},\beta_{0}) &- \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} \frac{1}{r(x_{kj};\theta_{0})} \frac{\partial r(x_{kj};\theta_{0})}{\partial \theta^{T}} u n^{-1/3} + \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} \frac{1}{\omega_{k}(x_{kj};\theta_{0})} \frac{\partial \omega_{k}(x_{kj};\theta_{0})}{\partial \theta^{T}} u n^{-1/3} \\ &- \frac{1}{2} u^{T} \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} \left\{ -\frac{1}{r^{2}(x_{kj};\theta_{0})} \frac{\partial r(x_{kj};\theta_{0})}{\partial \theta} \frac{\partial r(x_{kj};\theta_{0})}{\partial \theta^{T}} + \frac{1}{r(x_{kj};\theta_{0})} \frac{\partial^{2} r(x_{kj};\theta_{0})}{\partial \theta^{T}} \right\} u n^{-1/3} \\ &- \frac{1}{2} u^{T} \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} \left\{ \frac{1}{\omega_{k}(x_{kj};\theta_{0})} \frac{\partial \omega_{k}(x_{kj};\theta_{0})}{\partial \theta} \frac{\partial \omega_{k}(x_{kj};\theta_{0})}{\partial \theta^{T}} - \frac{1}{\omega_{k}(x_{kj};\theta_{0})} \frac{\partial^{2} \omega_{k}(x_{kj};\theta_{0})}{\partial \theta^{T}} \right\} u n^{-1/3} \\ &= l_{2}(\lambda_{0},\beta_{0}) + \sum_{k=1}^{K} O((n_{k} \log \log n_{k})^{1/2}) u n^{-1/3} + \frac{1}{2} u^{T} e_{11} u n^{1/3} + o(n^{1/3}), \end{split}$$

where ||u|| = 1 and e_{11} is defined in (2.4.6). When using the law of the iterated logarithm, we note that the two means of the two first order terms are both zero. From the expression of the covariance matrix V in (2.4.7), we can see that $-e_{11} - \delta e_{12} e_{12}^T$ is positive definite. The negative definiteness of e_{11} follows since $\delta e_{12} e_{12}^T$ is positive definite. Thus, $l_2(\lambda, \beta) < l_2(\lambda_0, \beta_0)$, (a.s.) on the surface of $B(\theta_0)$. Proof of part 3: First we note that $l(\lambda, \beta, \alpha(\lambda, \beta)) = l_1(\lambda, \beta, \alpha(\lambda, \beta)) + l_2(\lambda, \beta)$, where l_1 and l_2 are given in (2.1.11). The result naturally follows.

2.4.2 Proof of Theorem (2.2.3)

The existence of a maximum semiparametric likelihood estimator in the $O(n^{-1/3})$ neighborhood of (λ_0, β_0) follows directly from part 3 of Lemma (2.2.2). To prove the normality, we start with the first and second order derivatives of the profile loglikelihood (2.1.1).

$$\frac{\partial l}{\partial \lambda}|_{\theta_0} = \rho_n \sum_{i=1}^n \frac{\omega(t_i; \beta_0) - 1}{r(t_i; \lambda_0, \beta_0)} - \sum_{k=1}^K \sum_{j=1}^{n_k} \frac{\omega(x_{kj}; \beta_0) - 1}{\omega_k(x_{kj}; \lambda_0, \beta_0)} e_k,$$

$$\frac{\partial^2 l}{\partial \lambda \partial \lambda^T}|_{\theta_0} = \rho_n \rho_n^T \sum_{i=1}^n \frac{[\omega(t_i; \beta_0) - 1]^2}{r^2(t_i; \lambda_0, \beta_0)} - \sum_{k=1}^K \sum_{j=1}^{n_k} \frac{[\omega(x_{kj}; \beta_0) - 1]^2}{\omega_k^2(x_{kj}; \lambda_0, \beta_0)} e_k e_k^T,$$

$$\frac{\partial^2 l}{\partial \lambda \partial \beta^T}|_{\theta_0} = \rho_n \sum_{i=1}^n \frac{1}{r^2(t_i; \lambda_0, \beta_0)} \frac{\partial \omega(t_i; \beta_0)}{\partial \beta^T} - \sum_{k=1}^K \sum_{j=1}^{n_k} \frac{1}{\omega_k^2(x_{kj}; \lambda_0, \beta_0)} e_k \frac{\partial \omega(x_{kj}; \beta_0)}{\partial \beta^T},$$

$$\frac{\partial l}{\partial \beta}|_{\theta_0} = -\xi \sum_{i=1}^n \frac{1}{r(t_i; \lambda_0, \beta_0)} \frac{\partial \omega(t_i; \beta_0)}{\partial \beta} + \sum_{k=1}^K \sum_{j=1}^{n_k} \frac{1 - \lambda_k}{\omega_k^2(x_{kj}; \lambda_0, \beta_0)} \frac{\partial \omega(x_{kj}; \beta_0)}{\partial \beta},$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta \partial \beta^T}|_{\theta_0} &= -\xi \sum_{i=1}^n \left[\frac{\partial^2 \omega(t_i; \beta_0)}{\partial \beta \partial \beta^T} \frac{1}{r(t_i; \lambda_0, \beta_0)} - \left\{ \frac{\partial \omega(t_i; \beta_0)}{\partial \beta} \right\}^{\otimes 2} \frac{\xi}{r^2(x_{kj}; \lambda_0, \beta_0)} \right] \\ &+ \sum_{k=1}^K \sum_{j=1}^{n_k} \left[\frac{\partial^2 \omega(x_{kj}; \beta_0)}{\partial \beta \partial \beta^T} \frac{1 - \lambda_k}{\omega_k(x_{kj}; \lambda_{k0}, \beta_0)} - \left\{ \frac{\partial \omega(x_{kj}; \beta_0)}{\partial \beta} \right\}^{\otimes 2} \frac{(1 - \lambda_k)^2}{\omega_k^2(x_{kj}; \lambda_{k0}, \beta_0)} \right],\end{aligned}$$

$$\frac{\partial l}{\partial \alpha}|_{\theta_0} = -\sum_{i=1}^n g(t_i; \lambda_0, \beta_0),$$

$$\frac{\partial^2 l}{\partial \alpha^2}|_{\theta_0} = \sum_{i=1}^n g^2(t_i; \lambda_0, \beta_0),$$

$$\frac{\partial^2 l}{\partial \alpha \partial \lambda} |_{\theta_0} = -\sum_{i=1}^n \frac{\partial g(t_i; \lambda_0, \beta_0)}{\partial \lambda^T} = -\rho_n \sum_{i=1}^n \frac{[\omega(t_i; \beta_0) - 1]^2}{r^2(t_i; \lambda_0, \beta_0)},$$
$$\frac{\partial^2 l}{\partial \alpha \partial \beta^T} |_{\theta_0} = -\sum_{i=1}^n \frac{\partial g(t_i; \lambda_0, \beta_0)}{\partial \beta^T} = -\sum_{i=1}^n \frac{1}{r^2(t_i; \lambda_0, \beta_0)} \frac{\partial \omega(t_i; \beta_0)}{\partial \beta^T}.$$

$$Q_{n} = \frac{1}{n} \frac{\partial l}{\partial \theta} \Big|_{\theta_{0}} = \begin{pmatrix} \frac{1}{n} \frac{\partial l}{\partial \lambda} \\ \frac{1}{n} \frac{\partial l}{\partial \beta} \\ \frac{1}{n} \frac{\partial l}{\partial \alpha} \end{pmatrix}_{\theta_{0}}, \qquad (2.4.3)$$

$$S_{n} = \frac{1}{n} \frac{\partial^{2} l}{\partial \theta \partial \theta^{T}} \Big|_{\theta_{0}} = \begin{pmatrix} \frac{1}{n} \frac{\partial^{2} l}{\partial \lambda \partial \lambda^{T}} & \frac{1}{n} \frac{\partial^{2} l}{\partial \lambda \partial \beta^{T}} & \frac{1}{n} \frac{\partial^{2} l}{\partial \lambda \partial \alpha} \\ \frac{1}{n} \frac{\partial^{2} l}{\partial \beta \partial \lambda^{T}} & \frac{1}{n} \frac{\partial^{2} l}{\partial \beta \partial \beta^{T}} & \frac{1}{n} \frac{\partial^{2} l}{\partial \beta \partial \alpha} \\ \frac{1}{n} \frac{\partial^{2} l}{\partial \alpha \partial \lambda^{T}} & \frac{1}{n} \frac{\partial^{2} l}{\partial \alpha \partial \beta^{T}} & \frac{1}{n} \frac{\partial 2^{2} l}{\partial \beta \partial \alpha^{2}} \end{pmatrix}_{\theta_{0}}. \qquad (2.4.4)$$

Denote the (i, j) element of S_n by s_{nij} , we have

$$\begin{split} s_{n11} &= \frac{1}{n} \rho_n \rho_n^T \sum_{k=1}^K \sum_{j=1}^{n_k} \frac{[\omega(t_i; \beta_0) - 1]^2}{r^2(t_i; \lambda_0, \beta_0)} - \frac{1}{n} \sum_{k=1}^K \sum_{j=1}^{n_k} \frac{[\omega(x_{kj}; \beta_0) - 1]^2}{\omega_k^2(x_{kj}; \lambda_0, \beta_0)} e_k e_k^T \\ &\to \rho \rho^T \sum_{k=1}^K \rho_k \int \frac{[\omega(t; \beta_0) - 1]^2}{R^2(t; \lambda_0, \beta_0)} \omega_k(x; \lambda_{k0}, \beta_0) dF - \sum_{k=1}^K \rho_k \frac{[\omega(t; \beta_0) - 1]^2}{\omega_k^2(x; \lambda_0, \beta_0)} e_k e_k^T dF \\ &= \int [1 - \omega(x; \beta_0)]^2 \left[\frac{\rho \rho^T}{R(x; \lambda_0, \beta_0)} - \sum_{k=1}^K \frac{\rho_k}{\omega_k(x; \lambda_0, \beta_0)} e_k e_k^T \right] dF \equiv s_{11}, \\ s_{n12} \to \int \left[\frac{\rho}{R(x; \lambda_0, \beta_0)} - \sum_{k=1}^K \frac{\rho_k}{\omega_k(x; \lambda_0, \beta_0)} e_k \right] \frac{\partial \omega(x; \beta_0)}{\partial \beta^T} dF \equiv s_{12}, \\ s_{n13} \to - \int \frac{[\omega(x; \beta_0) - 1]^2}{R(x; \lambda_0, \beta_0)} \rho dF \equiv s_{13}, \\ s_{n22} \to \int \left[\sum_{k=1}^K \frac{\rho_k \lambda_k(1 - \lambda_k)}{\omega_k(x; \lambda_0, \beta_0)} - \frac{\phi(1 - \phi)}{R(x; \lambda_0, \beta_0)} \right] \frac{\partial^2 \omega(x; \beta_0)}{\partial \beta \partial \beta^T} dF \equiv s_{22}, \\ s_{n23} \to - \int \frac{1}{R(x; \lambda_0, \beta_0)} \frac{\partial \omega(x; \beta_0)}{\partial \beta} dF \equiv s_{23}, \\ s_{n33} \to \int \frac{[\omega(x; \beta_0) - 1]^2}{R(x; \lambda_0, \beta_0)} dF \equiv s_{33}. \end{split}$$

Let

Write

$$Q_n = \begin{pmatrix} Q_{n1} \\ Q_{n2} \\ Q_{n3} \end{pmatrix},$$

where

$$\begin{aligned} Q_{n1} &= \frac{1}{n} \frac{\partial l}{\partial \lambda} |_{\theta_0} \\ &= \frac{1}{n} \left\{ \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{\left[\omega(x_{kj}; \beta_0) - 1 \right]}{r(t_i; \lambda, \beta) x_{kj}} \rho_n + \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{1 - \omega(x_{kj}; \beta_0)}{\omega_k(x_{kj}; \lambda_{k0}, \beta_0)} e_k \right\} \\ Q_{n2} &= \frac{1}{n} \frac{\partial l}{\partial \beta} |_{\theta_0} \\ &= \frac{1}{n} \left\{ -\sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{1}{r(x_{kj}; \lambda_0, \beta_0)} \xi \frac{\partial \omega(x_{kj}; \beta_0)}{\partial \beta} + \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{1 - \lambda_k}{\omega_k(x_{kj}; \lambda_{k0}, \beta_0)} \frac{\partial \omega(x_{kj}; \beta_0)}{\partial \beta} \right\} \\ Q_{n3} &= \frac{1}{n} \frac{\partial l}{\partial \alpha} |_{\theta_0} \\ &= \frac{1}{n} \left\{ -\sum_{k=1}^{K} \sum_{j=1}^{n_k} g(x_{kj}; \lambda_0, \beta_0) \right\} \end{aligned}$$

Rewrite Q_n as

$$Q_n = \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} q_k(x_{kj}),$$

where

$$q_k(x_{kj}) = -\begin{pmatrix} \frac{1-\omega(x_{kj};\beta_0)}{r(x_{kj};\lambda_0,\beta_0)}\rho_n - \frac{1-\omega(x_{kj};\beta_0)}{\omega(x_{kj};\lambda_{k0},\beta_0)}e_k\\ \frac{\xi}{r(x_{kj};\lambda_0,\beta_0)}\frac{\partial\omega(x_{kj};\beta_0)}{\partial\beta} - \frac{1-\lambda_k}{\omega_k(x_{kj};\lambda_{k0},\beta_0)}\frac{\partial\omega(x_{kj};\beta_0)}{\partial\beta}\\ g(x_{kj};\lambda_0,\beta_0) \end{pmatrix}, \qquad k = 1, 2, \cdots, K.$$

Write

$$Q_n = \rho_{n1}Q_n^{(1)} + \rho_{n2}Q_n^{(2)} + \dots + \rho_{nK}Q_n^{(K)}$$

where

$$Q_n^{(k)} = \frac{1}{n_k} \sum_{j=1}^{n_k} q_k(x_{kj}), \qquad k = 1, 2, \cdots, K.$$

By CLT, it holds true that

$$\sqrt{n_k}(Q_n^{(k)} - \mu^{(k)}) \longrightarrow N(0, V^{(k)}), \qquad k = 1, 2, \cdots, K,$$

where

$$\mu^{(k)} = E[q_k(X_k)] = \int q_k(x)\omega(x;\beta_0)dF, \quad \text{and} \quad V^{(k)} = var(q_k(X_k)).$$

It's easy to show that

$$\sqrt{n}(Q_n - \mu) \longrightarrow N(0, V),$$

where

$$\mu = \sum_{k=1}^{K} \rho_{nk} \mu^{(k)} = \int \sum_{k=1}^{K} \rho_{nk} q_k(x) \omega_k(x; \lambda_{k0}, \beta_0) dF = 0, \text{ and,}$$
$$V = \sum_{k=1}^{K} \rho_{nk} V^{(k)}.$$

Noting that $V^{(k)}$ is symmetric, write it as

$$\left(\begin{array}{cccc} v_{11}^{(k)} & v_{12}^{(k)} & v_{13}^{(k)} \\ v_{12}^{(k)^T} & v_{22}^{(k)} & v_{23}^{(k)} \\ v_{13}^{(k)^T} & v_{23}^{(k)^T} & v_{33}^{(k)} \end{array}\right),$$

where

$$\begin{aligned} v_{11}^{(k)} &= var \left\{ \frac{1 - \omega(X_k; \beta_0)}{r(X_k; \lambda_0, \beta_0)} \rho - \frac{1 - \omega(X_k; \beta_0)}{\omega_k(X_k; \lambda_{k0}, \beta_0)} e_k \right\} \\ &= E \left\{ \frac{1 - \omega(X_k; \beta_0)}{r(X_k; \lambda_0, \beta_0)} \rho - \frac{1 - \omega(X_k; \beta_0)}{\omega_k(X_k; \lambda_{k0}, \beta_0)} e_k \right\}^{\otimes 2} - \left\{ E \left[\frac{1 - \omega(X_k; \beta_0)}{r(X_k; \lambda_0, \beta_0)} \rho \right] \right\}^{\otimes 2} \\ &= \rho \rho_n^T E \left\{ \left[\frac{1 - \omega(X_k; \beta_0)}{r(X_k; \lambda_0, \beta_0)} \right]^2 \right\} - \rho e_k^T E \left\{ \frac{[1 - \omega(X_k; \beta_0)]^2}{r(X_k; \lambda_0, \beta_0) \omega_k(X_k; \lambda_{k0}, \beta_0)} \right\} - e_k \rho^T E \left\{ \frac{[1 - \omega(X_k; \beta_0)]^2}{r(X_k; \lambda_0, \beta_0) \omega_k(X_k; \lambda_{k0}, \beta_0)} \right\} + e_k e_k^T E \left\{ \left[\frac{1 - \omega(X_k; \beta_0)}{\omega_k(X_k; \lambda_{k0}, \beta_0)} \right]^2 \right\} - \rho_n \rho^T \left\{ E \left[\frac{1 - \omega(X_k; \beta_0)}{r(X_k; \lambda_0, \beta_0)} \right] \right\}^2. \end{aligned}$$

Similar expressions can be developed for other $v_{ij}^{\left(k\right)},\!\mathrm{s}.$

Note that V is symmetric. Write V as

$$\left(\begin{array}{cccc} v_{11} & v_{12} & v_{13} \\ v_{12}^T & v_{22} & v_{23} \\ v_{13}^T & v_{23}^T & v_{33} \end{array}\right), \quad \text{where}$$

$$v_{11} = \sum_{k=1}^{K} \rho_{nk} v_{11}^{(k)}$$

$$= \sum_{k=1}^{K} \rho_{k} \rho \rho^{T} E \left\{ \frac{[1 - \omega(X_{k}; \beta_{0})]^{2}}{r^{2}(X_{k}; \lambda_{0}, \beta_{0})} \right\} - \sum_{k=1}^{K} \rho_{k} \rho e_{k}^{T} E \left\{ \frac{[1 - \omega(X_{k}; \beta_{0})]^{2}}{r(X_{k}; \lambda_{0}, \beta_{0})\omega_{k}(X_{k}; \lambda_{k0}, \beta_{0})} \right\} - \sum_{k=1}^{K} \rho_{k} e_{k} \rho^{T} E \left\{ \frac{[1 - \omega(X_{k}; \beta_{0})]^{2}}{r(X_{k}; \lambda_{0}, \beta_{0})\omega_{k}(X_{k}; \lambda_{k0}, \beta_{0})} \right\} + \sum_{k=1}^{K} \rho_{k} e_{k} e_{k}^{T} E \left\{ \frac{[1 - \omega(X_{k}; \beta_{0})]^{2}}{\omega_{k}^{2}(X_{k}; \lambda_{k0}, \beta_{0})} \right\} - \sum_{k=1}^{K} \rho_{k} \rho \rho_{n}^{T} \left\{ E \left[\frac{1 - \omega(X_{k}; \beta_{0})}{r(X_{k}; \lambda_{0}, \beta_{0})} \right] \right\}^{2}$$

$$= \rho \rho^{T} \int \frac{[1 - \omega(x; \beta_{0})]^{2}}{r(x; \lambda_{0}, \beta_{0})} dF - \rho \rho^{T} \int \frac{[1 - \omega(x; \beta_{0})]^{2}}{r(x; \lambda_{0}, \beta_{0})} dF - \rho \rho^{T} \int \frac{[1 - \omega(x; \beta_{0})]^{2}}{r(x; \lambda_{0}, \beta_{0})} dF + \sum_{k=1}^{K} \rho_{k} e_{k} e_{k}^{T} \int \frac{[1 - \omega(x; \beta_{0})]^{2}}{\omega_{k}(x; \lambda_{k0}, \beta_{0})} dF - \sum_{k=1}^{K} \rho_{nk} \rho \rho^{T} \left[\int \frac{1 - \omega(x; \beta_{0})}{r(X_{k}; \lambda_{0}, \beta_{0})} \omega_{k}(x; \lambda_{k0}, \beta_{0}) dF \right]^{2} = -\rho \rho^{T} \int \frac{[1 - \omega(x; \beta_{0})]^{2}}{r(x; \lambda_{0}, \beta_{0})} dF + \sum_{k=1}^{K} \rho_{nk} e_{k} e_{k}^{T} \int \frac{[1 - \omega(x; \beta_{0})]^{2}}{\omega_{k}(x; \lambda_{k0}, \beta_{0})} dF - \sum_{k=1}^{K} \rho_{k} \rho \rho^{T} \left[\int \frac{1 - \omega(x; \beta_{0})}{r(x; \lambda_{0}, \beta_{0})} \omega_{k}(x; \lambda_{k0}, \beta_{0}) dF \right]^{2}.$$

Similarly for other v_{ij} 's.

$$\begin{split} v_{12} &= \xi \rho \int \frac{1 - \omega(x; \beta_0)}{r(x; \lambda_0, \beta_0)} \frac{\partial \omega(x; \beta_0)}{\partial \beta^T} dF \\ &+ \sum_{k=1}^K \rho_k (1 - \lambda_k) e_k \int \frac{1 - \omega(x; \beta_0)}{\omega_k(x; \lambda_{k0}, \beta_0)} \frac{\partial \omega(x; \beta_0)}{\partial \beta^T} dF \\ &- \xi \rho \sum_{k=1}^K \rho_k \int \frac{1 - \omega(x; \beta_0)}{r(x; \lambda_0, \beta_0)} \omega_k(x; \lambda_{k0}, \beta_0) dF \int \frac{\partial \omega(x; \beta_0)}{\partial \beta^T} \frac{\omega_k(x; \lambda_{k0}, \beta_0)}{r(x; \lambda_0, \beta_0)} dF, \\ v_{13} &= \sum_{k=1}^K \rho_k \rho \left[\int g(x; \lambda_0, \beta_0) \omega_k(x; \lambda_{k0}, \beta_0) F \right]^2, \\ v_{22} &= -\xi^2 \int \left\{ \frac{\partial \omega(x; \beta_0)}{\partial \beta} \right\}^{\otimes 2} \frac{1}{r(x; \lambda_0, \beta_0)} dF + \sum_{k=1}^K \int \left\{ \frac{\partial \omega(x; \beta_0)}{\partial \beta} \right\}^{\otimes 2} \frac{\rho_k(1 - \lambda_k)^2}{\omega_k(x; \lambda_{k0}, \beta_0)} dF \\ &- \sum_{k=1}^K \rho_k \int \frac{\partial \omega(x; \beta_0)}{\partial \beta} \omega_k(x; \lambda_{k0}, \beta_0) dF \int \frac{\partial \omega(x; \beta_0)}{\partial \beta^T} \frac{1}{r(x; \lambda_0, \beta_0)} \omega_k(x; \lambda_{k0}, \beta_0) dF, \\ v_{23} &= -\xi \sum_{k=1}^K \rho_k \int \frac{\partial \omega(x; \beta_0)}{\partial \beta} \frac{1}{r(x; \lambda_0, \beta_0)} dF \int g(x; \lambda_0, \beta_0) \omega_k(x; \lambda_{k0}, \beta_0) dF, \\ v_{33} &= \int r(x; \lambda_0, \beta_0) g^2(x; \lambda_0, \beta_0) dF - \sum_{k=1}^K \rho_k \left[g(x; \lambda_0, \beta_0) \omega_k(x; \lambda_{k0}, \beta_0) dF \right]^2. \end{split}$$

Following the arguments in Qin (1993, 1999), the asymptotic normality (2.2.3) of

 $(\widetilde{\lambda},\widetilde{\beta},\widetilde{\alpha})$ is shown by noting that

$$\begin{pmatrix} \widetilde{\lambda} - \lambda_0 \\ \widetilde{\beta} - \beta_0 \\ \widetilde{\alpha} - 0 \end{pmatrix} = -S_n^{-1}Q_n + o_p(n^{-1/2})$$
(2.4.5)

along with the Slutsky's theorem, where S_n and Q_n are given in (2.4.4) and (2.4.3). It's easy to see that the asymptotic covariance matrix U in (2.2.3) is $S^{-1}VS^{-1}$. In the following lines, we show that U can be expressed block diagonally. We first show the following lemmas (Special cases for which $\lambda_k = 0$ or 1 will be used in various situations).

Lemma 2.4.1.
$$\int_{-\infty}^{s} G(x; \lambda_0, \beta_0) \omega_k(x; \lambda_0, \beta_0) dF = G(s) - F(s) + (1 - \lambda_k - \phi)\tau(s).$$

Lemma 2.4.2.

$$\int_{-\infty}^{s} \frac{\omega_k(x;\lambda_0,\beta_0)\omega_l(x;\lambda_0,\beta_0)}{R(x;\lambda_0,\beta_0)} dF = G(s) + (1-\lambda_k-\lambda_l-\xi)[G(s)-F(s)] + (1-\lambda_l-\phi)(1-\lambda_k-\phi)\tau(s).$$

Lemma 2.4.3. $\int_{-\infty}^{s} \frac{\omega_k(x;\lambda_0,\beta_0)}{R(x;\lambda_0,\beta_0)} \frac{\partial \omega(x;\beta_0))}{\partial \beta} dF(x) = \frac{1-\phi-\lambda_k}{\phi} \varsigma_1(s).$

Proof of Lemma (2.4.1). Note that

$$R(x;\lambda_0,\beta_0) = 1 + \phi[\omega(x;\beta_0) - 1],$$
$$\omega_k(x;\lambda_{k0},\beta_0) = \lambda_k + (1 - \lambda_k)(\omega(x;\beta_0) - 1),$$
$$\int_{-\infty}^s [\omega(x;\beta_0) - 1]dF = G(s) - F(s),$$
$$G(x;\lambda_0,\beta_0) = \frac{\omega(x;\beta_0) - 1}{R(x;\lambda_0,\beta_0)}.$$

Now,

$$\begin{split} &\int_{-\infty}^{s} G(x;\lambda_{0},\beta_{0})\omega_{k}(x;\lambda_{0},\beta_{0})dF \\ &= \int_{-\infty}^{s} \frac{\omega(x;\beta_{0}) - 1}{R(x;\lambda_{0},\beta_{0})} [1 + (1 - \lambda_{k})(\omega(x;\beta_{0}) - 1)]dF \\ &= \int_{-\infty}^{s} \frac{\omega(x;\beta_{0}) - 1}{R(x;\lambda_{0},\beta_{0})} \Big[r(x;\lambda_{0},\beta_{0}) - \phi(\omega(x;\beta_{0}) - 1) \Big] dF + (1 - \lambda_{k}) \int_{-\infty}^{s} \frac{(\omega(x;\beta_{0}) - 1)^{2}}{R(x;\lambda_{0},\beta_{0})} dF \\ &= G(s) - F(s) + (1 - \lambda_{k} - \phi) \int_{-\infty}^{s} \frac{[\omega(x;\beta_{0}) - 1]^{2}}{R(x;\lambda_{0},\beta_{0})} dF. \end{split}$$

The proof of other Lemmas is similar and thus omitted. By Lemma 2.4.2,

$$\begin{split} v_{11} &= -\rho\rho^T \int \frac{[\omega(x;\beta_0)-1]^2}{R(x;\lambda_0,\beta_0)} dF + \sum_{k=1}^K \rho_k e_k e_k^T \int \frac{[\omega(x;\beta_0)-1]^2}{\omega_k(x;\lambda_0,\beta_0)} dF \\ &\quad -\sum_{k=1}^K \rho_k \rho\rho^T (1-\lambda_k-\phi) \left[\int \frac{[\omega(x;\beta_0)-1]^2}{R(x;\lambda_0,\beta_0)} dF \right]^2, \\ v_{12} &= -\rho \int \frac{\partial \omega(x;\beta_0)}{\partial\beta^T} \frac{1}{R(x;\lambda_0,\beta_0)} dF + \sum_{k=1}^K \rho_k e_k \int \frac{\partial \omega(x;\beta_0)}{\partial\beta^T} \frac{1}{\omega_k(x;\lambda_0,\beta_0)} dF \\ &\quad -\rho \sum_{k=1}^K \rho_k (1-\lambda_k-\phi)^2 \int \frac{[\omega(x;\beta_0)-1]^2}{R(x;\lambda_0,\beta_0)} dF \int \frac{\partial \omega(x;\beta_0)}{\partial\beta^T} \frac{1}{R(x;\lambda_0,\beta_0)} dF, \end{split}$$

$$\begin{split} v_{13} &= \sum_{k=1}^{K} \rho_k \rho \left[\int G(x; \lambda_0, \beta_0) \omega_k(x; \lambda_0, \beta_0) dF \right]^2 \\ &= \sum_{k=1}^{K} \rho_k \rho (1 - \lambda_k - \phi)^2 \left[\int \frac{[\omega(x; \beta_0) - 1]^2}{R(x; \lambda_0, \beta_0)} dF \right], \\ v_{22} &= -\phi^2 \int \frac{\partial \omega(x; \beta_0)}{\partial \beta} \frac{\partial \omega(x; \beta_0)}{\partial \beta^T} \frac{1}{R(x; \lambda_0, \beta_0)} dF \\ &+ \sum_{k=1}^{K} \rho_k (1 - \lambda_k) \int \frac{\partial \omega(x; \beta_0)}{\partial \beta} \frac{\partial \omega(x; \beta_0)}{\partial \beta^T} \frac{1}{\omega_k(x; \lambda_0, \beta_0)} dF \\ &- \sum_{k=1}^{K} \rho_k (1 - \lambda_k - \phi)^2 \left\{ \int \frac{\partial \omega(x; \beta_0)}{\partial \beta} \frac{1}{R(x; \lambda_0, \beta_0)} dF (1 - \lambda_k - \phi) \int \frac{[\omega(x; \beta_0) - 1]^2}{R(x; \lambda_0, \beta_0)} dF \right\} \\ v_{23} &= -\phi \sum_{k=1}^{K} \rho_k \int \frac{\partial \omega(x; \beta_0)}{\partial \beta} \frac{\omega_k(x; \lambda_0, \beta_0)}{R(x; \lambda_0, \beta_0)} dF (1 - \lambda_k - \phi) \int \frac{[\omega(x; \beta_0) - 1]^2}{R(x; \lambda_0, \beta_0)} dF \\ &= \sum_{k=1}^{K} \rho_k (1 - \lambda_k - \phi)^2 \int \frac{\partial \omega(x; \beta_0)}{\partial \beta} \frac{1}{R(x; \lambda_0, \beta_0)} dF \int \frac{[\omega(x; \beta_0) - 1]^2}{R(x; \lambda_0, \beta_0)} dF, \end{split}$$

$$v_{33} = \int \frac{[\omega(x;\beta_0) - 1]^2}{R(x;\lambda_0,\beta_0)} dF - \sum_{k=1}^K \rho_k (\lambda_k + \phi - 1)^2 \left[\int \frac{[\omega(x;\beta_0) - 1]^2}{R(x;\lambda_0,\beta_0)} dF \right]^2.$$

Write

$$S = \left(\begin{array}{cc} e_{11} & e_{12} \\ \\ e_{12}^T & e_{22} \end{array}\right)$$

where

$$e_{11} = \begin{pmatrix} s_{11} & s_{12} \\ s_{12}^T & s_{22} \end{pmatrix}, \quad e_{12} = \begin{pmatrix} s_{13} \\ s_{23} \end{pmatrix}, \quad e_{22} = s_{33}.$$
(2.4.6)

Write

$$V = \left(\begin{array}{cc} u_{11} & u_{12} \\ \\ u_{12}^T & u_{22} \end{array}\right)$$

where

$$u_{11} = \left(\begin{array}{cc} a & b \\ \\ c & d \end{array}\right)$$

with

$$a = -\rho\rho^{T}\tau + diag(\rho)diag(\eta) - \rho\rho^{T}\sum_{k=1}^{K}\rho_{k}(1-\lambda_{k}-\phi)^{2}\tau^{2},$$

$$b = -\rho\varsigma_{1}^{T} + rbind(\rho_{k}\psi_{1k}^{T}, \ k = 1, 2, \cdots, K) - \rho\sum_{k=1}^{K}\rho_{k}(1-\lambda_{k}-\phi)^{2}\tau\varsigma_{1},$$

$$c = -\varsigma_{1}\rho^{T} + cbind(\rho_{k}\psi_{1k}, \ k = 1, 2, \cdots, K) - \varsigma_{1}^{T}\tau\sum_{k=1}^{K}\rho_{k}(1-\lambda_{k}-\phi)^{2}\rho^{T},$$

$$d = -\phi^{2}\varsigma_{2} + \sum_{k=1}^{K}\rho_{k}(1-\lambda_{k})^{2}\psi_{2k} - \sum_{k=1}^{K}\rho_{k}(1-\lambda_{k}-\phi)^{2}\varsigma_{1}\varsigma_{1}^{T},$$

$$u_{12} = \begin{pmatrix} \rho \sum_{k=1}^{K} \rho_k (1 - \lambda_k - \phi)^2 \tau^2 \\ \sum_{k=1}^{K} \rho_k (1 - \lambda_k - \phi)^2 \varsigma_1 \tau \end{pmatrix}, \quad u_{22} = \tau - \sum_{k=1}^{K} \rho_k (1 - \lambda_k - \phi)^2 \tau^2.$$

It's easy to see that

$$V = \begin{pmatrix} -e_{11} - \delta e_{12} e_{12}^T & -\delta e_{12} s_{33} \\ -\delta e_{12}^T s_{33} & s_{33} - \delta s_{33}^2 \end{pmatrix}, \qquad (2.4.7)$$

where $\delta = \sum_{k=1}^{K} \rho_{nk} (1 - \lambda_k - \xi)^2$. The inverse of the matrix S is

$$S^{-1} = \begin{pmatrix} s_D^{-1} & 0 \\ 0 & s_{33}^{-1} e_{12}^T s_D^{-1} e_{12} \end{pmatrix} + \begin{pmatrix} 0 & -s_D^{-1} e_{12} s_{33}^{-1} \\ -e_{12}^T s_D^{-1} s_{33}^{-1} & s_{33}^{-1} \end{pmatrix}$$

where $s_D = e_{11} - e_{12} s_{33}^{-1} e_{12}^T$. Tedious matrix multiplication gives

$$S^{-1}V = \begin{pmatrix} -s_D^{-1}e_{11} & -s_D^{-1}e_{12} \\ e_{12}^Ts_D^{-1}e_{11}s_{33}^{-1} - \delta e_{12}^T & e_{12}^Ts_D^{-1}e_{12}s_{33}^{-1} + 1 - s_{33} \end{pmatrix},$$
$$U = S^{-1}VS^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \sigma_{22} \end{pmatrix}$$

with

$$\begin{split} \Sigma_{11} &= -s_D^{-1} e_{11} s_D^{-1} + s_D^{-1} e_{12} s_{33}^{-1} e_{12}^T s_D^{-1}, \\ \Sigma_{12} &= -s_D^{-1} e_{12} s_{33}^{-1} e_{12}^T s_D^{-1} e_{12} s_{33}^{-1} + s_D^{-1} e_{11} s_D^{-1} e_{12} s_{33}^{-1} - s_D^{-1} e_{12} s_{33}^{-1}, \\ \sigma_{22} &= e_{12}^T s_D^{-1} e_{12} s_{33}^{-1} e_{12}^T s_D^{-1} e_{12} s_{33}^{-2} + e_{12}^T s_D^{-1} e_{12} s_{33}^{-2} - e_{12}^T s_D^{-1} e_{11} s_D^{-1} e_{12} s_{33}^{-2} \\ &+ e_{12}^T s_D^{-1} e_{12} s_{33}^{-1} + s_{33}^{-1} - \delta. \end{split}$$

These can be further simplified as

$$\Sigma_{11} = -s_D^{-1}(e_{11} - e_{12}s_{33}^{-1}e_{12}^T)s_D^{-1} = -s_D^{-1}s_Ds_D^{-1} = -s_D^{-1},$$

$$\Sigma_{12} = s_D^{-1} (-e_{12} s_{33}^{-1} e_{12}^T + e_{11} - s_D) s_D^{-1} e_{12} s_{33}^{-1} = s_D^{-1} (s_D - s_D) s_D^{-1} e_{12} s_{33}^{-1} = 0,$$

$$\sigma_{22} = e_{12}^T s_D^{-1} (e_{12} s_{33}^{-1} e_{12}^T + s_D - e_{11}) s_D^{-1} e_{12} s_{33}^{-2} + s_D^{-1} e_{12}^T s_D^{-1} e_{12} s_{33}^{-1} + s_{33}^{-1} - \delta$$
$$= s_{33}^{-1} e_{12}^T s_D^{-1} e_{12} s_{33}^{-1} + s_{33}^{-1} - \delta.$$

Therefore,

$$U = \begin{pmatrix} \Sigma_{11} & 0\\ 0 & \sigma_{22} \end{pmatrix}, \qquad (2.4.8)$$

where

$$\Sigma_{11} = -s_D^{-1} = -e_{11}^{-1} - e_{11}^{-1}e_{12}(s_{33} - e_{12}^T e_{11}^{-1}e_{12})^{-1}e_{12}^T e_{11}^{-1},$$

$$\sigma_{22} = -s_{33}^{-1}e_{12}^T \Sigma_{11}e_{12}s_{33}^{-1} + s_{33}^{-1} - \delta;$$

that is, $(\widetilde{\lambda}, \widetilde{\beta})$ and $\widetilde{\alpha}$ are asymptotically independent.

2.4.3 Proof of Theorem (2.3.1)

Denote

$$\gamma(t;\theta) = \frac{1}{r(t;\lambda,\beta) + \alpha[\omega(t;\beta) - 1]}, \quad \Gamma(t;\theta) = \frac{1}{R(t;\lambda,\beta) + \alpha[\omega(t;\beta) - 1]},$$

where $\theta = (\lambda, \beta, \alpha)$ and denote $\theta_0 = (\lambda_0, \beta_0, 0)$. To show that $\widetilde{F}(t)$ has the claimed representation, we first expand it into a Taylor series. Using the established large sample result, we can show that the remaining term in the Taylor series are negligible in probability. We can write $\widetilde{F}(t) = \frac{1}{n} \sum_{i=1}^{n} \gamma(t_i; \widetilde{\theta}) I(t_i \leq t)$. Expanding $\widetilde{F}(t)$ at θ_0 yields

$$\widetilde{F}(t) = \frac{1}{n} \sum_{i=1}^{n} \gamma(t_i; \widetilde{\theta}) I(t_i \le t)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \gamma(t_i; \theta_0) I(t_i \le t) + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \gamma}{\partial \widetilde{\theta}} (t_i; \theta_0) I(t_i \le t) (\widetilde{\theta} - \theta_0) + R_{1n}(t)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{r(t_i; \lambda_0, \beta_0)} I(t_i \le t) - r_1(t)^T S^{-1} Q_n + R_{2n}(t),$$

where $R_{1n}(t)$, i = 1, 2, satisfy $\sup_{t \in (-\infty, +\infty)} |R_{in}(t)| = o_p(n^{-1/2})$, and as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\gamma}{\partial\widetilde{\theta}}(t_i;\theta_0)I(t_i\leq t)\to r_1(t),$$

where

$$r_{1}(t) = \int \left[\sum_{k=1}^{K} \rho_{k} \omega_{k}(u; \lambda_{0}, \beta_{0}) \right] \frac{\partial \Gamma}{\partial \tilde{\theta}}(u; \theta_{0}) I(u \leq t) dF(u), \text{ a.s.}$$

$$= \int R(u; \lambda_{0}, \beta_{0}) \frac{\partial \gamma}{\partial \tilde{\theta}}(u; \theta_{0}) I(u \leq t) dF(u)$$

$$= \int_{-\infty}^{t} R(-\frac{1}{R^{2}} \frac{\partial R}{\partial \lambda^{T}}, -\frac{1}{R^{2}} \frac{\partial R}{\partial \beta^{T}}, -\frac{\omega - 1}{R^{2}}) dF$$

$$= \int_{-\infty}^{t} (G\rho^{T}, -\frac{\phi}{R} \frac{\partial \omega}{\partial \beta^{T}}, -G)^{T} dF$$

$$= \begin{pmatrix} [G(t) - F(t) - \phi\tau(t)]\rho \\ -\phi\varsigma_{1}(t) \\ -G(t) + F(t) + \phi\tau(t) \end{pmatrix}.$$

Similarly, we have

$$\begin{split} \widetilde{G}(t) &= \frac{1}{n} \sum_{i=1}^{n} \gamma(t_i; \widetilde{\theta}) \omega(t_i; \widetilde{\beta}) I(t_i \leq t) \\ &= \frac{1}{n} \sum_{i=1}^{n} \gamma(t_i; \theta_0) \omega(t_i; \beta_0) I(t_i \leq t) \\ &+ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial \gamma(t_i; \theta_0)}{\partial \widetilde{\theta}} \omega(t_i; \beta_0) + \frac{\partial \omega(t_i; \beta_0)}{\partial \widetilde{\theta}} \gamma(t_i; \theta_0) \right] I(t_i \leq t) (\widetilde{\theta} - \theta_0) + R_{1n}(t) \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{\omega(t_i; \beta_0)}{r(t_i; \lambda_0, \beta_0)} I(t_i \leq t) - u_1^T(t) S^{-1} Q_n + R_{2n}(t), \end{split}$$

where $R_{1n}(t)$, i = 1, 2, satisfy $\sup_{t \in (-\infty, +\infty)} |R_{in}(t)| = o_p(n^{-1/2})$, and as $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial \gamma(t_i; \theta_0)}{\partial \widetilde{\theta}} \omega(t_i; \beta_0) + \frac{\partial \omega(t_i; \beta_0)}{\partial \widetilde{\theta}} \gamma(t_i; \theta_0) \right] I(t_i \le t) \to u_1(t), \quad \text{where}$$

$$\begin{split} u_{1}(t) &= \int \left[\sum_{k=1}^{K} \rho_{k} \omega_{k}(u;\lambda_{0},\beta_{0}) \right] \left[\frac{\partial \gamma(u;\theta_{0})}{\partial \tilde{\theta}} \omega(u;\beta_{0}) + \frac{\partial \omega(u;\beta_{0})}{\partial \tilde{\theta}} \frac{1}{r(u;\lambda_{0},\beta_{0})} \right) \right] I(u \leq t) dF(u) \\ &= \int R(u;\lambda_{0},\beta_{0}) \left[\frac{\partial \gamma(u;\theta_{0})}{\partial \tilde{\theta}} \omega(u;\beta_{0}) + \frac{\partial \omega(u;\beta_{0})}{\partial \tilde{\theta}} \frac{1}{r(u;\lambda_{0},\beta_{0})} \right) \right] I(u \leq t) dF(u) \\ &= \int_{-\infty}^{t} \left[R(-\frac{1}{R^{2}} \frac{\partial R}{\partial \lambda^{T}}, -\frac{1}{R^{2}} \frac{\partial R}{\partial \beta^{T}}, -\frac{\omega-1}{R^{2}} \right)^{T} \omega + (0, \frac{\partial \omega}{\partial \beta^{T}}, 0)^{T} \right] dF \\ &= \int_{-\infty}^{t} (G\rho^{T}, -\frac{\phi}{R} \frac{\partial \omega}{\partial \beta^{T}}, -G)^{T} dG \\ &= \begin{pmatrix} [G(t) - F(t) + (1-\phi)\tau(t)]\rho \\ (1-\phi)\varsigma_{1}(t) \\ -G(t) + F(t) - (1-\phi)\tau(t) \end{pmatrix}. \end{split}$$

Finally, we write

$$\begin{aligned} \widetilde{H}_{k}(t) &= \frac{1}{n} \sum_{i=1}^{n} \gamma(t_{i}; \widetilde{\theta}) \omega_{k}(t_{i}; \widetilde{\lambda}, \widetilde{\beta}) I(t_{i} \leq t) \\ &= \frac{1}{n} \sum_{i=1}^{n} \gamma(t_{i}; \theta_{0}) \omega_{k}(t_{i}; \lambda_{0}, \beta_{0}) I(t_{i} \leq t) \\ &+ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial \gamma(t_{i}; \theta_{0})}{\partial \widetilde{\theta}} \omega_{k}(t_{i}; \lambda_{0}, \beta_{0}) + \frac{\partial \omega_{k}(t_{i}; \lambda_{0}, \beta_{0})}{\partial \widetilde{\theta}} \gamma(t_{i}; \theta_{0}) \right] I(t_{i} \leq t) (\widetilde{\theta} - \theta_{0}) + R_{1n}(t) \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{\omega_{k}(t_{i}; \lambda_{0}, \beta_{0})}{r(t_{i}; \lambda_{0}, \beta_{0})} I(t_{i} \leq t) - v_{1k}^{T}(t) S^{-1}Q_{n} + R_{2n}(t), \end{aligned}$$

$$(2.4.9)$$

where $R_{in}(t)$, i = 1, 2, satisfy $\sup_{t \in (-\infty, +\infty)} |R_{in}(t)| = o_p(n^{-1/2})$, and as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial\gamma(t_{i};\theta_{0})}{\partial\widetilde{\theta}}\omega_{k}(t_{i};\lambda_{0},\beta_{0})+\frac{\partial\omega_{k}(t_{i};\lambda_{0},\beta_{0})}{\partial\widetilde{\theta}}\gamma(t_{i};\theta_{0})\right]I(t_{i}\leq t)\rightarrow v_{1k}(t), \text{ where,}$$

$$\begin{split} v_{1k}(t) &= \int_{-\infty}^{t} \left[\sum_{l=1}^{K} \rho_{l} \omega_{l}(u; \lambda_{0}, \beta_{0}) \right] \left[\frac{\partial \gamma(u; \theta_{0})}{\partial \tilde{\theta}} \omega_{k}(t) + \frac{\partial \omega_{k}(t)}{\partial \tilde{\theta}} \frac{1}{r(u; \lambda_{0}, \beta_{0})} \right) \right] dF(u) \\ &= \int_{-\infty}^{t} R \left(-\frac{\omega_{k}}{R^{2}} \frac{\partial R}{\partial \lambda^{T}} + \frac{1-\omega}{R} e_{k}, -\frac{\omega_{k}}{R^{2}} \frac{\partial R}{\partial \beta^{T}} + \frac{1}{R} \frac{\partial \omega_{k}}{\partial \beta^{T}}, -\frac{\omega_{k}(\omega-1)}{R^{2}} \right)^{T} dF \\ &= \int_{-\infty}^{t} \left(-\frac{\omega_{k}}{R} \frac{\partial R}{\partial \lambda^{T}}, -\frac{\omega_{k}}{R} \frac{\partial R}{\partial \beta^{T}}, -\frac{\omega_{k}(\omega-1)}{R} \right)^{T} dF \\ &= \int_{-\infty}^{t} \left(G\rho^{T}, -\frac{\phi}{R} \frac{\partial \omega}{\partial \beta^{T}}, -G \right)^{T} dH_{k} \\ &= \left(\begin{bmatrix} G(t) - F(t) + (1-\lambda_{k}-\phi)\tau(t) \end{bmatrix}^{P} \\ (1-\lambda_{k}-\phi)\varsigma_{1}(t) \\ -G(t) + F(t) - (1-\lambda_{k}-\phi)\tau(t) \end{bmatrix} \right). \end{split}$$

Proof of Theorem (2.3.2).

Let $r_2(t) = -r_1(t)^T S^{-1}$ and let

$$\varepsilon_k^F(x_{kj};t) = \frac{I(x_{kj} \le t)}{r(x_{kj};\lambda_0,\beta_0)} - \int_{-\infty}^t \frac{\omega_k(x;\lambda_0,\beta_0)}{R(x;\lambda_0,\beta_0)} dF(x),$$

then

$$\begin{aligned} \widetilde{F}(t) &= \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{1}{r(x_{kj};\lambda_0,\beta_0)} I(x_{kj} \le t) + r_2(t)Q_n + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \left[\varepsilon_k^F(x_{kj};t) + \int_{-\infty}^t \frac{\omega_k(x;\lambda_0,\beta_0)}{R(x;\lambda_0,\beta_0)} dF(x) \right] + r_2(t)Q_n + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \left[\varepsilon_k^F(x_{kj};t) + r_2(t)q_k(x_{kj}) \right] + \frac{1}{n} \int_{-\infty}^t \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{\omega_k(x;\lambda_0,\beta_0)}{R(x;\lambda_0,\beta_0)} dF(x) + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \left[\varepsilon_k^F(x_{kj};t) + r_2(t)q_k(x_{kj}) \right] + F(t) + o_p(n^{-1/2}). \end{aligned}$$

The last equality is due to the fact that

$$\sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{\omega_k(x;\lambda_0,\beta_0)}{R(x;\lambda_0,\beta_0)} = n \sum_{k=1}^{K} \rho_{nk} \omega_k(x;\lambda_0,\beta_0) \frac{1}{R(x;\lambda_0,\beta_0)} = n.$$

By the criteria of Billingsley (1968, p. 128), we can show that $\sqrt{n}(\tilde{F}(t) - F(t) \rightarrow B(t)$ weakly in $D[-\infty, +\infty]$, where B(t) is a mean zero Gaussian process with continuous paths and covariance structure

$$\Sigma(s,t) = \sum_{k=1}^{K} \rho_k \text{cov}(\varepsilon_k^F(X_k;s) + r_2(s)q_k(X_k), \varepsilon_k^F(X_k;t) + r_2(t)q_k(X_k)), \ s \le t.$$

Let $u_2(t) = -u_1(t)^T S^{-1}$ and let

$$\varepsilon_k^G(x_{kj};t) = \frac{\omega(x_{kj};\beta_0)}{r(x_{kj};\lambda_0,\beta_0)} I(x_{kj} \le t) - \int_{-\infty}^t \frac{\omega_k(x;\lambda_0,\beta_0)}{R(x;\lambda_0,\beta_0)} \omega(x;\beta_0) dF(x),$$

then

$$\begin{split} \widetilde{G}(t) &= \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{\omega(x_{kj}; \beta_0)}{r(x_{kj}; \lambda_0, \beta_0)} I(x_{kj} \le t) + u_2(t) Q_n + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \left[\varepsilon_k^G(x_{kj}; t) + \int_{-\infty}^t \frac{\omega_k(x; \lambda_0, \beta_0)}{R(x; \lambda_0, \beta_0)} \omega(x; \beta_0) dF(x) \right] + u_2(t) Q_n + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \left[\varepsilon_k^G(x_{kj}; t) + u_2(t) q_k(x_{kj}) \right] + \frac{1}{n} \int_{-\infty}^t \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{\omega_k(x; \lambda_0, \beta_0)}{R(x; \lambda_0, \beta_0)} dG(x) + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \left[\varepsilon_k^G(x_{kj}; t) + u_2(t) q_k(x_{kj}) \right] + G(t) + o_p(n^{-1/2}). \end{split}$$

The last equality is due to the fact that

$$\sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{\omega_k(x;\lambda_0,\beta_0)}{R(x;\lambda_0,\beta_0)} = n \sum_{k=1}^{K} \rho_{nk} \omega_k(x;\lambda_0,\beta_0) \frac{1}{R(x;\lambda_0,\beta_0)} = n.$$

By the criteria of Billingsley (1968, p. 128), we can show that $\sqrt{n}(\tilde{G}(t) - G(t) \rightarrow B(t)$ weakly in $D[-\infty, +\infty]$, where B(t) is a mean zero Gaussian process with continuous paths and covariance structure

$$\Sigma(s,t) = \sum_{k=1}^{K} \rho_k \text{cov}(\varepsilon_k^G(X_k;s) + u_2(s)q_k(X_k), \varepsilon_k^G(X_k;t) + u_2(t)q_k(X_k)), \ s \le t.$$

Let $v_{2k}(t) = -v_{1k}^T(t)S^{-1}$ and let

$$\varepsilon_l^{H_k}(x_{lj};t) = \frac{\omega_k(x_{lj};\lambda_0,\beta_0)}{r(x_{lj};\lambda_0,\beta_0)} I(x_{lj} \le t) - \int_{-\infty}^t \frac{\omega_k(x;\lambda_0,\beta_0)}{R(x;\lambda_0,\beta_0)} \omega_l(x;\lambda_0,\beta_0) dF(x), \ l = 1, 2, \cdots, K,$$

then

$$\begin{split} \widetilde{H}_{k}(t) &= \frac{1}{n} \sum_{l=1}^{K} \sum_{j=1}^{n_{l}} \frac{\omega_{k}(x_{lj}; \lambda_{0}, \beta_{0})}{r(x_{lj}; \lambda_{0}, \beta_{0})} I(x_{lj} \leq t) + v_{2k}^{T}(t)Q_{n} + o_{p}(n^{-1/2}) \\ &= \frac{1}{n} \sum_{l=1}^{K} \sum_{j=1}^{n_{l}} \left[\varepsilon_{l}^{H_{k}}(x_{lj}; t) + \int_{-\infty}^{t} \frac{\omega_{k}(x)\omega_{l}(x)}{R(x; \lambda_{0}, \beta_{0})} dF(x) \right] + v_{2k}^{T}(t)Q_{n} + o_{p}(n^{-1/2}) \\ &= \frac{1}{n} \sum_{l=1}^{K} \sum_{j=1}^{n_{l}} \left[\varepsilon_{l}^{H_{k}}(x_{lj}; t) + v_{2k}^{T}(t)q_{l}(x_{lj}) \right] + \frac{1}{n} \int_{-\infty}^{t} \sum_{l=1}^{K} \sum_{j=1}^{n_{l}} \frac{\omega_{l}(x)dH_{k}(x)}{R(x; \lambda_{0}, \beta_{0})} + o_{p}(n^{-1/2}) \\ &= \frac{1}{n} \sum_{l=1}^{K} \sum_{j=1}^{n_{l}} \left[\varepsilon_{l}^{H_{k}}(x_{lj}; t) + v_{2k}^{T}(t)q_{k}(x_{lj}) \right] + H_{k}(t) + o_{p}(n^{-1/2}) \\ &= H_{1k}(t) + H_{k}(t) + o_{p}(n^{-1/2}), \end{split}$$

$$(2.4.10)$$

where we have denoted $\omega_k(x; \lambda_0, \beta_0)$ by $\omega_k(x)$ to save space. The last equality is due to the fact that

$$\sum_{l=1}^{K} \sum_{j=1}^{n_k} \frac{\omega_k(x; \lambda_0, \beta_0)}{R(x; \lambda_0, \beta_0)} = n \sum_{l=1}^{K} \rho_{nk} \omega_k(x; \lambda_0, \beta_0) \frac{1}{R(x; \lambda_0, \beta_0)} = n.$$

By the criteria of Billingsley (1968, p. 128), we can show that $\sqrt{n}(\tilde{H}_k(t) - H_k(t) \rightarrow B(t)$ weakly in $D[-\infty, +\infty]$, where B(t) is a mean zero Gaussian process with con-

tinuous paths and covariance structure

$$\Sigma_k(s,t) = \sum_{l=1}^{K} \rho_l \text{cov}(\varepsilon_l^{H_k}(X_l;s) + v_{2k}(s)q_l(X_l), \varepsilon_l^{H_k}(X_l;t) + v_{2k}(t)q_l(X_l)), \ s \le t.$$

Chapter 3

EM and MM Algorithms for the K2M Model and Its Special Cases

3.1 MM Algorithms for Model 2 and Model 3

We first consider Model 2. An EM algorithm has been proposed by Zhang (2002). An MM algorithm can be developed as follows. Note that the profiled log-likelihood (Qin and Zhang, 1997) is

$$l_{p}(\lambda,\beta_{1},\beta_{2}) = -n\log(n) - \sum_{i=1}^{n}\log\{1 + \eta[\exp(\beta_{1} + \varphi(t_{i};\beta_{2})) - 1]\} + \sum_{j=1}^{n_{2}}[\beta_{1} + \varphi(x_{2j};\beta_{2})] + \sum_{k=1}^{n_{3}}\log[\lambda + (1-\lambda)\exp(\beta_{1} + \varphi(x_{3k};\beta_{2})]]$$

where η is the Lagrange multiplier. Zhang (2002) has shown that, $\eta = [n_2 + n_3(1 - \lambda)]/n$ when l_p achieves its maximum.

A minorizing function for $\log[\lambda + (1 - \lambda) \exp(\beta_1 + \varphi(x_{3k}; \beta_2))]$ is

$$\frac{\lambda^{(m)}}{\lambda^{(m)} + (1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(x_{3k}; \beta_2^{(m)}))} \log\left\{\frac{\lambda^{(m)} + (1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(x_{3k}; \beta_2^{(m)}))}{\lambda^{(m)}} \lambda\right\} \\ + \frac{(1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(x_{3k}; \beta_2^{(m)}))}{\lambda^{(m)} + (1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(x_{3k}; \beta_2^{(m)}))} \log\left\{\frac{\lambda^{(m)} + (1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(x_{3k}; \beta_2^{(m)}))}{(1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(x_{3k}; \beta_2^{(m)}))} \cdot (1 - \lambda) \exp(\beta_1 + \varphi(x_{3k}; \beta_2))\right\}.$$

A minorizing function for $-\log\{1 + \eta[\exp(\beta_1 + \varphi(t_i; \beta_2)) - 1]\}$ is

$$-\log\{1+\eta^{(m)}[\exp(\beta_1^{(m)}+\varphi(t_i;\beta_2^{(m)}))-1]\}$$

$$-\frac{\eta[\exp(\beta_1+\varphi(t_i;\beta_2))-1]-\eta^{(m)}[\exp(\beta_1^{(m)}+\varphi(t_i;\beta_2^{(m)}))-1]}{1+\eta^{(m)}[\exp(\beta_1^{(m)}+\varphi(t_i;\beta_2^{(m)}))-1]}$$

Therefore, a minorizing function for $l(\lambda, \beta_1, \beta_2)$ is

$$\begin{split} &\sum_{j=1}^{n_2} [\beta_1 + \varphi(x_{2j};\beta_2)] + \sum_{k=1}^{n_3} \left\{ \frac{\lambda^{(m)}}{\lambda^{(m)} + (1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(x_{3k};\beta_2^{(m)}))} \log(\lambda) \right. \\ &+ \frac{(1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(x_{3k};\beta_2^{(m)}))}{\lambda^{(m)} + (1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(x_{3k};\beta_2^{(m)}))} \log[(1 - \lambda) \exp(\beta_1 + \varphi(x_{3k};\beta_2))] \right\} \\ &- \sum_{i=1}^n \frac{\eta[\exp(\beta_1 + \varphi(t_i;\beta_2)) - 1]}{1 + \eta^{(m)}[\exp(\beta_1^{(m)} + \varphi(t_i;\beta_2^{(m)})) - 1]} + c, \end{split}$$

which is denoted by $g(\theta|\theta^{(m)})$, where $\theta = (\lambda, \beta_1, \beta_2)$, $\eta^{(m)} = (n_2 + n_3(1 - \lambda^{(m)})/n$ and c is a constant.

Maximization of $g(\theta|\theta^{(m)})$ is completed as follows. Let $(\lambda^{(m+1)}, \eta^{(m+1)}, \beta_1^{(m+1)}, \beta_2^{(m+1)})$

be the next iterate which satisfies

$$\lambda^{(m+1)} = \frac{1}{n_3} \sum_{k=1}^{n_3} \frac{\lambda^{(m)}}{\lambda^{(m)} + (1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(x_{3k}; \beta_2^{(m)}))},$$
$$\eta^{(m+1)} = (n_2 + n_3(1 - \lambda^{(m+1)})/n,$$
$$n_2 + n_3(1 - \lambda^{(m+1)}) - \eta^{(m+1)} \sum_{i=1}^{n} \frac{\exp(\beta_1^{(m+1)} + \varphi(t_i; \beta_2^{(m+1)}))}{1 + \eta^{(m)} [\exp(\beta_1^{(m)} + \varphi(t_i; \beta_2^{(m)})) - 1]} = 0,$$

$$\sum_{j=1}^{n_2} x_{2j} + \sum_{k=1}^{n_3} x_{3k} - \sum_{k=1}^{n_3} \frac{\lambda^{(m)}}{\lambda^{(m)} + (1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(x_{3k}; \beta_2^{(m)}))} x_{3k} - \eta^{(m+1)} \sum_{i=1}^{n} \frac{\exp(\beta_1^{(m+1)} + \varphi(t_i; \beta_2^{(m+1)}))}{1 + \eta^{(m)} [\exp(\beta_1^{(m)} + \varphi(t_i; \beta_2^{(m)})) - 1]} t_i = 0,$$

where $\beta_1^{(m+1)}$ and $\beta_2^{(m+1)}$ can be obtained from the last two equations.

Remark 3.1.1. $\lambda^{(m)}$ and $\eta^{(m)}$ in the last two equations can be replaced by $\lambda^{(m+1)}$ and $\eta^{(m+1)}$.

Remark 3.1.2. $\beta_1^{(m+1)}$ can be written in terms of $\beta_2^{(m+1)}$, while $\beta_2^{(m+1)}$ can be obtained by first cancelling $\beta_1^{(m+1)}$ from the last two equations and solving an equation of this form $\sum_{i=1}^n c_i \exp[\varphi(t_i; \beta_2^{(m+1)})] = 0$, where c_i 's are known.

Next, we apply MM algorithm to Model 3. Similar to the arguments above, the iteration equations are

$$\lambda^{(m+1)} = \frac{1}{n_2} \sum_{j=1}^{n_2} \frac{\lambda^{(m)}}{\lambda^{(m)} + (1 - \lambda^{(m)}) \exp(\beta_1^{(m)} + \varphi(t_i; \beta_2^{(m)}))},$$
$$\eta^{(m+1)} = n_2 (1 - \lambda^{(m+1)})/n,$$

$$\sum_{i=1}^{n} \frac{\eta \left[\exp(\beta_{1}^{(m+1)} + \varphi(t_{i}; \beta_{2}^{(m+1)})) - 1\right]}{1 + \eta^{(m)} \left[\exp(\beta_{1}^{(m)} + \varphi(t_{i}; \beta_{2}^{(m)})) - 1\right]} = n_{2}(1 - \lambda^{(m+1)}),$$

$$\sum_{i=1}^{n} \frac{\eta t_{i} \left[\exp(\beta_{1}^{(m+1)} + \varphi(t_{i}; \beta_{2}^{(m+1)})) - 1\right]}{1 + \eta^{(m)} \left[\exp(\beta_{1}^{(m)} + \varphi(t_{i}; \beta_{2}^{(m)})) - 1\right]} = \sum_{j=1}^{n_{2}} \frac{(1 - \lambda^{(m)}) x_{2j} \exp(\beta_{1}^{(m)} + \varphi(x_{2j}; \beta_{2}^{(m)}))}{\lambda^{(m)} + (1 - \lambda^{(m)}) \exp(\beta_{1}^{(m)} + \varphi(x_{2j}; \beta_{2}^{(m)}))}$$

Similar to Zhang (2002), an EM algorithm can be proposed for Model 3.

3.2 EM and MM Algorithms for the *K*2*M* Model

Before we apply EM and MM algorithms to the K2M model, we first take a look at how MM algorithm works for K-sample two-component normal mixtures. The corresponding EM algorithm was proposed by McLachlan et al (1982).

3.2.1 An MM Algorithm for *K*-Sample Normal Mixtures

Suppose that the densities of the two components in assumption A_1 take parametric forms: $f(x; \theta_1)$ and $g(x; \theta_2)$. Based on a set of samples, the log-likelihood is

$$l(\theta) = \sum_{k=1}^{K} \sum_{j=1}^{n_k} \log\{\lambda_k f(x_{kj}; \theta_1) + (1 - \lambda_k)g(x_{kj}; \theta_2)\}.$$
 (3.2.1)

where $\theta = (\theta_1^T, \theta_2^T)^T$.

At iteration m, each summand of (3.2.1) can be minorized by

$$g_{kj} = \frac{\lambda_k^{(m)} f(x_{kj}; \theta_1^{(m)})}{H_k(x_{kj}; \theta^{(m)})} \log \left\{ \frac{H_k(x_{kj}; \theta^{(m)})}{\lambda_k^{(m)} f(x_{kj}; \theta_1^{(m)})} \lambda_k f(x_{kj}; \theta_1) \right\} \\ + \frac{\lambda_k^{(m)} f(x_{kj}; \theta_2^{(m)})}{H_k(x_{kj}; \theta^{(m)})} \log \left\{ \frac{H_k(x_{kj}; \theta^{(m)})}{\lambda_k^{(m)} f(x_{kj}; \theta_2^{(m)})} (1 - \lambda_k) f(x_{kj}; \theta_2) \right\},$$

where $H_k(x; \theta^{(m)}) = \lambda_k^{(m)} f(x; \theta_1^{(m)}) + (1 - \lambda_k^{(m)}) g(x; \theta_2^{(m)}).$

By the additive property of minorization, $l(\theta)$ is minorized by

$$g(\theta|\theta^{(m)}) = \sum_{k=1}^{K} \sum_{j=1}^{n_k} \left\{ \frac{\lambda_k^{(m)} f(x_{kj}; \theta_1^{(m)})}{H_k(x_{kj}; \theta^{(m)})} \log[\lambda_k f(x_{kj}; \theta_1)] + \frac{\lambda_k^{(m)} f(x_{kj}; \theta_2^{(m)})}{H_k(x_{kj}; \theta^{(m)})} \log[(1 - \lambda_k) f(x_{kj}; \theta_2)] \right\} + c,$$

where c is a constant.

Maximizing the surrogate function $g(\theta|\theta^{(m)})$ leads to the following equations:

$$\lambda_k = \frac{1}{n_k} \sum_{j=1}^{n_k} \frac{\lambda_k^{(m)} f(x_{kj}; \theta_1^{(m)})}{H_k(x_{kj}; \theta^{(m)})}, \qquad (3.2.2)$$

$$\sum_{k=1}^{K} \sum_{j=1}^{n_k} \left\{ \frac{\lambda_k^{(m)} f(x_{kj}; \theta_1^{(m)})}{H_k(x_{kj}; \theta^{(m)})} \cdot \frac{\partial \log f(x_{kj}; \theta_1)}{\partial \theta_1} \right\} = 0,$$
(3.2.3)

$$\sum_{k=1}^{K} \sum_{j=1}^{n_k} \left\{ \frac{(1-\lambda_k^{(m)})g(x_{kj};\theta_2^{(m)})}{H_k(x_{kj};\theta^{(m)})} \cdot \frac{\partial \log g(x_{kj};\theta_2)}{\partial \theta_2} \right\} = 0.$$
(3.2.4)

Let's further consider the situation where f and g are normal densities with parameters $\theta_1 = (\mu_1, \Sigma_1)$ and $\theta_2 = (\mu_2, \Sigma_2)$, respectively. Then the equation (3.2.3) yields

$$0 = \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{\lambda_k^{(m)} f(x_{kj}; \ \mu_1^{(m)}, \ \Sigma_1^{(m)})}{H_k(x_{kj}; \theta^{(m)})} \{-\Sigma_1^{-1}(x_{kj} - \mu_1)\},\$$

$$0 = \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{\lambda_k^{(m)} f(x_{kj}; \ \mu_1^{(m)}, \ \Sigma_1^{(m)})}{H_k(x_{kj}; \theta^{(m)})} \{ -\frac{1}{2} \Sigma_1^{-T} + \frac{1}{2} \Sigma_1^{-T} (x_{kj} - \mu_1) (x_{kj} - \mu_1)^T \Sigma_1^{-T} \},\$$

which give

$$\mu_1^{(m+1)} = \frac{\sum_{k=1}^K \sum_{j=1}^{n_k} \frac{\lambda_k^{(m)} f(x_{kj}; \mu_1^{(m)}, \Sigma_1^{(m)})}{H_k(x_{kj}; \theta^{(m)})} x_{kj}}{\sum_{k=1}^K \sum_{j=1}^{n_k} \frac{\lambda_k^{(m)} f(x_{kj}; \mu_1^{(m)}, \Sigma_1^{(m)})}{H_k(x_{kj}; \theta^{(m)})}},$$

$$\Sigma_{1}^{(m+1)} = \frac{\sum_{k=1}^{K} \sum_{j=1}^{n_{k}} \frac{\lambda_{k}^{(m)} f(x_{kj}; \mu_{1}^{(m)}, \Sigma_{1}^{(m)})}{H_{k}(x_{kj}; \theta^{(m)})} (x_{kj} - \mu_{1}) (x_{kj} - \mu_{1})^{T}}{\sum_{k=1}^{K} \sum_{j=1}^{n_{k}} \frac{\lambda_{k}^{(m)} f(x_{kj}; \mu_{1}^{(m)}, \Sigma_{1}^{(m)})}{H_{k}(x_{kj}; \theta^{(m)})}}$$

Similarly, the equation (3.2.4) yields

$$\mu_{2}^{(m+1)} = \frac{\sum_{k=1}^{K} \sum_{j=1}^{n_{k}} \frac{\lambda_{k}^{(m)} f(x_{kj}; \mu_{2}^{(m)}, \Sigma_{2}^{(m)})}{H_{k}(x_{kj}; \theta^{(m)})} x_{kj}}{\sum_{k=1}^{K} \sum_{j=1}^{n_{k}} \frac{\lambda_{k}^{(m)} f(x_{kj}; \mu_{2}^{(m)}, \Sigma_{2}^{(m)})}{H_{k}(x_{kj}; \theta^{(m)})}},$$

$$\Sigma_{2}^{(m+1)} = \frac{\sum_{k=1}^{K} \sum_{j=1}^{n_{k}} \frac{\lambda_{k}^{(m)} f(x_{kj}; \mu_{2}^{(m)}, \Sigma_{2}^{(m)})}{H_{k}(x_{kj}; \theta^{(m)})} (x_{kj} - \mu_{2}) (x_{kj} - \mu_{2})^{T}}{\sum_{k=1}^{K} \sum_{j=1}^{n_{k}} \frac{\lambda_{k}^{(m)} f(x_{kj}; \mu_{2}^{(m)}, \Sigma_{2}^{(m)})}{H_{k}(x_{kj}; \theta^{(m)})}}.$$
(3.2.5)

Remark 3.2.1. When K = 2, the iteration equations are the same as those obtained by McLachlan (1981) with EM algorithm.

Remark 3.2.2. When $\Sigma_1 = \Sigma_2 = \Sigma$, the $(m+1)^{th}$ iteration produces

$$\Sigma^{(m+1)} = \frac{1}{n} \left\{ \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{\lambda_k^{(m)} f(x_{kj}; \mu_1^{(m)}, \Sigma^{(m)})}{H_k(x_{kj}; \theta^{(m)})} (x_{kj} - \mu_1) (x_{kj} - \mu_1)^T + \frac{(1 - \lambda_k^{(m)}) f(x_{kj}; \mu_2^{(m)}, \Sigma^{(m)})}{H_k(x_{kj}; \theta^{(m)})} (x_{kj} - \mu_2) (x_{kj} - \mu_2)^T \right\}.$$

3.2.2 An EM Algorithm for the *K*2*M* Model

Define indicators Y_{kj} , k = 1, 2, ..., K, $j = 1, 2, ..., n_K$ as

$$Y_{kj} = \begin{cases} 0 & \text{if } X_{kj} \text{ is from drawn } F(x); \\ 1 & \text{if } X_{kj} \text{ is drawn from } G(x). \end{cases}$$

The likelihood of the complete data $(X_{kj}, Y_{kj}), k = 1, 2, ..., j = 1, 2, ..., n_K$ is

$$L_{c}(\theta, F) = \prod_{k=1}^{K} \prod_{j=1}^{n_{k}} [p(Y_{kj} = y_{kj}) dH(x_{kj} | Y_{kj} = y_{kj})]$$

$$= \prod_{k=1}^{K} \prod_{j=1}^{n_{k}} [\lambda_{k}^{1-y_{kj}} (1-\lambda_{k})^{y_{kj}} dF(x_{kj}) u(x_{kj})^{y_{kj}}]$$

$$= [\prod_{i=1}^{n} p_{i}] [\prod_{k=1}^{K} \prod_{j=1}^{n_{k}} \lambda_{k}^{1-y_{kj}} (1-\lambda_{k})^{y_{kj}}] [\prod_{k=1}^{K} \prod_{j=1}^{n_{k}} u(x_{kj})^{y_{kj}}],$$

where $(t_1, t_2, ..., t_n)$ consist of all samples and $p_i = dF(t_i), i = 1, 2, ..., n$, are nonnegative jumps with total mass unity. Let $\psi = (\theta, F)$, and

$$l_{c}(\varphi) = \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \log(p_{i}) + \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} y_{kj} [\beta_{1} + \varphi(x_{kj};\beta_{2})] + \sum_{i=1}^{N} \sum_{j=1}^{n_{k}} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \sum_{j=1}^{n_{k}} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \log(p_{i}) + \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \log(p_{i}) + \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \log(p_{i}) + \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \log(p_{i}) + \sum_{k=1}^{N} \sum_{j=1}^{n_{k}} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \log(p_{i}) + \sum_{k=1}^{N} \sum_{j=1}^{n_{k}} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \log(p_{i}) + \sum_{k=1}^{N} \sum_{j=1}^{n_{k}} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \log(p_{i}) + \sum_{i=1}^{n} \sum_{j=1}^{n_{k}} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{j=1}^{n} \sum_{j=1}^{n} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{j=1}^{n} \sum_{j=1}^{n} \sum_{j=1}^{n} [y_{kj} \log(1-\lambda_{k}) + (1-y_{kj}) \log(\lambda_{k})] + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{j=1$$

which is linear in the unobservable data y_{kj} , k = 1, 2, ..., K, $j = 1, 2, ..., n_k$. The EM algorithm is an iterative method. Let $\psi^{(m)} = (\theta^{(m)}, F^{(m)}) = (\beta_1^{(m)}, \beta_2^{(m)}, \lambda^{(m)}, F^{(m)})$ be the *m*th iterate. On the next iteration, the EM algorithm takes two steps.

E-step. The conditional expectation of $l_c(\varphi)$ given the observed data $x = \{x_{kj}, k = 1, 2, ..., K, j = 1, 2, ..., n_k\}$, is

$$Q(\psi;\psi^{(m)} = E_{\psi^{(m)}}[l_c(\psi)|x] = l_1^{(m)}(\lambda) + \lambda_2^{(m)}(\beta_1, \beta_2, F),$$

where

$$l_1^{(m)}(\lambda) = \sum_{k=1}^K \sum_{j=1}^{n_k} [(1 - y_{kj}^{(m)} \log(\lambda_k) + y_{kj}^{(m)} \log(1 - \lambda_k)],$$

$$l_2^{(m)}(\beta_1, \beta_2, F) = \sum_{i=1}^n \log(p_i) + \beta_1 \sum_{k=1}^K \sum_{j=1}^{n_k} y_{kj}^{(m)} + \sum_{k=1}^K \sum_{j=1}^{n_k} y_{kj}^{(m)} \varphi(x_{kj}; \beta_2)$$

with $y_{kj}^{(m)}$, k = 1, 2, ..., K, $j = 1, 2, ..., n_k$, being

$$E_{\psi^{(m)}}(Y_{kj}|x) = p_{\psi^{(m)}}(Y_{kj} = 1|x_{kj})$$

$$= \frac{(1 - \lambda_k^{(m)})\exp(\beta_1^{(m)} + \varphi(x_{kj};\beta_2^{(m)}))}{\lambda_k^{(m)} + (1 - \lambda_k^{(m)})\exp(\beta_1^{(m)} + \varphi(x_{kj};\beta_2^{(m)}))}$$

$$= \frac{\exp(\beta_1^* + \varphi(x_{kj};\beta_2^{(m)})}{1 + \exp(\beta_1^* + \varphi(x_{kj};\beta_2^{(m)})},$$

where $\beta_1^* = \beta_1^{(m)} + \log \frac{1 - \lambda_k^{(m)}}{\lambda_k^{(m)}}$.

M-step. The surrogate function $Q(\psi; \psi^{(m)})$ is maximized w.r.t. ψ and the maximizer is denoted $\varphi^{(m+1)} = (\beta_1^{(m+1)}, \beta_2^{(m+1)}, \lambda^{(m+1)}, F^{(m+1)})$. It's easy to see that

$$\lambda_k^{(m+1)} = 1 - \bar{y}_k^{(m)} = 1 - \frac{1}{n_k} \sum_{j=1}^{n_k} y_{kj}^{(m)}, \ k = 1, 2, ..., K$$

maximize $l_1^{(m)}(\lambda)$.

In order to maximize $l_2^{(m)}(\beta_1, \beta_2, F)$, we follow Qin and Zhang (1997). First, for fixed (β_1, β_2) , $\sum_{i=1}^n \log(p_i)$ is maximized subject to constraints $\sum_{i=1}^n p_i = 1$, $p_i \ge 0$, $\sum_{i=1}^n p_i[u(t_i) - 1] = 0$. The use of Lagrange procedure yields the maximizer

$$p_i^{(m+1)} = \frac{1}{n} \frac{1}{1 + \xi^{(m)}[u(t_i) - 1]}, \ i = 1, 2, ..., n,$$
(3.2.6)

where the lagrange multiplier $\xi^{(m)}$ satisfies the equation

$$\frac{1}{n}\sum_{i=1}^{n}\frac{u(t_i)-1}{1+\xi^{(m)}[u(t_i)-1]}=0.$$

Next we maximize

$$l_{2}^{(m)}(\beta_{1},\beta_{2}) = l_{2}^{(m)}(\beta_{1},\beta_{2},F^{(m+1)})$$

$$= \sum_{i=1}^{n} \log(p_{i}^{(m+1)}) + \beta_{1} \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} y_{kj}^{(m)} + \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} y_{kj}^{(m)} \varphi(x_{kj};\beta_{2})$$

$$= -n \log(n) - \sum_{i=1}^{n} \log[1 + \xi^{(m)}(u(t_{i}) - 1)] + \beta_{1} \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} y_{kj}^{(m)}$$

$$+ \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} y_{kj}^{(m)} \varphi(x_{kj};\beta_{2})$$

over (β_1,β_2) . The maximizer $(\beta_1^{(m+1)},\beta_2^{(m+1)})$ satisfies the system of score equations

$$\frac{\partial l_2^{(m)}}{\partial \beta_1} = -\sum_{i=1}^n \frac{\xi^{(m)} u(t_i)}{1 + \xi^{(m)} [u(t_i) - 1]} + \sum_{k=1}^K \sum_{j=1}^{n_k} y_{kj}^{(m)} = 0, \qquad (3.2.7)$$

$$\frac{\partial l_2^{(m)}}{\partial \beta_2} = -\sum_{i=1}^n \frac{\xi^{(m)} u(t_i) t_i}{1 + \xi^{(m)} [u(t_i) - 1]} + \sum_{k=1}^K \sum_{j=1}^{n_k} y_{kj}^{(m)} \frac{\partial \varphi}{\partial \beta_2} = 0, \quad (3.2.8)$$

It can be shown that equation (3.2.7) along with equation (3.2.6) implies that

$$\xi^{(m)} = \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} y_{kj}^{(m)}.$$

The E-step and M-step alternate until convergence is achieved.

In the special case where $\varphi(x; \beta_2) = \beta_2^T x$, Following Zhang (2002), we can make use

of the standard logistic software to solve the system of equations (3.2.7) and (3.2.8). The procedure is as follows. Suppose the current iterate is $(\beta_1^{(m)}, \beta_2^{(m)}, \lambda_k^{(m)},), k = 1, 2, ...$ The next iterate is to calculate $(\beta_1^{(m+1)}, \beta_2^{(m+1)}, \lambda_k^{(m+1)},), k = 1, 2, ...$ in the following order. First calculate

$$y_{kj}^{(m)} = \frac{\exp(\beta_1^* + \beta_2^{(m)} x_{kj})}{1 + \exp(\beta_1^* + \beta_2^{(m)} x_{kj})}, \text{ where } \beta_1^* = \beta_1^{(m)} + \log(\frac{1 - \lambda_k^{(m)}}{\lambda_k^{(m)}}), \text{ and } \lambda_k^{(m+1)} = 1 - \bar{y}_k^{(m)} = 1 - \frac{1}{n_k} \sum_{j=1}^{n_k} y_{kj}^{(m)}.$$

Then with the help of standard software for logistic regression, $(\beta_1^*, \beta_2^{(m+1)})$ solves the system

$$\sum_{i=1}^{n} \frac{\exp(\beta_{1}^{*} + \beta_{2}^{(m+1)}t_{i})}{1 + \exp(\beta_{1}^{*} + \beta_{2}^{(m+1)}t_{i})} = \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} y_{kj}^{(m)},$$
$$\sum_{i=1}^{n} \frac{\exp(\beta_{1}^{*} + \beta_{2}^{(m+1)}t_{i})}{1 + \exp(\beta_{1}^{*} + \beta_{2}^{(m+1)}t_{i})}t_{i} = \sum_{k=1}^{K} \sum_{j=1}^{n_{k}} y_{kj}^{(m)}x_{kj},$$

which are the score equations of a logistic regression. Recall that the standard logistic procedure works for paired data (x, y), where y can be any values between 0 and 1. For our case, the paired data are (x_{kj}, y_{kj}) , where $k = 1, 2, \dots, K$ and $j = 1, 2, \dots, n_k$. The EM algorithm is completed by setting $\beta_1^{(m+1)} = \beta_1^* - \log \frac{\xi^{(m)}}{1-\xi^{(m)}}$, where $\xi^{(m)} = \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} y_{kj}^{(m)}$.

3.2.3 An MM Algorithm for the *K*2*M* Model

To apply the MM algorithm to the K2M model, we first write the empirical log-likelihood of the observed data.

$$l = \sum_{i=1}^{n} \log(p_i) + \sum_{k=1}^{K} \sum_{j=1}^{n_k} \log[\lambda_k + (1 - \lambda_k) \exp(\beta_1 + \psi(x_{kj}; \beta_2))]$$

The function l is minorized by

$$Q(\theta|\theta^{(m)}) = \sum_{i=1}^{n} \log(p_i) + \sum_{k=1}^{K} \sum_{j=1}^{n_k} \{ \frac{\lambda_k^{(m)}}{\lambda_k^{(m)} + (1-\lambda_k^{(m)}) \exp(\beta_1^{(m)} + \beta_2^{(m)} x_{kj})} \log(\lambda_k) + \frac{(1-\lambda_k^{(m)}) \exp(\beta_1^{(m)} + \beta_2^{(m)} x_{kj})}{\lambda_k^{(m)} + (1-\lambda_k^{(m)}) \exp(\beta_1^{(m)} + \psi(x_{kj};\beta_2))} [\log(1-\lambda_k) + \beta_1 + \psi(x_{kj};\beta_2)] \}.$$

Denote the current iterate as $\theta^{(m)}$. At next iteration, instead of maximizing the function $l(\theta)$, where $\theta = (a, b, \lambda^T, F)^T$, subject to $\sum_{i=1}^n p_i = 1$, $p_i \ge 0$, $\sum_{i=1}^n p_i [\exp(\beta_1 + \psi(t_i; \beta_2)) - 1] = 0$, we maximize the surrogate function $Q(\theta|\theta^{(m)})$, subject to the same conditions. Construct the lagrangian

$$m = Q(\theta|\theta^{(m)}) - \zeta(\sum_{i=1}^{n} p_i - 1) - n\eta \sum_{i=1}^{n} p_i[\exp(\beta_1 + \psi(t_i; \beta_2)) - 1].$$
(3.2.9)

Then the first order conditions are

$$0 = \frac{\partial m}{\partial \beta_1} = \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{(1 - \lambda_k^{(m)}) \exp(\beta_1^{(m)} + \psi(x_{kj}; \beta_2^{(m)}))}{\lambda_k^{(m)} + (1 - \lambda_k^{(m)}) \exp(\beta_1^{(m)} + \psi(x_{kj}; \beta_2^{(m)}))} - n\eta \sum_{i=1}^{n} p_i \exp(\beta_1 + \psi(t_i; \beta_2)),$$
(3.2.10)

$$0 = \frac{\partial m}{\partial \beta_2} = \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{(1 - \lambda_k^{(m)}) \exp(\beta_1^{(m)} + \psi(x_{kj}; \beta_2^{(m)}))}{\lambda_k^{(m)} + (1 - \lambda_k^{(m)}) \exp(\beta_1^{(m)} + \psi(x_{kj}; \beta_2^{(m)}))} x_{kj}$$
$$- n\eta \sum_{i=1}^{n} p_i t_i \exp(\beta_1 + \psi(t_i; \beta_2)), \qquad (3.2.11)$$

$$0 = \frac{\partial m}{\partial \lambda_k} = \frac{1}{\lambda_k} \sum_{j=1}^{n_k} \frac{\lambda_k^{(m)}}{\lambda_k^{(m)} + (1 - \lambda_k^{(m)}) \exp(\beta_1^{(m)} + \psi(x_{kj}; \beta_2^{(m)}))} - \frac{1}{1 - \lambda_k} \sum_{j=1}^{n_k} \frac{(1 - \lambda_k^{(m)}) \exp(\beta_1^{(m)} + \psi(x_{kj}; \beta_2^{(m)}))}{\lambda_k^{(m)} + (1 - \lambda_k^{(m)}) \exp(\beta_1^{(m)} + \psi(x_{kj}; \beta_2^{(m)}))},$$
(3.2.12)

$$0 = \frac{\partial m}{\partial p_i} = \frac{1}{p_i} - \zeta - n\eta [\exp(\beta_1 + \psi(t_i; \beta_2)) - 1], \qquad (3.2.13)$$

$$0 = \frac{\partial m}{\partial \zeta} = 0$$
, that is $\sum_{i=1}^{n} p_i - 1 = 0$ (3.2.14)

$$0 = \frac{\partial m}{\partial \eta} = 0, \text{ that is } \sum_{i=1}^{n} p_i \exp(\beta_1 + \psi(t_i; \beta_2)) = 1.$$
 (3.2.15)

The equations (3.2.13), (3.2.14) and (3.2.15) imply that

$$\zeta = n, \tag{3.2.16}$$

$$p_i = \frac{1}{n} \frac{1}{1 + \eta(\exp(\beta_1 + \psi(t_i; \beta_2)) - 1)}.$$
(3.2.17)

Equations (3.2.10) and (3.2.15) produce

$$\eta = \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{(1 - \lambda_k^{(m)}) \exp(\beta_1^{(m)} + \beta_2^{(m)} x_{kj})}{\lambda_k^{(m)} + (1 - \lambda_k^{(m)}) \exp(\beta_1^{(m)} + \psi(x_{kj}; \beta_2^{(m)}))},$$
(3.2.18)
which implies that $0 \le \eta \le 1$. Solving equation (3.2.12) gives

$$\lambda_k = \frac{1}{n_k} \sum_{j=1}^{n_k} \frac{\lambda_k^{(m)}}{\lambda_k^{(m)} + (1 - \lambda_k^{(m)}) \exp(\beta_1^{(m)} + \psi(x_{kj}; \beta_2^{(m)}))}, \quad k = 1, 2, ..., K. \quad (3.2.19)$$

The equation (3.2.18) is equivalent to

$$\eta = \frac{1}{n} \left[n - \sum_{k=1}^{K} \lambda_k^{(m)} \sum_{j=1}^{n_k} \frac{1}{\lambda_k^{(m)} + (1 - \lambda_k^{(m)}) \exp(\beta_1^{(m)} + \psi(x_{kj}; \beta_2^{(m)}))} \right], \quad (3.2.20)$$

Equations (3.2.19) and (3.2.20) produce

$$\eta = \frac{1}{n} \sum_{k=1}^{K} n_k (1 - \lambda_k).$$
(3.2.21)

Solving equations (3.2.10) and (3.2.11) gives the next iterates $\beta_1^{(m+1)}$ and $\beta_2^{(m+1)}$.

Remark 3.2.3. To generate $\beta_1^{(m+1)}$ and $\beta_2^{(m+1)}$, Remark 3.1.2 applies here.

3.3 Simulation Studies

We will study through simulations how well the proposed methods behave for different mixtures. For each estimate, the average asymptotic variance and empirical variance will be reported. Specifically, we study normal, exponential, and poisson mixtures. We focus ourselves on the case of K = 3. From each mixture, a sample of size 150 or 250 is simulated. 1000 simulations are run. In each run, we compute the model-based estimate and its standard error for each parameter and report the average values based on all 1000 runs. For comparison, the empirical standard deviation is also provided. The simulation result is reported in Table (3.1). In Table (3.1), we have considered six cases with different pairs of distributions, different mixing proportions, and different sample sizes. Specifically, the following cases are considered.

Case 1: f is N(0,1) and g is N(2,1); Case 2: f is N(0,1) and g is N(1,1); Case 3: f is Exp(1) and g is Exp(3); Case 4: f is Exp(1) and g is Exp(1.5); Case 5: f is Pois(1) and g is Pois(3); Case 6: f is Pois(1) and g is Pois(1.5).

The results show that the empirical or sample standard deviations match the asymptotic version well, except for some cases. The limited simulation also shows that, when a mixing proportion is closer to 0 or 1, the precisions of all estimates are higher. In the table, "-" stands for divergence.

3.4 Real Examples

Examples are based on data sets of Anderson (1979), McLachlan et al (1982), Smith and Vounatsou (1997), and Nagelkerke et al (2001).

Example 3.4.1. Anderson (1979) simulated a small data set containing three samples, of which one sample is from the mixture of 0.25N(2,1) + 0.75N(0,1), and the other two are from the two components separately.

Specifically, Anderson generated these three samples: 1.15, 0.25, 2.31, 2.44, 3.28, 3.44; 0.74, -0.50, 1.08, 1.34, -0.74, 0.15; -0.23, 0.71, 0.92, -0.53, -0.68, 1.04, 0.61, -0.88, -0.61, 0.59, 2.96, 2.59; randomly from N(2, 1), N(0, 1), and 0.25N(2, 1) + 0.75N(0, 1), respectively.

						$\widetilde{\lambda}_1$			$\widetilde{\lambda}_2$			$\widetilde{\lambda}_3$	
Cases	λ_1	λ_2	λ_3	n	ave	var1	var2	ave	var1	var2	ave	var1	var2
1	.99	.00	.25	150	.984	.020	.019	.009	.019	.017	.255	.045	.041
				250	.986	.018	.012	.007	.012	.014	.255	.042	.037
	.35	.50	.75	150	.293	.136	.268	.460	.132	.173	.728	.226	.253
				250	.309	.119	.149	.480	.100	.109	.754	.167	.145
2	.99	.00	.25	150	.924	.094	.126	.068	.094	.123	.283	.115	.118
				250	.935	.101	.112	.057	.090	.067	.278	.086	.101
	.35	.50	.75	150	.339	.270	.261	.495	.201	.189	.744	.159	.144
				250	.328	.239	.254	.483	.180	.177	.741	.149	.141
3	.99	.00	.25	150	.985	.077	.088	.061	.028	.041	.288	.069	.057
				250	.951	.055	.068	.011	.019	.031	.249	.053	.047
	.35	.50	.75	150	.333	.143	.175	.509	.201	.144	.797	.211	.158
				250	.281	.105	.154	.471	.157	.164	.774	.189	.127
4	.99	.00	.25	150	.512	.428	.402	.660	.257	.245	.699	.127	.131
				250	-	-	-	-	-	-	-	-	-
	.35	.50	.75	150	-	-	-	-	-	-	-	-	-
				250	-	-	-	-	-	-	-	-	-
5	.99	.00	.25	150	.942	.067	.057	.014	.026	.003	.249	.066	.054
				250	.953	.055	.036	.01	.019	.022	.248	.053	.055
	.35	.50	.75	150	.348	.214	.221	.528	.151	.146	.816	.181	.177
				250	-	-	-	-	-	-	-	-	-
6	.99	.00	.25	150	-	-	-	-	-	-	-	-	-
				250	-	-	-	-	-	-	-	-	-
	.35	.50	.75	150	.445	.312	.336	.534	.272	.251	.791	.145	.163
	.35	.50	.75	250	.405	.301	.323	.523	.250	.247	.795	.144	.158

Table 3.1: A simulation: estimation of mixture proportions from 1000 samples.

ave: average of $\widetilde{\lambda}$, var1: empirical variance of $\widetilde{\lambda}$, ave2: average of $\widetilde{\sigma}_{\widetilde{\lambda}}$

The results are listed in Table 3.2. As we can see, The parametric method – EM(normal) is fastest. It's interesting that it gives the worst answer. Also, the EM and MM methods are compatible based on our semiparametric mixture model 1 and 2, the proposed EM and/or MM algorithms give the same estimation of 0.189 for the mixing proportion, which is the same as those in Qin (1999) and Zhang (2002).

Table 3.2: Tolerance $\epsilon = 10^{-10}$; Pentium 4 CPU 2.53GHz 504 Mb of RAM

$\hat{ heta}$	startings	EM(1)	$EM(2)^*$	MM(1)	MM(2)	EM(normal)
$\hat{\lambda}_1$	0.1	1	-	1	1	0.664
$\hat{\lambda}_2$	0.2	0	-	0	0	0
$\hat{\lambda}_3$	0.3	0.189	0.189	0.189	0.189	0.165
\hat{eta}_1	-0.5	1.94	1.94	1.94	1.94	-
\hat{eta}_2	-1	-1.75	-1.75	-1.75	-1.75	-
iterations	-	52	-	572	586	32
CPU time(sec.)	-	1.05	-	1.11	0.75	0.03

*See Zhang (2002)

In the table, EM(1) and EM(normal) stand for EM algorithms for model 1 and normal mixtures, respectively.

Example 3.4.2. Smith and Vounatsou (1997) analyzed data from a study that quantifies the capacity of mouse cells in culture to transfer small molecules between one another. They considered a mixture of two components with an uncategorized sample from the mixture and a categorized sample from one of the two components.

Smith and Vounatsou proposed four methods for their problem. We fit a 2-mixture model for the same problem. The two mixing proportions are found to be 0 and 0.73, which are very close to their best results. The standard error for the two estimates are 0.02 and 0.07, respectively.

Example 3.4.3. McLachlan et al (1982) analyzed data representing survival times in weeks for two sets of rats which were given dosages of cytoxan at a concentration of 60 mg/kg. The second set was given the full dosage once weekly, while the first received half the dosage twice weekly. They compared the toxicity of the chemical agent at the two dosage levels, assuming that f and g are both normal with different means and different variances. Two samples of respective size 40 and 44 are available, each from a mixture of two homogeneous normal components.

Without the normality assumption, our EM and MM algorithms give estimations of 0.80 and 0.27 for the two mixing proportions, which are close to McLachlan's result. We noted that standard errors were not able to compute. This might be due to the problem that the density ratio model is not valid for the data. Numeric results show that β_1 and β_2 are estimated by 1094.7 and -139.0, respectively.

Example 3.4.4. Nagelkerke et al (2001) studied 42 populations (year by age), each of which is a mixture of two types of infection - Mycobacterium tuberculosis (TB) or environmental mycobacteria (EM). Mantoux tests are used for establishing tuberculous infection. The test results for year 1965 to 1995, after excluding some cases, are viewed as samples from mixtures of two distributions. The ratios of the two component densities for different populations are the same up to a population specific constant.

We analyze the same data by treating the two components to be homogeneous across all populations. Here are our results based on the semiparametric model (S), a fully parametric normal mixture model (P), and Nagerlkerke's results (in parentheses). year85:

(S) 0.136 0.217 0.610 0.922 0.995 1.000
(P) 0.289 0.357 0.641 0.872 0.924 0.940
(0.164 0.243 0.604 0.892 0.960 0.971)

year90:

(S) 0.128 0.367 0.692 0.936 0.825 0.893
(P) 0.331 0.561 0.746 0.928 0.850 0.903
(0.214 0.474 0.749 0.969 0.882 0.937)

year95:

(S) 0.037 0.046 0.414 0.960 0.900 1.000
(P) 0.542 0.448 0.654 0.977 0.923 0.991
(0.043 0.050 0.389 0.903 0.850 0.945)

We take the year 1995 results for interpretation. In year 1995, six populations of different age groups are subject to the mantoux test. The infection rates due to TB are 0.037, 0.046, 0.414, 0.960, 0.900, 1.000, respectively, based on our methods. The histograms of the data for the year 90 are shown in Figure (3-1). The fitted CDF's and the non-parametric CDF's are shown in Figure (3-2), where ECDF stands for Empirical Cumulative Distribution Function, and SCDF stands for Semiparametric Cumulative Distribution Function. The two graphs well support each other. On the one hand, the histograms for the first two age groups are flat when the induration size is greater than 4, while others show a clear second mode around induration size 17. On the other hand, the fitted cumulative distribution curves match the empirical counterparts well except for the first two age groups. We will report the goodness of fit result in Chapter 5.



Figure 3-1: Histograms of the six tuberculin data for 1990.



Figure 3-2: Fitted CDF's and the non-parametric CDF's for 1990 tuberculin data.

Example 3.4.5. The WHO diagnostic criteria for diabetes mellitus were determined in part by evidence that in some populations the plasma glucose level 2-h after an oral glucose load is a mixture of two distinct distributions. Thompson et al (1998) model the two underlying distributions as generalized linear models and the mixture probabilities as logistic regression models. The population is divided into 4 subpopulations. The sample sizes are 300, 276, 161, and 182, respectively.

We analyze the same data by treating the two components to be homogeneous across all populations. The mixing proportions based on normal mixtures are estimated to be 0.634, 0.429, 0.712, 0.548, while based on our methods, the estimates are 0.753, 0.988, 0.474, 0.761. The results from the two methods are quite different. From the histograms, we can see the normal mixtures fit well.



Figure 3-3: Histograms of the 2-h plasma glucose data.

Chapter 4

Statistical Inference on Parameters

In this chapter, we consider the inference problem about our parameter of interest. All inferences are based on the assumption that $\varphi(x;\beta_2) \neq 0$ for some x, since otherwise there will exist the so called irregularity likelihood problem, as noted in the work of Zou, Fine and Yandell (2002), who consider inferences based on a partial likelihood without the restriction of $\beta_2 \neq 0$ in the case of $\varphi(x;\beta_2) = \beta_2^T x$.

4.1 Confidence Intervals

Confidence intervals can be constructed by using the normal theory for large samples or by inverting likelihood ratio statistics. It is also a good practice to use bootstrap methods for small samples, which will not be considered here.

4.1.1 A Normal-theory-based Confidence Region for λ

Consider an individual proportion, λ_1 say. The 95% confidence interval reads:

$$\widetilde{\lambda}_1 \pm 1.96 \sqrt{var(\widetilde{\lambda}_1)},$$

where the variance of $\tilde{\lambda}_1$, $var(\tilde{\lambda}_1)$, is extracted from the asymptotic variance matrix U in (2.2.1), which is estimated consistently using the plug-in principle.

A 95% joint confidence interval for $\lambda = (\lambda_1, \dots, \lambda_K)^T$ can be formed by inverting the χ^2_K quantity

$$(\widetilde{\lambda} - \lambda)^T \widetilde{\Sigma_{\widetilde{\lambda}}} (\widetilde{\lambda} - \lambda),$$

where $\Sigma_{\tilde{\lambda}}$ is the variance of $\tilde{\lambda}$ and is extracted from the asymptotic variance matrix U in (2.2.1), which is estimated using the plug-in principle.

4.1.2 A Likelihood Ratio-based Confidence Region for λ

Alternatively, we can construct a confidence interval for λ by inverting the semiparametric empirical likelihood ratio test statistic

$$R(\lambda) = 2[\sup_{\lambda,\beta} l(\lambda,\beta,\alpha(\lambda,\beta)) - \sup_{\beta} l(\lambda,\beta,\alpha(\lambda,\beta))].$$
(4.1.1)

This is based on the following theorem (4.1.1).

Theorem 4.1.1. Under the conditions of (2.2.3), $R(\lambda) = 2[\sup_{\lambda,\beta} l(\lambda, \beta, \alpha(\lambda, \beta)) - \sup_{\beta} l(\lambda, \beta, \alpha(\lambda, \beta))] \rightarrow \chi^2_K$, when $H_0: \lambda = \lambda_0$ is true.

A proof of this result is given in the proof section. Therefore, a 95% confidence interval for λ is constructed as

$$\{\lambda | R(\lambda) \le \chi_K^2 (1 - 0.95)\}. \tag{4.1.2}$$

4.1.3 A Normal-theory-based Confidence Interval for β_2

The 95% confidence interval for β_2 is:

$$\widetilde{\beta}_2 \pm 1.96 \sqrt{var(\widetilde{\beta}_2)},$$

where the variance of $\tilde{\beta}_2$ is extracted from the asymptotic variance matrix U in (2.2.1), which is estimated by the plug-in principle.

4.1.4 A Likelihood Ratio-based Confidence Interval for β_2

Alternatively, we can construct a confidence interval for β_2 by inverting the semiparametric empirical likelihood ratio test statistic

$$R(\beta_2) = 2[\sup_{\lambda,\beta} l(\lambda,\beta,\alpha(\lambda,\beta)) - \sup_{\lambda,\beta_1} l(\lambda,\beta,\alpha(\lambda,\beta))].$$
(4.1.3)

Theorem 4.1.2. Let β_{20} be the true value of β and $p = dim(\beta_2)$. Under the conditions of (2.2.3), $R(\beta_{20}) \rightarrow \chi_p^2$. Therefore, a 95% confidence interval for β_2 is constructed as

$$\{\beta_2 | R(\beta_2) \le \chi_p^2 (1 - 0.95)\}. \tag{4.1.4}$$

4.2 Empirical Likelihood Ratio Tests

An empirical likelihood ratio test of $H_0: \lambda = \lambda_0$ against $H_1: \lambda \neq \lambda_0$ is based on the Theorem (4.1.1). For given significance level 0.05, reject H_0 if $R(\lambda_0) > \chi_K^2(1-0.05)$. Similarly, an empirical likelihood ratio test of $H_0: \beta_2 = \beta_{20}$ vs $H_1: \beta_2 \neq \beta_{20}$ is based on the Theorem (4.1.2). For given significance level 0.05, reject H_0 if $R(\beta_{20}) > \chi_p^2(1-0.05)$.

4.3 Simulation Studies

In this section, we first demonstrate the distribution of the semiparametric empirical likelihood ratio statistics through Q-Q plots, then we study the proposed confidence intervals/regions through simulations. We will, for different situations, be focused on the length and coverage of confidence intervals. Powers of the likelihood ratio-based χ^2 -tests will not be reported.

4.3.1 Q-Q plots for Empirical Likelihood Ratio Statistics

For large sample sizes, we have shown that empirical likelihood ratio statistics for λ and β are both χ^2 distributed. In this section, we demonstrate the small-tomedium-sized sample property of those statistics via Q-Q plots. The procedure is as follows. we generate a set of three samples, each from one of the K mixtures, then use the proposed semiparametric method with these samples, and finally calculate the empirical likelihood ratio statistic. Repeating this procedure 1000 times to have 1000 replicates of the the likelihood ratio statistic. A Q-Q plot is then constructed by plotting the quantiles of those values against the standard χ^2 quantiles.

we consider three normal mixtures: 0.10N(0, 1)+0.90N(2, 1), 0.50N(0, 1)+0.50N(2, 1)and 0.90N(0, 1)+0.10N(2, 1). We generate 1000 sets of three samples of size 100 each from the three mixtures respectively. It's easy to see that $\beta_1 = -2$ and $\beta_2 = 2$. For $n_1 = n_2 = n_3 = 100$, at the true value $\lambda_1 = 0.10, \lambda_2 = 0.50$ and $\lambda_3 = 0.90$, the Q-Q plot for the 1000 replications of $R(\lambda)$ versus the standard χ_3^2 is shown in Figure (4.3.1). At the true value $\beta_2 = 2$, the Q-Q plot for the 1000 replications of $R(\beta_2)$ versus the standard χ_1^2 is shown in Figure (4-2). The χ^2 approximation to each of the two statistics appears satisfactory.

4.3.2 Coverage of Confidence Intervals

To evaluate the empirical likelihood ratio-based confidence regions for the mixing proportions λ and the parameter β_2 , we consider three normal mixtures: $\lambda_1 N(0, 1) + (1 - \lambda_1)N(\mu_2, 1)$, $\lambda_2 N(0, 1) + (1 - \lambda_2)N(\mu_2, 1)$ and $\lambda_3 N(0, 1) + (1 - \lambda_3)N(\mu_2, 1)$. For comparison we also consider fully parametric inferences; that is, F and G are $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. For the mixing proportions, confidence intervals are calculated semiparametrically (S) and parametrically (P). We report the



Figure 4-1: Q-Q plot for λ

simulation results in Table (4.1), where the nominal coverage level is 90%.

To evaluate the likelihood ratio test statistic for the parameter β_2 , We consider two sets of proportions and two values for μ_2 : 0.5 and 2.0 so that β_2 equals 0.5 and 2.0, respectively. We generate 1000 sets of three samples from the three mixtures respectively. For comparison we also consider the fully parametric case that F and Gare $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. The value of β_2 is $\mu_2 - \mu_1$. For β_2 , under a nominal coverage of 90%, confidence intervals are calculated semiparametrically (S) and parametrically (P). We report the simulation results in Table (4.2).

Method	λ	$\operatorname{Cov.}(\%)$	Av. Len.	Av. Midpt.
		$n_1 = n_2 = n_3 = 100$		
S	0.50, 0.60, 0.70	86.3	0.3544	0.4975
P	0.50, 0.60, 0.70	88.4	0.3316	0.4989
S	0.30, 0.60, 0.70	87.5	0.3284	0.2851
P	0.30, 0.50, 0.70	89.1	0.3113	0.2925
S	0.10, 0.40, 0.70	89.7	0.2905	0.1104
P	0.10, 0.40, 0.70	90.3	0.2853	0.0994
S	0.10, 0.50, 0.90	87.8	0.2962	0.0895
P	0.10, 0.50, 0.90	89.2	0.2901	0.1051

Table 4.1: Average length, midpoint and coverage for λ_1 from 1000 samples.

Cov. = Coverage, Av. Len. = Average Length, Av. Midpt. = Average Midpoint

Table 4.2: Average length, midpoint and coverage for β_2 from 1000 samples

β_2	λ	Method	$\operatorname{Cov.}(\%)$	Av. Len.	Av. Midpt.
		$n_1 = n_2 = n_3 = 100$			
0.5	0.40, 0.50, 0.60	S	86.2	0.4441	0.5154
0.5	0.40, 0.50, 0.60	Р	88.3	0.4211	0.4985
0.5	0.10, 0.50, 0.90	S	86.0	0.3986	0.4881
0.5	0.10, 0.50, 0.90	Р	91.4	0.3590	0.4918
2.0	0.40, 0.50, 0.60	S	87.7	1.2436	2.0123
2.0	0.40, 0.50, 0.60	Р	92.3	1.0183	2.0117
2.0	0.10, 0.50, 0.90	S	88.9	1.1953	2.0024
2.0	0.10, 0.50, 0.90	Р	91.1	0.9703	1.9902

Cov. = Coverage, Av. Len. = Average Length, Av. Midpt. = Average Midpoint



Figure 4-2: Q-Q plot for β_2

4.4 Proofs

4.4.1 Proof of Theorem (4.1.2)

. Let $\widetilde{\xi} = (\widetilde{\lambda} - \lambda_0, \widetilde{\beta}_1 - \beta_{10}, \widetilde{\alpha} - 0)^T$. We have

$$\begin{pmatrix} \tilde{\beta}_2 - \beta_{20} \\ \tilde{\xi} \end{pmatrix} = \begin{pmatrix} \tilde{\beta}_2 - \beta_{20} \\ \tilde{\lambda} - \lambda_0 \\ \tilde{\beta}_1 - \beta_{10} \\ \tilde{\alpha} - 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & I_p & 0 \\ I_K & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\lambda} - \lambda_0 \\ \tilde{\beta}_1 - \beta_{10} \\ \tilde{\beta}_2 - \beta_{20} \\ \tilde{\alpha} - 0 \end{pmatrix}$$
$$= -\begin{pmatrix} 0 & 0 & I_p & 0 \\ I_K & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} S^{-1}Q_n + o_p(n^{-1/2}).$$

Let $\hat{\lambda}$ and $\hat{\beta}_1$ maximize $l(\lambda, \beta_1, \beta_{20}, \alpha(\lambda, \beta_1, \beta_{20}))$. We have $\frac{\partial l}{\partial \lambda} = \frac{\partial l}{\partial \beta_1} = \frac{\partial l}{\partial \alpha} = 0$, all evaluated at $(\hat{\lambda}, \hat{\beta}_1, \beta_{20}, \hat{\alpha})$, where $\hat{\alpha} = \alpha(\hat{\lambda}, \hat{\beta}_1, \beta_{20})$. Expanding the above at $\theta_0 = (\lambda_0, \beta_{10}, \beta_{20}, 0)$ yields

$$\hat{\xi} \equiv \begin{pmatrix} \hat{\lambda} - \lambda_0 \\ \hat{\beta}_1 - \beta_{10} \\ \hat{\alpha} - 0 \end{pmatrix} = -C^{-1} \begin{pmatrix} \frac{\partial l}{\partial \lambda} \\ \frac{\partial l}{\partial \beta_1} \\ \frac{\partial l}{\partial \alpha} \end{pmatrix} + o_p(n_{-1/2}),$$

where

$$C = \begin{pmatrix} \frac{\partial^2 l}{\partial \lambda \partial \lambda^T} & \frac{\partial^2 l}{\partial \lambda \partial \beta_1} & \frac{\partial^2 l}{\partial \lambda \partial \alpha} \\ \frac{\partial^2 l}{\partial \beta_1 \partial \lambda^T} & \frac{\partial^2 l}{\partial \beta_1 \partial \beta_1} & \frac{\partial^2 l}{\partial \beta_1 \partial \alpha} \\ \frac{\partial^2 l}{\partial \alpha \partial \lambda^T} & \frac{\partial^2 l}{\partial \alpha \partial \beta_1} & \frac{\partial^2 l}{\partial \alpha^2} \end{pmatrix},$$

all partial derivatives being evaluated at θ_0 . Denote $Q_{1n} = \frac{1}{n} \frac{\partial l}{\partial \beta_2}$ and $Q_{2n} = (\frac{1}{n} \frac{\partial l}{\partial \lambda}, \frac{1}{n} \frac{\partial l}{\partial \beta_1}, \frac{1}{n} \frac{\partial l}{\partial \alpha})^T$, both evaluated at θ_0 . Note that

$$\begin{pmatrix} \beta_{20} - \tilde{\beta}_{2} \\ \hat{\xi} - \tilde{\xi} \end{pmatrix} = \begin{pmatrix} 0 \\ \hat{\xi} \end{pmatrix} - \begin{pmatrix} \tilde{\beta}_{2} - \beta_{20} \\ \tilde{\xi} \end{pmatrix}$$
$$= -\begin{pmatrix} 0 & 0 \\ 0 & C^{-1} \end{pmatrix} \begin{pmatrix} Q_{1n} \\ Q_{2n} \end{pmatrix} + \begin{pmatrix} 0 & 0 & I_{p} & 0 \\ I_{K} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} S^{-1}Q_{n} + o_{p}(n^{-1/2})$$

$$\begin{split} &= -\left(\begin{array}{ccc} 0 & 0 \\ 0 & C^{-1} \end{array}\right) \left(\begin{array}{ccc} 0 & 0 & I_p & 0 \\ I_K & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) Q_n + \left(\begin{array}{ccc} 0 & 0 & I_p & 0 \\ I_K & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) S^{-1}Q_n + o_p(n^{-1/2}) \\ &= \left[\left(\begin{array}{ccc} 0 & 0 \\ 0 & C^{-1} \end{array}\right) \left(\begin{array}{ccc} 0 & 0 & I_p & 0 \\ I_K & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) S \left(\begin{array}{ccc} 0 & 0 & I_p & 0 \\ I_K & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)^{-1} \\ &= \left(\begin{array}{ccc} I_p & 0 \\ 0 & I_{K+2} \end{array}\right) \left[\left(\begin{array}{ccc} \tilde{\beta}_2 - \beta_{20} \\ \tilde{\xi} \end{array}\right) + o_p(n^{-1/2}) \\ &= \left(\begin{array}{ccc} -I_p \\ C^{-1b} \end{array}\right) (\tilde{\beta}_2 - \beta_{20}) + o_p(n^{-1/2}), \end{split}$$

where $b^T = (s_{13}^T, s_{23}^T, s_{43}^T)$, and

$$\begin{pmatrix} \hat{\lambda} - \widetilde{\lambda} \\ \hat{\beta}_1 - \widetilde{\beta}_1 \\ \beta_{20} - \widetilde{\beta}_2 \\ \hat{\alpha} - \widetilde{\alpha} \end{pmatrix} = \begin{pmatrix} 0 & I_K & 0 & 0 \\ 0 & 0 & 1 & 0 \\ I_p & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_{20} - \widetilde{\beta}_2 \\ \hat{\xi} - \widetilde{\xi} \end{pmatrix}.$$

From Taylor expansion, $l(\hat{\lambda}, \hat{\beta}_1, \beta_2, \hat{\alpha})$ equals

$$\begin{split} &l(\widetilde{\lambda},\widetilde{\beta}_{1},\widetilde{\beta}_{2},\widetilde{\alpha}) + \frac{n}{2}((\widehat{\lambda} - \widetilde{\lambda})^{T},\widehat{\beta}_{1} - \widetilde{\beta}_{1},(\beta_{20} - \widetilde{\beta}_{2})^{T},\widehat{\alpha} - \widetilde{\alpha})S\begin{pmatrix} \widehat{\lambda} - \widetilde{\lambda} \\ \widehat{\beta}_{1} - \widetilde{\beta}_{1} \\ \beta_{20} - \widetilde{\beta}_{2} \\ \widehat{\alpha} - \widetilde{\alpha} \end{pmatrix} + o_{p}(1) \\ &= \frac{n}{2}((\beta_{20} - \widetilde{\beta}_{2})^{T},(\widehat{\xi} - \widetilde{\xi})^{T})\begin{pmatrix} s_{33} & b^{T} \\ b & C \end{pmatrix}\begin{pmatrix} \beta_{20} - \widetilde{\beta}_{2} \\ \widehat{\xi} - \widetilde{\xi} \end{pmatrix} + o_{p}(1) \\ &= \frac{n}{2}\Big[\begin{pmatrix} -I_{p} \\ C^{-1}a \end{pmatrix}(\widetilde{\beta}_{2} - \beta_{20})\Big]^{T}\begin{pmatrix} s_{33} & b^{T} \\ b & C \end{pmatrix}\Big[\begin{pmatrix} -I_{p} \\ C^{-1}a \end{pmatrix}(\widetilde{\beta}_{2} - \beta_{20})\Big] + o_{p}(1) \\ &= \frac{n}{2}(\widetilde{\beta}_{2} - \beta_{20})^{T}[s_{33} - b^{T}C^{T}b](\widetilde{\beta}_{2} - \beta_{20}) + o_{p}(1). \end{split}$$

Therefore, $l(\hat{\lambda}, \hat{\beta}_1, \beta_{20}, \hat{\alpha}) - l(\tilde{\lambda}, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\alpha})$ converges to $-\frac{1}{2}\chi^2_{p+1}$ in distribution, since the asymptotic variance of $\sqrt{n}(\tilde{\beta}_2 - \beta_{20})$ is $-(s_{33} - b^T C^T b)^{-1}$.

4.4.2 Proof of Theorem (4.1.1)

. Let $\widetilde{\xi} = (\widetilde{\beta} - \beta_0, \widetilde{\alpha} - 0)^T$. We have

$$\begin{pmatrix} \widetilde{\lambda} - \lambda_0 \\ \widetilde{\xi} \end{pmatrix} = -S^{-1}Q_n + o_p(n^{-1/2}) = \begin{pmatrix} s_{11} & c_{12} \\ c_{12}^T & c_{22} \end{pmatrix} \begin{pmatrix} Q_{1n} \\ Q_{2n} \end{pmatrix}.$$

where

$$c_{12} = (s_{12}, s_{13}), \quad c_{22} = \begin{pmatrix} s_{22} & s_{23} \\ s_{32} & s_{33} \end{pmatrix},$$

and

$$Q_{1n} = \frac{1}{n} \frac{\partial l}{\partial \lambda}, \quad Q_{2n} = (\frac{1}{n} \frac{\partial l}{\partial \beta}, \frac{1}{n} \frac{\partial l}{\partial \alpha})^T,$$

both evaluated at θ_0 . Denote the maximizer of $l(\lambda_0, \beta, \alpha(\lambda_0, \beta))$ by $\hat{\beta}$, which satisfies

$$\frac{\partial l}{\partial \beta}(\hat{\lambda_0}, \hat{\beta}, \hat{\alpha}) = \frac{\partial l}{\partial \alpha}(\hat{\lambda_0}, \hat{\beta}, \hat{\alpha}) = 0,$$

where $\hat{\alpha} = \alpha(\hat{\lambda}, \hat{\beta}_1, \beta_{20})$. Expanding the above equations at $\theta_0 = (\lambda_0, \beta_0, 0)$ yields

$$\widehat{\xi} \equiv \begin{pmatrix} \widehat{\beta} - \beta_0 \\ \widehat{\alpha} - 0 \end{pmatrix} = -c_{22}^{-1}Q_{2n} + o_p(n^{-1/2}).$$

Note that

$$\begin{pmatrix} \lambda_0 - \tilde{\lambda} \\ \hat{\xi} - \tilde{\xi} \end{pmatrix} = \begin{pmatrix} 0 \\ \hat{\xi} \end{pmatrix} - \begin{pmatrix} \tilde{\lambda} - \lambda_0 \\ \tilde{\xi} \end{pmatrix}$$

$$= -\begin{pmatrix} 0 & 0 \\ 0 & c_{22}^{-1} \end{pmatrix} \begin{pmatrix} Q_{1n} \\ Q_{2n} \end{pmatrix} + \begin{pmatrix} s_{11} & c_{12} \\ c_{12}^T & c_{22} \end{pmatrix}^{-1} \begin{pmatrix} Q_{1n} \\ Q_{2n} \end{pmatrix} + o_p(n^{-1/2})$$

$$= \left\{ \begin{pmatrix} 0 & 0 \\ 0 & c_{22}^{-1} \end{pmatrix} \begin{pmatrix} s_{11} & c_{12} \\ c_{12}^T & c_{22} \end{pmatrix} - \begin{pmatrix} I_K & 0 \\ 0 & I_{p+2} \end{pmatrix} \right\} \begin{pmatrix} \tilde{\lambda} - \lambda_0 \\ \tilde{\xi} \end{pmatrix} + o_p(n^{-1/2})$$

$$= \begin{pmatrix} -I_K \\ c_{22}^{-1}c_{12}^T \end{pmatrix} (\tilde{\lambda} - \lambda_0) + o_p(n^{-1/2}),$$

Expanding $l(\hat{\lambda_0}, \hat{\beta}, \hat{\alpha})$ at $(\widetilde{\lambda}, \widetilde{\beta}, \widetilde{\alpha})$ yields

$$\begin{split} l(\hat{\lambda_0}, \hat{\beta}, \hat{\alpha}) &- l(\widetilde{\lambda}, \widetilde{\beta}, \widetilde{\alpha}) = \frac{n}{2} (\hat{\lambda_0} - \widetilde{\lambda})^T, \hat{\beta} - \widetilde{\beta})^T, \hat{\alpha} - \widetilde{\alpha}) S \begin{pmatrix} \hat{\lambda_0} - \widetilde{\lambda} \\ \hat{\beta} - \widetilde{\beta} \\ \hat{\alpha} - \widetilde{\alpha} \end{pmatrix} + o_p(1) \\ \\ &= \frac{n}{2} ((\lambda_0 - \widetilde{\lambda})^T, (\hat{\xi} - \widetilde{\xi})^T) \begin{pmatrix} s_{11} & c_{12} \\ c_{12}^T & c_{22} \end{pmatrix} \begin{pmatrix} \lambda_0 - \widetilde{\lambda} \\ \hat{\xi} - \widetilde{\xi} \end{pmatrix} + o_p(1) \\ \\ &= \frac{n}{2} \Big[\begin{pmatrix} -I_K \\ c_{22}^{-1} c_{12}^T \end{pmatrix} (\widetilde{\lambda} - \widetilde{\lambda}) \Big]^T \begin{pmatrix} s_{11} & c_{12} \\ c_{12}^T & c_{22} \end{pmatrix} \Big[\begin{pmatrix} -I_K \\ c_{21}^{-1} c_{12}^T \end{pmatrix} (\widetilde{\lambda} - \widetilde{\lambda}) \Big] + o_p(1) \\ \\ &= \frac{n}{2} (\widetilde{\lambda} - \widetilde{\lambda})^T (s_{11} - c_{12} c_{22}^{-1} c_{12}^T) (\widetilde{\lambda} - \widetilde{\lambda}) + o_p(1) \end{split}$$

which converges to $-\frac{1}{2}\chi^2_{p+1}$ in distribution, since the asymptotic variance of $\sqrt{n}(\tilde{\lambda}-\lambda_0)$ is $-(s_{11}-c_{12}c_{22}^{-1}c_{12}^T)^{-1}$.

Chapter 5

A Kolmogrov-Smirnov-type GOF Test for the Density Ratio Model

5.1 Introduction

The estimation and inference procedure rely on whether or not the underlying model is valid. An invalid model basically leads to inconsistent estimators. The validity of a model can be assessed by goodness-of-fit tests. In this chapter, we propose an omnibus test which is of Kolmogorov-Smirnov (K-S) type. We first establish a weak convergence result, based on which a KS test is then proposed. The KS statistic measures the discrepancy between the empirical version and the model based estimates of the population distribution. The asymptotic analytic distribution of the test statistic is not available so the critical value and p-value are obtained through the bootstrap procedure. Finally, small sample properties are studied via simulation and real examples are presented. Denote $D_n(t) = \sqrt{n}(\tilde{H}_1(t) - \hat{H}_1(t), \dots, \tilde{H}_K(t) - \hat{H}_K(t))^T$. We first develop the weak convergence result.

5.2 Weak Convergence of $D_n(t)$

The construction of the K-S goodness-of-fit test is based on the following weak convergence theorem.

Theorem 5.2.1. Under Model (I) and suitable regularity conditions, as $n \to \infty$,

$$\sqrt{n} \begin{pmatrix} \widetilde{H}_{1}(t) - \widehat{H}_{1}(t) \\ \widetilde{H}_{2}(t) - \widehat{H}_{2}(t) \\ \vdots \\ \widetilde{H}_{K}(t) - \widehat{H}_{K}(t) \end{pmatrix} \rightarrow \begin{pmatrix} W_{1}(t) \\ W_{2}(t) \\ \vdots \\ W_{K}(t) \end{pmatrix} \equiv W(t)$$
(5.2.1)

weakly in $D[-\infty, \infty]$, where W(t) is a Gaussian process with mean zero and covariance structure $\Sigma(t, s)$, $t \leq s$ given in the proof.

5.3 A Kolmogorov-Smirnov Goodness-of-Fit Test

For large sample sizes, if the density ratio model is valid, both the model-based semiparametric estimator $\tilde{H}_k(x)$ and the model-free empirical estimator of $\hat{H}_k(x)$ should be close to $H_k(x)$, since both are consistent. Therefore, a formal goodness-offit test of the semiparametric density ratio model can be constructed based on some sort of distance. We consider the well known K-S type distance.

5.3.1 The K-S Test Statistic: Δ_n

For the classical K-S statistic, the construction is straightforward. For our case, however, the K distances involved should be combined to form a statistic in such a way that each distance contributes a suitable proportion to the constructed statistic. To construct such a statistic, we use the weighted average of those distances. For ease of exposition, we focus our attention on the test

$$H_0: \log \frac{g(x)}{f(x)} = \beta_1 + \beta_2 x \quad \text{vs} \quad H_1: \text{Not } H_0.$$
 (5.3.1)

Let $\Delta_{nk}(x) = \sqrt{n} [\widetilde{H}_k(x) - \widehat{H}_k(x)]$ and $\Delta_{nk} = \sup_{-\infty \le x \le \infty} |\Delta_{nk}(x)|$, for $k = 1, 2, \cdots, K$. Denote the K-S statistic as Δ_n , which is defined as

$$\Delta_n = \frac{1}{K} \sum_{k=1}^{K} \rho_{nk} \Delta_{nk}.$$
(5.3.2)

Remark 5.3.1. An optimal choice of the weights is open. By optimal, we mean that the choice has to be made so that the power of the proposed K-S test approaches to 1 for given an alternative as n gets large. To overcome this difficulty, we will propose some alternative goodness-of-fit tests in the future.

For the proposed test, a large value of Δ_n indicates that the density ratio model is inappropriate. Note that, by the Continuous Mapping Theorem (Billingsley, 1968), we have

$$\lim_{n \to \infty} P(\Delta_n \ge \delta_{1-q}) = \lim_{n \to \infty} P(\frac{1}{K} \sum_{k=1}^K \rho_{nk} \sup_{-\infty \le t \le \infty} \sqrt{n} |\Delta_{nk}(t)| \ge \delta_{1-q})$$
$$= P(\frac{1}{K} \sum_{k=1}^K \rho_k \sup_{-\infty \le t \le \infty} |W_k(t)| \ge \delta_{1-q})$$
$$= q,$$

thus, the q-quantile of Δ_n can be obtained through the limiting process when n is large. Since there is no explicit benchmark available, a resampling procedure such as bootstrap has to be called for.

5.3.2 A Bootstrap Procedure for Approximating the *P*-value

A bootstrap procedure is used to approximate the critical value of the proposed Kolmogorov-Smirnov-type goodness-of-fit test. We have developed semiparametric estimators $\tilde{H}_k(x)$ for the mixtures $H_k(x)$, $k = 1, 2, \dots, K$, which can be shown to be more efficient than their nonparametric counterparts, therefore, in stead of $\hat{H}_1(x), \dots, \hat{H}_K(x)$, the empirical distribution, we generate bootstrap samples from $\tilde{H}_1(x), \dots, \tilde{H}_K(x)$, respectively. To be specific, we go with the following loop.

STEP 1. Generate a bootstrap sample for each $k = 1, 2, \cdots, K$.

Let $x_{k1}^*, x_{k2}^*, \dots, x_{kn_k}^*$ be a random sample from $\widetilde{H}_k(x)$. Assume further that $x_{11}^*, x_{12}^*, \dots, x_{1n_1}^*, x_{21}^*, x_{22}^*, \dots, x_{2n_2}^*, \dots, x_{K1}^*, x_{K2}^*, \dots, x_{Kn_K}^*$ are jointly independent. Let $t_1^*, t_2^*, \dots, t_n^*$ be the pooled bootstrap sample, where $n = n_1 + n_2 + \dots + n_K$.

STEP 2. Repeat the estimation procedure proposed in Chapter two.

With the generated data in Step 1, repeat the empirical likelihood-based semiparametric procedure proposed in Chapter two to get the bootstrap version estimates of the parameters and the cumulative distribution functions. To be specific, let $(\tilde{\lambda}^*, \tilde{\beta}^*, \tilde{\alpha}^*)$ be solutions to the bootstrap version score equations. Then the bootstrap version estimates of F(x), G(x) and $H_k(x)$ for $k = 1, 2, \dots, K$, are constructed

as

$$\widetilde{F}^*(x) = \sum_{i=1}^n \widetilde{p}_i^* I(t_i^* \le x)$$

$$= \frac{1}{n} \sum_{i=1}^n \frac{1}{\gamma(t_i^*; \widetilde{\lambda}^*, \widetilde{\beta}^*)} \frac{1}{1 + \widetilde{\alpha}^* [\omega(t_i^*; \widetilde{\beta}^*) - 1] / \gamma(t_i^*; \widetilde{\lambda}^*, \widetilde{\beta}^*)} I(t_i^* \le x),$$

$$\widetilde{G}^*(x) = \sum_{i=1}^n \widetilde{p}_i^* \omega(t_i^*; \widetilde{\beta}^*) I(t_i^* \le x), \text{ and}$$

$$\widetilde{H}^*_k(x) = \sum_{i=1}^n \widetilde{p}_i^* [\widetilde{\lambda}^*_k + (1 - \widetilde{\lambda}^*_k) \omega(t_i^*; \widetilde{\beta}^*)] I(t_i^* \le x), \quad k = 1, \cdots, K,$$

where

$$\widetilde{p}_i^* = \frac{1}{n} \frac{1}{r(t_i^*; \widetilde{\lambda}^*, \widetilde{\beta}^*)} \frac{1}{1 + \widetilde{\alpha}^*[\omega(t_i^*; \widetilde{\beta}^*) - 1]/r(t_i^*; \widetilde{\lambda}^*, \widetilde{\beta}^*)},$$

where the functions $r(t; \lambda, \beta)$ and $\omega(t; \beta)$ are defined in Chapter two.

The empirical distribution function $\hat{H}_k^*(x)$, is formed corresponding to the bootstrap sample $x_{k1}^*, x_{k2}^*, \dots, x_{kn_k}^*$.

Step 3. Calculate Δ_{nk}^* , the bootstrap version of Δ_{nk} .

We define the bootstrap version of $\Delta_{nk}(x)$ and Δ_{nk} as

$$\Delta_{nk}^*(x) = \sqrt{n} [\widetilde{H}_k^*(x) - \widehat{H}_k^*(x)]$$
$$\Delta_{nk}^* = \sup_{-\infty \le x \le \infty} |\Delta_{nk}^*(x)|,$$

Step 4. Calculate Δ_n^* , the bootstrap version of Δ_n .

We weight Δ_{nk}^* , $k = 1, 2, \dots, K$, to get the bootstrap version of Δ_n^* as

$$\Delta_n^* = \frac{1}{K} \sum_{k=1}^K \rho_{nk} \Delta_{nk}^*.$$

Step 5. Repeat the above procedure B = 1000 times to get Δ_n^{*b} , $b = 1, 2, \dots, B$. The quantile of Δ_n is then approximate by that of Δ_n^{*b} , $b = 1, 2, \dots, B$. We use the following

$$\frac{\#\{\Delta_n^{*b} \ge \Delta_n\}}{B}$$

to approximate the P-value of the test, where Δ_n is the observed value of the test statistic based on the original samples.

5.3.3 The Power under Local Alternatives

Suppose the two densities in the density ratio model indeed are related by

$$\log \frac{g(x)}{f(x)} = \beta_1 + \beta_2 x + \gamma x^2.$$
 (5.3.3)

We consider the normal densities for both g(x) and f(x) and K = 2 or 3. Given β_2 , for some choices of γ , the corresponding β_1 can be determined. The power of the test (5.3.1) under the above local alternative can be simulated as in the following procedure:

 For given λ_k's and β₂, for each choice of γ, generate a set of K samples, each from H₁(x), H₂(x), · · · , H_K(x).

- Use the estimation procedure in Chapter two on the data generated above, and find the *P*-value as described in last subsection. A *P*-values smaller than a nominal significance level, say 0.05, indicates rejection of hypothesis (5.3.1).
- 3. Repeat the first two steps r = 1000 times and find the number of rejections in the r simulations, say r_0 . The power is approximated by r_0/r .

5.3.4 A Comparison of Powers: K-S Test against t Test

If the model (5.3.3) is true, then testing the density ratio assumption amounts to testing that $\gamma = 0$. We compare the power of the proposed K-S test with that of the t test. The t statistic, T, is constructed as follows:

$$T = \sqrt{n} \frac{\tilde{\gamma}}{\tilde{\sigma}_{\gamma}},$$

where $\tilde{\sigma}_{\gamma}$ is the large sample standard deviation of $\tilde{\gamma}$, with unknown quantities replaced by consistent estimates. Under the null hypothesis that $\gamma = 0, T$ is approximately normally distributed. The critical value of T is thus the standard quantile for a given significance level. On the other hand, the distribution of T is unknown under the alternative that $\gamma \neq 0$, therefore, to find the power of the t test, we shall use simulation methods. Specifically, We take f(x) to be the standard normal density, and β_2 to be fixed as 0.5, so that, under the model (5.3.3), g(x) is a normal density $N(\mu, \sigma^2)$ with

$$\mu = \frac{\beta_2}{1 - 2\gamma}, \quad \sigma^2 = \frac{1}{1 - 2\gamma}$$

We report the achieved significance levels and powers in Table (5.1), for given mixing proportions $\lambda_1 = 0.2$ and $\lambda_2 = 0.7$. In the table, Δ is the K-S type GOF statistic, and T is the t statistic. When the density ratio assumption is true, the achieved significance levels are close to the nominal level. The farther the γ is away from 0, the greater the power of the test to detect the departure. Also, the power of the test is, as expected, less than that of the t test.

Table 5.1: Powers for testing $\gamma = 0$ when $\psi(x; \beta_2) = \beta_2 x + \gamma x^2$ with $\beta_2 = 0.5$

γ	sample sizes	Levels	Δ	T
0.0	$n_1 = n_2 = 100$	0.10	0.145	0.142
0.0	$n_1 = n_2 = 100$	0.05	0.081	0.073
0.0	$n_1 = n_2 = 100$	0.01	0.021	0.012
0.0	$n_1 = n_2 = 150$	0.10	0.156	0.137
0.0	$n_1 = n_2 = 150$	0.05	0.134	0.129
0.0	$n_1 = n_2 = 150$	0.01	0.029	0.025
-1.0	$n_1 = n_2 = 100$	0.10	0.577	0.621
-1.0	$n_1 = n_2 = 100$	0.05	0.486	0.511
-1.0	$n_1 = n_2 = 100$	0.01	0.289	0.324
-1.0	$n_1 = n_2 = 150$	0.10	0.833	0.887
-1.0	$n_1 = n_2 = 150$	0.05	0.703	0.748
-1.0	$n_1 = n_2 = 150$	0.001	0.501	0.601

5.4 Examples

Example 5.4.1. We continue the example of Nagelkerke et al (2001) studied in Chapter 3.

We evaluated the goodness-of-fit of the semiparametric model to the three year data. We found that the p-values for the three years are 0.044, 0.221 and 0.002, respectively. Therefore, the model fits well for year 90 data only.

Example 5.4.2. We continue the example of Thompson et al (1998) discussed in Chapter 3.

The p-value for the data is 0.017. Therefore, the model doesn't fit well based on our proposed test.

5.5 Proofs

Proof of theorem (5.2.1). According to (2.4.10), it suffices to show that $\sqrt{n}(H_{11}(t) + H_1(t) - \hat{H}_1(t), \cdots, H_{1K}(t) + H_K(t) - \hat{H}_K(t))^T \to W(t)$ weakly in $D[-\infty, \infty]$. It's easy to see that, for $k = 1, 2, \cdots, K$, $E\{\sqrt{n}[H_{1k}(t) + H_k(t) - \hat{H}_k(t)]\} = 0$. Moreover,

$$\begin{aligned} &\cos(\sqrt{n}\{\widetilde{H}_{i}(t) - \widehat{H}_{i}(t)\}, \sqrt{n}\{\widetilde{H}_{j}(s) - \widehat{H}_{j}(s)\}) \\ &= n\cos(\{\widetilde{H}_{i}(t) - H_{i}(t)\} - \{\widehat{H}_{i}(t) - H_{i}(t)\}, \{\widetilde{H}_{j}(s) - H_{j}(s)\} - \{\widehat{H}_{j}(s) - H_{j}(s)\}) \\ &= n\cos(\{\widetilde{H}_{i}(t) - H_{i}(t)\}, \{\widetilde{H}_{j}(s) - H_{j}(s)\}) - n\cos(\{\widetilde{H}_{i}(t) - H_{i}(t)\}, \{\widehat{H}_{j}(s) - H_{j}(s)\}) \\ &- n\cos(\{\widehat{H}_{i}(t) - H_{i}(t)\}, \{\widetilde{H}_{j}(s) - H_{j}(s)\}) + n\cos(\{\widehat{H}_{i}(t) - H_{i}(t)\}, \{\widehat{H}_{j}(s) - H_{j}(s)\}), \end{aligned}$$

$$\begin{aligned} &\operatorname{cov}(\{\widetilde{H}_{i}(t) - H_{i}(t)\}, \{\widetilde{H}_{j}(s) - H_{j}(s)\}) \\ &= \operatorname{cov}(\frac{1}{n}\sum_{k=1}^{K}\sum_{l=1}^{n_{k}} \left[\varepsilon_{k}^{H_{i}}(x_{kl};t) + v_{2i}(t)q_{k}(x_{kl})\right], \frac{1}{n}\sum_{k=1}^{K}\sum_{l=1}^{n_{k}} \left[\varepsilon_{k}^{H_{j}}(x_{kl};s) + v_{2j}(s)q_{k}(x_{kl})\right]) \\ &= \frac{1}{n^{2}}\sum_{k=1}^{K}n_{k}\operatorname{cov}(\varepsilon_{k}^{H_{i}}(X_{k};t) + v_{2i}(t)q_{k}(X_{k}), \varepsilon_{k}^{H_{j}}(X_{k};s) + v_{2j}(s)q_{k}(X_{k}))) \\ &\operatorname{cov}(\{\widetilde{H}_{i}(t) - H_{i}(t)\}, \{\widehat{H}_{j}(s) - H_{j}(s)\}) = \frac{1}{n}\operatorname{cov}(\varepsilon_{j}^{H_{i}}(X_{j};t) + v_{2i}(t)q_{k}(X_{j}), I(X_{j} \leq s))) \\ &\operatorname{cov}(\{\widehat{H}_{i}(t) - H_{i}(t), \{\widetilde{H}_{j}(s) - H_{j}(s)\}) = \frac{1}{n}\operatorname{cov}(\varepsilon_{i}^{H_{j}}(X_{i};s) + v_{2j}(s)q_{i}(X_{j}), I(X_{i} \leq t))) \\ &\operatorname{cov}(\{\widehat{H}_{i}(t) - H_{i}(t)\}, \{\widehat{H}_{j}(s) - H_{j}(s)\}) = \frac{1}{n_{i}}\{H_{i}(t \wedge s) - H_{i}(t)H_{k}(s)\}I(i = j), \end{aligned}$$

where I(A) is the indicator function of A.

Therefore, for any $i, j = 1, 2, \cdots, K$, $\operatorname{cov}(\sqrt{n}\{\widetilde{H}_i(t) - \widehat{H}_i(t)\}, \sqrt{n}\{\widetilde{H}_j(s) - \widehat{H}_j(s)\}) = E\{W_i(s)W_j(t)\}$. This, together with the CLT for sample means and the Cramer-Wold device, implies that the finite dimensional distribution of $\sqrt{n}\{H_{1k} + H_k(t) - \widehat{H}_k(t)\}$ converges weakly to those of W(t). Thus, in order to show weak convergence of $(\widetilde{H}_1(t) - \widehat{H}_1(t), \cdots, \widetilde{H}_K(t) - \widehat{H}_K(t))^T$, it is enough to show that the process $\{\sqrt{n}\{H_{1k} + H_k(t) - \widehat{H}_k(t)\}, t \in [-\infty, \infty]\}$ is tight in $D[-\infty, \infty]$. However, this can be proved by the tightness criteria in Billingsley (1968, Ch.3). This completes the proof.

Another way to express the variance structure is to use the Theorem (2.3.1), according to which, it suffices to show that $\sqrt{n}(H_{11}(t) - \hat{H}_1(t) - H_{12}(t), \cdots, H_{K1}(t) - \hat{H}_K(t) - H_{K2}(t))^T \rightarrow W(t)$ weakly in $D[-\infty, \infty]$. It's easy to see that, for $k = 1, 2, \cdots, K$, $E(\sqrt{n}[H_{k1}(t) - \hat{H}_k(t) - H_{k2}(t)]) = 0$. To calculate $\operatorname{cov}(\sqrt{n}[H_{k1}(s) - H_{k2}(s)]) = 0$.

and

$$\widehat{H}_k(s) - H_{k2}(s)$$
, $\sqrt{n}[H_{l1}(t) - \widehat{H}_l(t) - H_{l2}(t)]$, we proceed as follows.

$$\begin{aligned} &\cos(\sqrt{n}[H_{k1}(s) - \hat{H}_{k}(s) - H_{k2}(s)], \sqrt{n}[H_{l1}(t) - \hat{H}_{l}(t) - H_{l2}(t)]) \\ &= &\cos(\sqrt{n}[H_{k1}(s) - \hat{H}_{k}(s)], \sqrt{n}[H_{l1}(t) - \hat{H}_{l}(t)]) - &\cos(\sqrt{n}[H_{k1}(s) - \hat{H}_{k}(s)], \sqrt{n}[H_{l2}(t)]) \\ &- &\cos(\sqrt{n}[H_{l1}(t) - \hat{H}_{l}(t)], \sqrt{n}[H_{k2}(s)]) + &\cos(\sqrt{n}[H_{k2}(s)], \sqrt{n}[H_{l2}(t)]), \end{aligned}$$

and

$$\begin{aligned} &\operatorname{cov}(\sqrt{n}[H_{k1}(s) - \widehat{H}_{k}(s)], \sqrt{n}[H_{l1}(t) - \widehat{H}_{l}(t)]) \\ &= n \sum_{m=1}^{K} n_{m} \operatorname{cov}\left\{ \left[\frac{\omega_{k}}{nr} - \frac{I(m=k)}{n_{k}} \right] I(X_{m} \leq s), \left[\frac{\omega_{l}}{nr} - \frac{I(m=l)}{n_{l}} \right] I(X_{m} \leq t) \right\} \\ &= n \sum_{m=1}^{K} n_{m} \int_{-\infty}^{s \wedge t} \left[\frac{\omega_{k}}{nr} - \frac{I(m=k)}{n_{k}} \right] \left[\frac{\omega_{l}}{nr} - \frac{I(m=l)}{n_{l}} \right] \omega_{m} dF \\ &- n \sum_{m=1}^{K} n_{m} \int_{-\infty}^{s} \left[\frac{\omega_{k}}{nr} - \frac{I(m=k)}{n_{k}} \right] \omega_{m} dF \int_{-\infty}^{t} \left[\frac{\omega_{l}}{nr} - \frac{I(m=l)}{n_{l}} \right] \omega_{m} dF \\ &= - \int_{-\infty}^{s \wedge t} \frac{\omega_{k}\omega_{l}}{r} dF + \frac{1}{\rho_{k}} H_{k}(s \wedge t) I(k=l) - \sum_{m=1}^{K} \rho_{m} \int_{-\infty}^{s} \frac{\omega_{k}\omega_{m}}{r} dF \int_{-\infty}^{t} \frac{\omega_{l}\omega_{m}}{r} dF \\ &+ H_{l}(t) \int_{-\infty}^{s} \frac{\omega_{k}\omega_{l}}{r} dF + H_{k}(s) \int_{-\infty}^{t} \frac{\omega_{l}\omega_{k}}{r} dF - \frac{1}{\rho_{k}} H_{k}(s) H_{k}(t) I(k=l) \end{aligned}$$

$$= -G(s \wedge t) - (1 - \xi - \lambda_k - \lambda_l)(G(s \wedge t) - F(s \wedge t)) - (1 - \xi - \lambda_k)(1 - \xi - \lambda_l)\tau(s \wedge t) + H_l(t)[-G(s) - (1 - \xi - \lambda_k - \lambda_l)(G(s) - F(s)) - (1 - \xi - \lambda_k)(1 - \xi - \lambda_l)\tau(s)] - \sum_{m=1}^{K} \rho_m[-G(s) - (1 - \xi - \lambda_k - \lambda_m)(G(s) - F(s)) - (1 - \xi - \lambda_k)(1 - \xi - \lambda_m)\tau(s)] [-G(t) - (1 - \xi - \lambda_l - \lambda_m)(G(t) - F(t)) - (1 - \xi - \lambda_l)(1 - \xi - \lambda_m)\tau(t)] + \frac{1}{\rho_k}[H_k(s \wedge t) - H_k(s)H_k(t)]I(k = l).$$

Using (2.4.9), we have

$$\operatorname{cov}(\sqrt{n}[H_{k1}(s) - \widehat{H}_{k}(s)], \sqrt{n}[H_{l2}(t)])$$

$$= \begin{pmatrix} [G(t) - F(t) + (1 - \lambda_{l} - \phi)\tau(t)]\rho \\ (1 - \lambda_{l} - \phi)\varsigma_{1}(t) \\ -G(t) + F(t) - (1 - \lambda_{l} - \phi)\tau(t) \end{pmatrix}^{T} S^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix},$$

where

$$a = -\rho_n [G(s) - F(s) + (1 - \xi - \lambda_k)\tau(s)] + [G(s) - F(s)]e_k + \rho_n \tau (1 - \lambda_k - \xi)H_k(s)$$

$$- \rho_n \tau \sum_{m=1}^K \rho_m (1 - \xi - \lambda_m)[G(s) + (1 - \lambda_k - \lambda_m - \xi)(G(s) - F(s))]$$

$$b = -(1 - \lambda_k - \xi)\varsigma_1(s) + \varsigma_1(1 - \lambda_k - \xi)H_k(s)$$

$$- \varsigma_1 \sum_{m=1}^K \rho_m (1 - \xi - \lambda_m)[G(s) + (1 - \lambda_k - \lambda_m - \xi)(G(s) - F(s))]$$

$$c = \tau (1 - \lambda_k - \xi)H_k(s) - \tau \sum_{m=1}^K \rho_m (1 - \xi - \lambda_m)[G(s) + (1 - \lambda_k - \lambda_m - \xi)(G(s) - F(s))],$$

 $\begin{aligned} \cos(\sqrt{n}[H_{k2}(s)], \sqrt{n}[H_{l2}(t)]) \\ &= \begin{pmatrix} [G(s) - F(s) + (1 - \lambda_k - \phi)\tau(s)]\rho \\ (1 - \lambda_k - \phi)\varsigma_1(s) \\ -G(s) + F(s) - (1 - \lambda_k - \phi)\tau(s) \end{pmatrix}^T S^{-1}VS^{-1} \begin{pmatrix} [G(t) - F(t) + (1 - \lambda_l - \phi)\tau(t)]\rho \\ (1 - \lambda_l - \phi)\varsigma_1(t) \\ -G(t) + F(t) - (1 - \lambda_l - \phi)\tau(t) \end{pmatrix} \\ &= \begin{pmatrix} [G(s) - F(s) + (1 - \lambda_k - \phi)\tau(s)]\rho \\ (1 - \lambda_k - \phi)\varsigma_1(s) \end{pmatrix}^T \Sigma_{11} \begin{pmatrix} [G(t) - F(t) + (1 - \lambda_l - \phi)\tau(t)]\rho \\ (1 - \lambda_l - \phi)\varsigma_1(t) \end{pmatrix} \\ &+ \sigma_{22}[-G(s) + F(s) - (1 - \lambda_k - \phi)\tau(s)]^2, \quad \text{using } (2.4.8). \end{aligned}$

Chapter 6

Conclusions and Future Research

In this dissertation work, we discussed K mixtures from each of which a sample is available. we have developed an empirical likelihood-based estimation and inference procedure for parameters of interest. By noting that, under semiparametric models, estimation and inference may be more robust than those under parametric models and more efficient than those under nonparametric models, we impose in our procedure an assumption that the two component densities differ by a parametric multiplicative factor. This assumption is called a density ratio model or exponential tilting model, which includes many pairs of familiar densities as special cases. K mixtures are useful in a variety of contexts such as Genetics and Life testing. In Genetics, the mixing proportions may be found under some genetic assumptions. Zou, et al (2002) considered this situation with known mixing proportions using empirical likelihood. Our procedure considered the more general case. Moreover, we proposed EM/MM algorithms to estimate unknown quantities and a formal goodness of fit test to check the validity of the postulated density ratio model. We noted that the Newton-Raphson method is not a good choice for high dimensional problem, although it's well known for its quadratic convergence rate. Besides the KS goodness of fit procedure, other procedures of goodness of fit are possible. We will consider the following questions in the future.

- Base estimation and inference on the partial likelihood and study the relative efficiency of partial likelihood-based estimators to corresponding estimators under the full empirical likelihood;
- Develop alternative GOF tests, such as the information test and the χ² test;
 Develop a global test under the setting of Zou, Fine and Yandell (2002);
- Study the effect of misspecification of $\varphi(x; \beta_2)$;
- Use two-sample two-component mixture models to tackle the verification bias problem in diagnostic medicine;
- Find applications of the K-sample 2(3)-component mixture models in quantitative genetics.
References

- Anderson, J.A. (1979). Multivariate logistic compounds. *Biometrika*. 66, 7-16.
- Choi, S.C. (1979). Two sample tests for compound distributions for homogeneity of mixing proportions. *Technometrics.* 21, 361-365.
- Doerge, R. W., Zeng, Z. B., and Weir, B. S. (1997). Statistical issues in the search for genes affecting quantitative traits in experimental populations. *Statistical Science.* 12, 195-219.
- Everitt, B.S. and Hand, D.J. (1999). *Finite mixture distributions*. London: Chapman and Hall.
- Fine, J. P., Zou, F. and Yandell, B., (2004). Nonparametric estimation of the effects of quantitative trait loci. *Biostatistics.* 5, 501-513.
- Gilbert, P., Lele, S. and Vardi, Y. (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*.86, 27-43.
- Hall, P. (1981). On the nonparametric estimation of mixture proportions. JRSSB.43, 147-156.

- Hall, P. and Titterington, D.M.(1984). Efficient nonparametric estimation of mixture proportions. JRSSB. 46, 465-473.
- Heller, G., and Qin, J. (2001). Pairwise rank-based likelihood for estimation and inference on the mixture proportion. *Biometrics.* 57, 813-817
- Hosmer, D.W. (1973). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. *Biometrics*, 29, 761-770.
- Hsieh, D. A., Manski, C. F. and McFadden, D. (1985). Estimation of response probabilities from augmented retrospective observations. J. Amer. Statist. Assoc.80, 651-662.
- Kay, and Little, (1987) Transformations of the expanatory variables in the logistic regression models for binary data. *Biometrika*. **74**, 495-501.
- Lange, K. (2004). Optimization. Springer.
- McLachlan, G.J., Lawoko, C.R.O. and Ganesalingam, S. (1982). On the likelihood ratio test for compound distributions for homogeneity of mixing proportions. Technometrics. 24, 331-334.
- McLachlan, G.J., and Krishnan, T. (1997). The EM algorithm and extentions. New York: Wiley.
- McLachlan, G.J., and Basford, K.E. (1988). *Mixture models: inference and applications to clustering.* New York: Marcel Dekker.

- Murry, G.D. and Titterington, D.M. (1978). Estimation problems with data from a mixture. Appl. Statist. 27, 325-334
- Nagelkerke N.J.D., Martien, W.B. and Kim, S.J. (2001). Logistic discrimination of mixtures of M. tuberculosis and non-specific tuberculin reactions. *Statistics in Medicine.* 20, 1113-1124.
- Neuenschwander, B., Kim, S., Zwahlen, M., Engel, R., Rieder, H. (2000). Trends in the prevalence of infection with Mycobacterium tuberculosis in Korea from 1965 to 1995 – an analysis of seven surveys by mixture models. *International Journal of Tuberculosis and Lung Disease.* 4, 719-729.
- Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*. **75**, 237-249.
- Owen, A. B. (1990). Empirical likelihood confidence regions. Ann. Statist. 18, 90-120.
- Owen, A. B. (2001). Empirical likelihood. Wiley.
- Prentice, R. L. and Pyke, R. (1979). Logistic disease incidence models and casecontrol studies. *Biometrika*.66, 403-411.
- Qin, J. (1993). Empirical likelihood in biased sample problems. Ann. Statist. 21, 1182-96.
- Qin, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika*. 85, 619-630.

- Qin, J. (1999). Empirical likelihood ratio based confidence intervals for mixture proportions. The Annals of Statistics. 99, 1368-1384.
- Qin, J. (2000). Combine parametric and empirical likelihoods. Biometrika. 87, 484-490.
- Qin, J., Leung, D. (2005). A Semiparametric Two-Component Compound Mixture Model and Its Application to Estimating Malaria Attributable Fractions. *Biometrics.* 61, 456-464.
- Qin, J., Leung, D., and Shao, J. (2002). Estimation with survey data under nonignorable nonresponse or informative sampling. JASA. 97, 193-200
- Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating functions. Ann. Statist. 22, 300-325.
- Qin, J. and Zhang, B. (1997). A goodness of .t test for logistic regression models based on case-control data. *Biometrika*. 84, 609-618.
- Qin, J. and Zhang, B. (2005). Marginal likelihood, conditional likelihood and empirical likelihood: Connections and applications. *Biometrika*.92, 251-270.
- Smith, T. and Vounatsou, P. (1997). Logistic regression ang latent class models for estimating positives in diagnostic assays with poor resolution Communications in Statistics, theory and methods. 27, 303-319.
- Vounatsou, P., Smith, T. and Smith, A.F.M. (1998). Bayesian analysis of twocomponent mixture distributions applied to estimating malaria attributable

fractions. Appl. Statist. 47, 575-587.

- Titterington, D.M., Smith, A.F.M., and Makov, U.E. (1985). Statistical analysis of finite mixture distributions. New York: Wiley.
- Vardi, Y. (1982). Nonparametric estimation in presence of length biasthe Annals of Statistics. 10, 616-620.
- Wacholder, S. and Weinberg, C. R. (1994). Flexible maximum likelihood methods for assessing joint e.ects in case-control studies with complex sampling. *Biometrics*. 50, 350-357.
- Weinberg, C. R. and Sandler, D. P. (1991). Randomized recruitment in case-control studies. Amer. J. Epidemiology. 134, 421-433.
- Weinberg, C. R. and Wacholder, S. (1993). Prospective analysis of case-control data under general multiplicative-intercept risk models. *Biometrika*. 80, 461-465.
- Zhang, B. (2000). A goodness of fit test for multiplicative-intercept risk models based on case-control data. *Statistica Sinica.* 10, 839-865.
- Zhang, B. (2002a). An em algorithm for a semiparametric finite mixture model. Journal of Statistical Computation and Simulation. 72, 791-802.
- Zhang, B. (2002b). Assessing goodness-of-fit of generalized logit models based on case-control data. Journal of Multivariate Analysis. 82, 17-38.
- Zou, F., Fine, J. P. (2002). A note on a partial empirical likelihood. *Biometrika*.89, 958-961.

Zou, F., Fine, J. P. and Yandell, B. S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika.* **89**, 61-75.