Empirical Likelihood Inference for Two-Sample Problems

by

Ying Yan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Statistics

Waterloo, Ontario, Canada, 2010

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

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

Abstract

In this thesis, we are interested in empirical likelihood (EL) methods for two-sample problems, with focus on the difference of the two population means. A weighted empirical likelihood method (WEL) for two-sample problems is developed . We also consider a scenario where sample data on auxiliary variables are fully observed for both samples but values of the response variable are subject to missingness. We develop an adjusted empirical likelihood method for inference of the difference of the two population means for this scenario where missing values are handled by a regression imputation method. Bootstrap calibration for WEL is also developed. Simulation studies are conducted to evaluate the performance of naive EL, WEL and WEL with bootstrap calibration (BWEL) with comparison to the usual two-sample t-test in terms of power of the tests and coverage accuracies. Simulation for the adjusted EL for the linear regression model with missing data is also conducted.

Acknowledgments

I would like to thank my supervisor Professor Changbao Wu, who devoted much time, patient guidance and great effort on my master's work, and encouraged me to overcome lots of obstacles on my statistical studies.

I would also like to thank Professors Grace Yi and Yulia Gel for their time in reading this thesis and their helpful suggestions.

Dedication

This is dedicated to my parents, sisters and Min.

Contents

List of Tables

Chapter 1

Introduction

Two-sample problems are common in applied statistics. Suppose we have two study populations and our interest is the difference between the two populations. Particularly we are interested in the difference of the two population means. For example, Zhou, Gao and Hui (1997) studied the effects of race on medical costs of patients. Their interest is whether the average medical costs for African American patients is the same as white patients. In this thesis, we focus on comparing the difference of two population means.

1.1 Two-sample problems

Suppose $\{y_{11}, \cdots, y_{1n_1}\}$ is a random sample from a population following the distribution F_0 and $\{y_{21}, \cdots, y_{2n_2}\}$ is a random sample from a population following the distribution G_0 and the two samples are independent. Let μ_1 and μ_2 be respectively the two population means and σ_1^2 and σ_2^2 be respectively the two population variances. Let $d = \mu_1 - \mu_2$. Let $\bar{y}_{i.} = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$, $i = 1, 2$ be the two sample means and $S_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$, $i = 1, 2$, be the two sample variances. We are interested in making inference on d. More specifically, we are interested in testing the hypothesis

$$H_0 : d = 0,$$

or equivalently,

$$H_0 : \mu_1 = \mu_2.$$

When F_0 and G_0 are from the same location-scale family of distributions, this two-sample problem is called the Behrens-Fisher problem. The restriction to a normal distribution is often made. In this case, the variances σ_1^2 and σ_2^2 for two samples are not necessarily the same.

1.2 Parametric methods

In this section we provide a brief review of several classic parametric methods to deal with two-sample problems.

Large sample test Under the null hypothesis $H_0 : \mu_1 = \mu_2$, we know that

$$T_1 = \frac{\bar{y}_{1.} - \bar{y}_{2.}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \to N(0, 1)$$

when both n_1 and n_2 go to infinity. As a result, we can construct the large sample test which has the rejection region $\{|T_1| > z_{\alpha/2}\}$. Here $z_{\alpha/2}$ is the $(1 - \alpha/2)$th quantile of the standard normal distribution.

Student's t-test and likelihood ratio test Suppose F_0 and G_0 are normal distributions, that is, $y_{11}, \cdots, y_{1n_1} \sim N(\mu_1, \sigma_1^2)$ and $y_{21}, \cdots, y_{2n_2} \sim N(\mu_2, \sigma_2^2)$, and the two population variances σ_1^2 and σ_2^2 are the same, say $\sigma_1^2 = \sigma_2^2 = \sigma^2$. we can use the parametric two-sample t-test. Define the pooled variance estimator as

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

and the Student's t-statistic is defined as

$$T_2 = \frac{\bar{y}_{1.} - \bar{y}_{2.}}{S \times \sqrt{(\frac{1}{n_1} + \frac{1}{n_2})}},$$

then

$$T_2 \sim t_{n_1+n_2-2},$$

and the standard two-sample t-test has the rejection region as $\{|T_2| > t_{n_1+n_2-2}(\alpha/2)\}$, where α is the size of the test and $t_{n_1+n_2-2}(\alpha/2)$ is the $(1-\alpha/2)$th quantile of a t-distribution with n_1+n_2-2 degrees of freedom. We reject H_0 if $|T_2| > t_{n_1+n_2-2}(\alpha/2)$.

The likelihood ratio test under the assumption that $\sigma_1^2 = \sigma_2^2$ will lead to the same result as the Student's t-test. Details can be found in standard textbooks.

Test for the Behrens-Fisher problem The most popular solution for the Behrens-Fisher problem was proposed by B.L. Welch (1938). He considered the statistic

$$T_1 = \frac{\bar{y}_{1.} - \bar{y}_{2.}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}},$$

which is the same as the one for the large sample test. However, there is a major difference in terms of limiting distributions. The key idea is that the distribution of $S_1^2/n_1 + S_2^2/n_2$ can be approximated by a scaled chi-squared distribution. The number of degrees of freedom can be estimated by using the Welch-Satterthwaite equation

$$\hat{v} = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{S_1^4/[n_1^2(n_1 - 1)] + S_2^4/[n_2^2(n_2 - 1)]}.$$

As a result, the distribution of T_1 can be approximated by a t distribution. When the sample sizes are small or the two population variances are equal or approximately equal, then the standard two-sample t-test is more accurate.

Chapter 2

Two-Sample Empirical Likelihood Methods

The empirical likelihood (EL) method, proposed by Owen (1988), has been one of the most popular inference methods in the last 20 years. It was first introduced to construct confidence intervals for a population mean of a single sample of independent and identically distributed observations. Owen (2001) provides an excellent overview of the EL method, including intuitive ideas, theoretical developments, and applications. In the section, we develop empirical likelihood methods for two-sample problems. We first provide a short review of the EL method for a single sample and then describe a naive EL method and establish a weighted EL method for two-sample problems. Performances of these methods are compared through simulation studies.

2.1 A brief review of empirical likelihood for a single sample

Suppose that we are interested in testing the population mean of one single sample using empirical likelihood, which was first proposed by Owen (1988). Let $X_1, \cdots, X_n$ be a random sample from cumulative distribution function (CDF) F_0 and $x_1, \cdots, x_n$ is the realization of the sample. Suppose the population mean is μ. The empirical cumulative distribution function (ECDF) of $X_1, ..., X_n$ is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq x\},$$

for $x \in \mathbb{R}$. The nonparametric likelihood of any CDF F (not necessarily F_0) is defined as

$$L(F) = \prod_{i=1}^{n} Pr(X_i = x_i).$$

Define $\mathcal{F}$ as the class of all distribution functions. Then from Kiefer and Wolfowitz (1956), the ECDF F_n uniquely maximizes $L(F)$. That is, for any $F \neq F_n$, $L(F) < L(F_n) = (\frac{1}{n})^n$. As a result, $L(F_n) = \sup\{L(F) : F \in \mathcal{F}\}$.

Consequently, F_n maximizes the nonparametric likelihood $L(F) = \prod_{i=1}^{n} p_i$ over $p_i \geq 0, i = 1, \cdots, n$, and $\sum_{i=1}^{n} p_i = 1$, where $p_i = Pr(X_i = x_i)$. That is,

$$L(F_n) = \sup\{L(F) : F \in \mathcal{F}_0\},$$

where $\mathcal{F}_0 \in \mathcal{F}$ contains all multinomial distributions which put weights $p_1, \cdots, p_n$ entirely on $x_1, \cdots, x_n$ respectively.

The (profile) empirical likelihood ratio function for μ is defined as

$$\begin{aligned}\mathcal{R}(\mu) &= \frac{\sup\{L(F) : F \in \mathcal{F}_0, \sum_{i=1}^{n} p_i x_i = \mu\}}{\sup\{L(F) : F \in \mathcal{F}_0\}} \\ &= \sup\{\prod_{i=1}^{n} np_i : \sum_{i=1}^{n} p_i = 1, p_i \geq 0, \sum_{i=1}^{n} p_i x_i = \mu\}.\end{aligned}$$

The above second equality holds because $F(x) = \sum_{i=1}^{n} p_i I\{X_i \leq x\}$ when $F \in \mathcal{F}_0$ and $\sup\{L(F) : F \in \mathcal{F}_0\} = L(F_n) = (\frac{1}{n})^n$. In the following, we state the important and fundamental result in the empirical likelihood theory, which was proved by Owen (1988).

Theorem 2.1 (Owen (1988)) Let $X_1, \cdots, X_n$ be an *i.i.d.* random sample from $F_0 \in \mathcal{F}$. Let $\mu_0 = \int x dF_0(x)$ and assume that $\sigma^2 = Var(X_i) < \infty$. Then

$$-2 \log \mathcal{R}(\mu_0) \rightarrow \chi_1^2.$$

This theorem is similar to Wilks' theorem in parametric settings, which can be used to test statistical hypothesis and construct confidence intervals on μ. Since Owen's pioneer work on empirical likelihood, the EL method has been extended and applied to different statistical problems. One of the applications is the two-sample problem described in the next section.

2.2 Naive two-sample empirical likelihood

In this section, we briefly introduce how to use Owen (1988)'s method to establish the naive two-sample empirical likelihood method. Owen (2001) provides a multi-sample empirical likelihood theorem, which includes the two-sample problem as a special case. This theorem is stated in the two-sample setting as follows:

Theorem 2.1 (Owen (2001)) Let $y_{11}, \cdots, y_{1n_1} \in \mathbb{R}^t \sim F_0$ and $y_{21}, \cdots, y_{2n_2} \in \mathbb{R}^t \sim G_0$, with all observations independent. Let $\theta \in \mathbb{R}^t$ be defined by $E[h(y_1, y_2, \theta)] = 0$, where $y_1 \sim F_0$, $y_2 \sim G_0$ and $h(y_1, y_2, \theta) \in R^t$. Define the (profile) empirical likelihood ratio function as

$$R(\theta) = \sup\{\prod_{i=1}^{n_1}(n_1 p_i)\prod_{j=1}^{n_2}(n_2 q_j) : \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} p_i q_j h(y_{1i}, y_{2j}, \theta) = 0\}.$$

Then $-2\log R(\theta_0)$ converges in distribution to a χ^2_t distribution random variables.

Proof: the proof can be found in Owen (2001).

Now we define $h(y_1, y_2, d) = y_1 - y_2 - d$ where $d = \mu_1 - \mu_2$, where μ_1 and μ_2 are two population means respectively. Then this theorem provides a naive two-sample empirical likelihood method for general two-sample problems:

Corollary 2.3 Define the naive two-sample empirical likelihood ratio function

$$R(d) = \sup\{\prod_{i=1}^{n_1}(n_1 p_i)\prod_{j=1}^{n_2}(n_2 q_j) : \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} p_i q_j (y_{1i} - y_{2j} - d) = 0\}.$$

Then when d is the true difference of two population means, $-2\log R(d)$ converges in distribution to a χ^2_t distribution random variables.

The above result can be derived under a different formulation, as outlined in Wu (2009). We consider $t = 1$. We maximize the empirical likelihood function

$$L(\mathbf{p}_1, \mathbf{p}_2) = \prod_{j=1}^{n_1}(n_1 p_{1j})\prod_{j=1}^{n_2}(n_2 p_{2j})$$

subject to constraints

$$\sum_{j=1}^{n_1} p_{1j} = 1, \sum_{j=1}^{n_1} p_{1j}(y_{1j} - \mu_0 - d) = 0,$$
$$\sum_{j=1}^{n_2} p_{2j} = 1, \sum_{j=1}^{n_2} p_{2j}(y_{2j} - \mu_0) = 0$$

for a fixed μ_0, where $\mathbf{p}_1 = (p_{11}, \cdots, p_{1n_1})'$ and $\mathbf{p}_2 = (p_{21}, \cdots, p_{2n_2})'$ are the discrete probability measures over these two samples respectively, and $\mu_0 = \mu_2 + O(n_2^{-1/2})$. This leads to a log empirical likelihood ratio function $r(\mu_0, d)$ with two parameters d and μ_0, where

$$r(\mu_0, d) = -\sum_{j=1}^{n_1} \log\{1 + \lambda_1(y_{1j} - \mu_0 - d)\} - \sum_{j=1}^{n_2} \log\{1 + \lambda_2(y_{2j} - \mu_0)\},$$

and λ_1 and λ_2 are the solutions to

$$\sum_{j=1}^{n_1} \frac{y_{1j} - \mu_0 - d}{1 + \lambda_1(y_{1j} - \mu_0 - d)} = 0$$

and

$$\sum_{j=1}^{n_2} \frac{y_{2j} - \mu_0}{1 + \lambda_2(y_{2j} - \mu_0)} = 0$$

respectively.

To derive the empirical likelihood ratio function for d, we find $\hat{\mu}_0$, which maximizes $r(\mu_0, d)$ for a fixed d. After plugging $\hat{\mu}_0$ into the EL ratio function, we get

$$-2r(d) = -2r(\hat{\mu}_0, d) = (\bar{y}_{1.} - \bar{y}_{2.} - d)^2 / (\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) + o_p(1).$$

Hence we have the following theorem:

Theorem 2.4 Suppose that the log empirical likelihood ratio function $r(d)$ is defined as above, then when d is the true difference of the two population means, $-2r(d)$ converges in distribution to a χ_1^2 random variable.

2.3 Weighted two-sample empirical likelihood

As shown by Owen (2001), EL can be applied to solve multi-sample problems, including two-sample problems. One can use the EL methods to construct confidence intervals without assuming the distributions of the samples. However, the EL method may encounter some problems when applied to multiple-sample problems. The major one is that, EL may not perform well when the distributions are quite skewed or sample sizes are not large or sample sizes from each population are quite different. Fu, Wang and Wu (2009) developed a weighted empirical likelihood

(WEL) method, and tried to alleviate the difficulties that EL may encounter. They showed through simulation studies that WEL performs better than the naive EL and some other traditional methods, when population distributions are skewed and sample sizes are small or moderate.

In this section, enlightened by Fu, Wang and Wu (2009)'s idea, we construct the weighted empirical likelihood methods for two sample problems.

We use the same setting as before. We suppose $\{y_{11}, \cdots, y_{1n_1}\}$ is a random sample from a population following the distribution F_0 and $\{y_{21}, \cdots, y_{2n_2}\}$ is a random sample from a population following the distribution G_0 and these two samples are independent. Their population means, μ_1 and μ_2, differ by d. We are interested in testing the hypothesis:

$$H_0 : d = 0,$$

or equivalently,

$$H_0 : \mu_1 = \mu_2.$$

Some inspiring techniques for dealing with multi-sample problems are proposed by Fu, Wang and Wu (2009). We use similar method to establish a Wilk's type theorem for two-sample problems under a weighted EL formulation. First, we define the weighted log empirical likelihood function as

$$l_w(\mathbf{p}_1, \mathbf{p}_2) = \frac{w_1}{n_1} \sum_{j=1}^{n_1} \log(p_{1j}) + \frac{w_2}{n_2} \sum_{j=1}^{n_2} \log(p_{2j}), \tag{2.1}$$

subject to $\sum_{j=1}^{n_i} p_{ij} = 1, i = 1, 2$, and $\sum_{j=1}^{n_1} p_{1j} y_{1j} - \sum_{j=1}^{n_2} p_{2j} y_{2j} = d$, where $w_1 = w_2 = \frac{1}{2}$, $p_{ij} = Pr(Y_{ij} = y_{ij})$, and $\mathbf{p}_1 = (p_{11}, \cdots, p_{1n_1})'$ and $\mathbf{p}_2 = (p_{21}, \cdots, p_{2n_2})'$. Note that σ_1^2 and σ_2^2 denote respectively the two population variances.

Theorem 2.5 Assume that both σ_1^2 and σ_2^2 are finite and that $n_1/n_2 \to c_0 \neq 0$ as $n = n_1 + n_2 \to \infty$. When d is the true difference of two population means, we have

$$-2r_w(d)/c_1 \to \chi_1^2, \tag{2.2}$$

where the weighted log empirical likelihood ratio $-2r_w(d)$ is defined in (2.3) and c_1 is an adjusting constant and is specified in (2.4).

Proof: we rewrite our constraints as

$$\sum_{i=1}^{2} w_i \sum_{j=1}^{n_i} p_{ij} = 1$$

and

$$\sum_{i=1}^{2} w_i \sum_{j=1}^{n_i} p_{ij} Z_{ij} = \eta,$$

or rewrite the above equation as

$$\sum_{i=1}^{2} w_i \sum_{j=1}^{n_i} p_{ij} u_{ij} = 0$$

where $Z_{1j} = (1, y_{1j}/w_1)'$, $Z_{2j} = (0, -y_{2j}/w_2)'$, $\eta = (w_1, d)'$, $u_{ij} = Z_{ij} - \eta$. Let

$$G = l_w - \lambda'(\sum_{i=1}^{2} w_i \sum_{j=1}^{n_i} p_{ij} u_{ij}) - \tau(1 - \sum_{i=1}^{2} w_i \sum_{j=1}^{n_i} p_{ij}).$$

Use the standard Lagrange multiplier method, we get

$$p_{ij} = \frac{1}{n_i(1 + \lambda' u_{ij})},$$

and λ is the solution to

$$g(\lambda) = \sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{u_{ij}}{1 + \lambda' u_{ij}} = 0.$$

Substituting $1/(1 + \lambda' u_{ij}) = 1 - \lambda' u_{ij}/(1 + \lambda' u_{ij})$ into the above formula, we have

$$[\sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{u_{ij} u_{ij}^{'}}{1 + \lambda' u_{ij}}]\lambda = \sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} u_{ij}.$$

Noting that

$$U = \sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} u_{ij} = (0, y_{1.} - y_{2.} - d)',$$

and using similar argument in Owen (2001), we get

$$\lambda = D^{-1} U + o_p(n^{-1/2}),$$

where $D = \sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} u_{ij} u'_{ij}$.

The two-sample weighted log empirical likelihood ratio can be defined as

$$r_w(d) = -\sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \log(1 + \lambda' u_{ij}), \tag{2.3}$$

then using the Taylor series expansion, we get

$$
\begin{aligned}
-2r_w(d_0) &= 2\sum_{i=1}^{2}\frac{w_i}{n_i}\sum_{j=1}^{n_i}\log(1+\lambda' u_{ij}) \\
&= 2\sum_{i=1}^{2}\frac{w_i}{n_i}\sum_{j=1}^{n_i}(\lambda' u_{ij}-\frac{1}{2}\lambda' u_{ij}u_{ij}^{'}\lambda)+o_p(n^{-1}) \\
&= U'D^{-1}U+o_p(n^{-1}) \\
&= d^{(22)}(y_{1.}-y_{2.}-d_0)^2+o_p(n^{-1}),
\end{aligned}
$$

where $d^{(22)}$ is the second diagonal element of D^{-1}. Let

$$c_1 = d^{(22)}\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right). \tag{2.4}$$

It follows that when d is the true difference of two means, $-2_w r(d)/c_1$ converges in distribution to a χ_1^2 random variable with one degree of freedom.

2.4 Simulation studies

In this section, we assess the power of the tests using the two-sample naive EL method and the two-sample WEL method, and compare them to the student's t-test using pooled sample variance, under small or moderate sample-size settings. We generate our data from four different types of distributions: normal, log-normal, exponential and uniform. Different variance structures and different sample sizes are also taken into account. The null hypothesis is set as

$$H_0 : d = \mu_1 - \mu_2 = 0,$$

and the alternative hypothesis is

$$H_1 : d = 1/\sqrt{m},$$

when data are generated by normal, exponential and uniform distributions, where m is chosen for different values. For lognormal distributions, the alternative hypothesis is

$$H_1 : d = e^{(\mu_1+\sigma_1^2/2)} - e^{(\mu_2+\sigma_2^2/2)}.$$

We pre-specify the nominal level of the test at $\alpha = 0.05$. The total number of simulations is $B = 2000$.

Under each simulation run, the data were generated under different distributions, which have different means, variances and sample sizes. The true parameters of μ_1 and μ_2, and hence d, were set under the alternative hypothesis. The proportion of times of rejecting the null hypothesis is the so-call empirical power of a test. Specifically, let $(\hat{L}_b, \hat{U}_b)$ be the confidence interval for d from the bth simulation. Then the empirical power is defined as

$$P = 1 - B^{-1} \sum_{b=1}^{B} I(\hat{L}_b < 0 < \hat{U}_b) \times 100,$$

where $\hat{L}_b$ and $\hat{U}_b$ were computed using bi-section methods which is described in the next section, and I is the indicator function.

Tables 2.1, 2.2, 2.3 and 2.4 report the comparison of the empirical powers of t-test, naive empirical likelihood (EL) and weighted empirical likelihood (WEL). In Table 2.1, sample 1 and sample 2 were both generated from normal distribution with different means, standard deviations and sample sizes, indicated by (μ_1, μ_2), (σ_1, σ_2) and (n_1, n_2) respectively.

From Table 2.1, we see that three tests have similar performance in terms of power in most settings. However, when $(\sigma_1, \sigma_2) = (1, 1.5)$ and $(n_1, n_2) = (15, 30)$ or $(30, 15)$, the empirical powers of EL an WEL are much higher than the t-test. For example, when $(\mu_1, \mu_2) = (1 + 1/\sqrt{5}, 1)$, $(\sigma_1, \sigma_2) = (1, 1.5)$ and $(n_1, n_2) = (15, 30)$, the powers of EL and WEL methods are 0.2440 and 0.2300, respectively, while the power of the t-test method is 0.1610. In this case the powers of EL and WEL are approximately 50% higher than that of t-test. When $(\mu_1, \mu_2) = (1 + 1/\sqrt{30}, 1)$, $(\sigma_1, \sigma_2) = (1, 1.5)$ and $(n_1, n_2) = (15, 30)$, the powers of EL and WEL are approximately 100% higher than that of t-test.

Table 2.1: Comparisons of powers of t-test, EL and WEL when $Y \sim N(\mu, \sigma^2)$

(σ_1, σ_2)	(n_1, n_2)	(μ_1, μ_2)	t-test	EL	WEL
(1,1)	(30,30)	$(1+1/\sqrt{5},1)$	0.3855	0.4055	0.3945
		$(1+1/\sqrt{10},1)$	0.2165	0.2325	0.2205
		$(1+1/\sqrt{30},1)$	0.1075	0.1150	0.1105
	(15,30)	$(1+1/\sqrt{5},1)$	0.2795	0.3095	0.2895
		$(1+1/\sqrt{10},1)$	0.1790	0.1995	0.1835
		$(1+1/\sqrt{30},1)$	0.1025	0.1205	0.1135
	(30,15)	$(1+1/\sqrt{5},1)$	0.2755	0.3090	0.2920
		$(1+1/\sqrt{10},1)$	0.1660	0.1950	0.1805
		$(1+1/\sqrt{30},1)$	0.0935	0.1140	0.1010
(1,1.5)	(30,30)	$(1+1/\sqrt{5},1)$	0.2585	0.2710	0.2640
		$(1+1/\sqrt{10},1)$	0.1555	0.1690	0.1595
		$(1+1/\sqrt{30},1)$	0.0805	0.0920	0.0860
	(15,30)	$(1+1/\sqrt{5},1)$	0.1610	0.2440	0.2300
		$(1+1/\sqrt{10},1)$	0.0945	0.1635	0.1520
		$(1+1/\sqrt{30},1)$	0.0510	0.1045	0.0935
	(30,15)	$(1+1/\sqrt{5},1)$	0.2485	0.2125	0.1975
		$(1+1/\sqrt{10},1)$	0.1745	0.1435	0.1320
		$(1+1/\sqrt{30},1)$	0.1195	0.1025	0.0905
(1.5,1)	(30,30)	$(1+1/\sqrt{5},1)$	0.2625	0.2810	0.2685
		$(1+1/\sqrt{10},1)$	0.1515	0.1600	0.1535
		$(1+1/\sqrt{30},1)$	0.0830	0.0925	0.0845
	(15,30)	$(1+1/\sqrt{5},1)$	0.2465	0.2110	0.1970
		$(1+1/\sqrt{10},1)$	0.1820	0.1475	0.1365
		$(1+1/\sqrt{30},1)$	0.1250	0.1025	0.0905
	(30,15)	$(1+1/\sqrt{5},1)$	0.1525	0.2315	0.2180
		$(1+1/\sqrt{10},1)$	0.0960	0.1545	0.1415
		$(1+1/\sqrt{30},1)$	0.0490	0.0945	0.0850

In Tables 2.2, 2.3 and 2.4, data were generated from skewed distributions: the lognormal distribution, the exponential distribution, and the uniform distribution. From these tables, we see that EL and WEL perform better or much better than t-test in most cases.

For instance, when both samples are generated from lognormal distributions with parameters $(\mu_1, \mu_2) = (1 + 1/\sqrt{5}, 1)$, $(\sigma_1, \sigma_2) = (1.5, 1)$ and $(n_1, n_2) = (30, 15)$, the powers of EL and WEL methods are 0.6815 and 0.7185, respectively, while the power of the t-test is 0.0930, which is unacceptably low.

Table 2.2: Comparisons of powers of t-test, EL and WEL when $Y \sim \text{lognorm}(\mu, \sigma^2)$

(σ_1, σ_2)	(n_1, n_2)	(μ_1, μ_2)	t-test	EL	WEL
(1,1)	(30,30)	$(1+1/\sqrt{5},1)$	0.2700	0.3860	0.3755
		$(1+1/\sqrt{10},1)$	0.1515	0.2465	0.2360
		$(1+1/\sqrt{30},1)$	0.0690	0.1355	0.1290
	(15,30)	$(1+1/\sqrt{5},1)$	0.2700	0.3055	0.2235
		$(1+1/\sqrt{10},1)$	0.1630	0.2020	0.1475
		$(1+1/\sqrt{30},1)$	0.0860	0.1365	0.1025
	(30,15)	$(1+1/\sqrt{5},1)$	0.1410	0.3805	0.4160
		$(1+1/\sqrt{10},1)$	0.0805	0.2775	0.3025
		$(1+1/\sqrt{30},1)$	0.0460	0.1775	0.2010
(1,1.5)	(30,30)	$(1+1/\sqrt{5},1)$	0.0400	0.1260	0.1210
		$(1+1/\sqrt{10},1)$	0.0350	0.1525	0.1480
		$(1+1/\sqrt{30},1)$	0.0560	0.2270	0.2170
	(15,30)	$(1+1/\sqrt{5},1)$	0.0390	0.1435	0.1335
		$(1+1/\sqrt{10},1)$	0.0185	0.1665	0.1790
		$(1+1/\sqrt{30},1)$	0.0125	0.2320	0.2550
	(30,15)	$(1+1/\sqrt{5},1)$	0.0730	0.1685	0.1630
		$(1+1/\sqrt{10},1)$	0.0745	0.1665	0.1535
		$(1+1/\sqrt{30},1)$	0.0945	0.1800	0.1540
(1.5,1)	(30,30)	$(1+1/\sqrt{5},1)$	0.3825	0.7540	0.7465
		$(1+1/\sqrt{10},1)$	0.2785	0.6355	0.6205
		$(1+1/\sqrt{30},1)$	0.1880	0.5075	0.4970
	(15,30)	$(1+1/\sqrt{5},1)$	0.4470	0.5740	0.4935
		$(1+1/\sqrt{10},1)$	0.3450	0.4800	0.3925
		$(1+1/\sqrt{30},1)$	0.2655	0.3770	0.3055
	(30,15)	$(1+1/\sqrt{5},1)$	0.0930	0.6815	0.7185
		$(1+1/\sqrt{10},1)$	0.0630	0.5920	0.6270
		$(1+1/\sqrt{30},1)$	0.0370	0.4940	0.5300

Table 2.3: Comparisons of powers of t-test, EL and WEL when $Y \sim \exp(1/\mu)$

(n_1,n_2)	$(\mu_1,\ \mu_2)$	t-test	EL	WEL
(30,30)	$(1+1/\sqrt{5},1)$	0.2575	0.3290	0.3165
	$(1+1/\sqrt{10},1)$	0.1550	0.2000	0.1925
	$(1+1/\sqrt{30},1)$	0.0820	0.1095	0.1050
(15,30)	$(1+1/\sqrt{5},1)$	0.2480	0.2550	0.1940
	$(1+1/\sqrt{10},1)$	0.1620	0.1735	0.1340
	$(1+1/\sqrt{30},1)$	0.1035	0.1195	0.0810
(30,15)	$(1+1/\sqrt{5},1)$	0.1470	0.3135	0.3420
	$(1+1/\sqrt{10},1)$	0.0920	0.2180	0.2395
	$(1+1/\sqrt{30},1)$	0.0530	0.1480	0.1585

Table 2.4: Comparisons of powers of t-test, EL and WEL when $Y \sim \text{unif}(0, 2\mu)$

(n_1,n_2)	$(\mu_1,\ \mu_2)$	t-test	EL	WEL
(30,30)	$(1+1/\sqrt{5},1)$	0.6360	0.6700	0.6560
	$(1+1/\sqrt{10},1)$	0.4080	0.4400	0.4245
	$(1+1/\sqrt{30},1)$	0.1795	0.1925	0.1860
(15,30)	$(1+1/\sqrt{5},1)$	0.5155	0.4765	0.4485
	$(1+1/\sqrt{10},1)$	0.3285	0.2945	0.2775
	$(1+1/\sqrt{30},1)$	0.1520	0.1495	0.1390
(30,15)	$(1+1/\sqrt{5},1)$	0.4320	0.5755	0.5550
	$(1+1/\sqrt{10},1)$	0.2535	0.3585	0.3450
	$(1+1/\sqrt{30},1)$	0.1140	0.1660	0.1560

2.5 Computational algorithms

For the naive two-sample empirical likelihood method, the major computational problem is to maximize the joint empirical log-likelihood

$$l(p_1, p_2) = \sum_{j=1}^{n_1} \log(p_{1j}) + \sum_{j=1}^{n_2} \log(p_{2j})$$

subject to $\sum_{j=1}^{n_1} p_{1j} = 1$, $\sum_{j=1}^{n_2} p_{2j} = 1$ and $\sum_{j=1}^{n_1} p_{1j}y_{1j} - \sum_{j=1}^{n_2} p_{2j}y_{2j} = d$. We adopt the technique used in the section on WEL to reformulate the constraints as

$$\frac{1}{2}\sum_{j=1}^{n_1} p_{1j} + \frac{1}{2}\sum_{j=1}^{n_2} p_{2j} = 1,$$
$$\frac{1}{2}\sum_{j=1}^{n_1} p_{1j}z_{1j} + \frac{1}{2}\sum_{j=1}^{n_2} p_{2j}z_{2j} = \eta,$$

where $z_{1j} = (1, y_{1j})'$, $z_{2j} = (0, -y_{2j})'$ and $\eta = (1/2, d/2)'$. We define $u_{1j} = z_{1j} - \eta$ and $u_{2j} = z_{2j} - \eta$ and $q_{ij} = 1/2p_{ij}$ and the computational task is equivalent to maximizing

$$l_1(q_1, q_2) = \sum_{i=1}^{2}\sum_{j=1}^{n_i} \log(q_{ij})$$

with the re-formulated constraints

$$\sum_{i=1}^{2}\sum_{j=1}^{n_i} q_{ij} = 1,$$
$$\sum_{i=1}^{2}\sum_{j=1}^{n_i} q_{ij}u_{ij} = 0.$$

The maximization problem thus is changed to a standard formulation. Hence we can use the standard computational procedure developed by Wu (2004) to deal with it. To be specific, we define the Lagrangian

$$G = l_1(q_1, q_2) - \lambda_1(\sum_{i=1}^{2}\sum_{j=1}^{n_i} q_{ij} - 1) - (n_1 + n_2)\lambda_2'(\sum_{i=1}^{2}\sum_{j=1}^{n_i} q_{ij}u_{ij}).$$

Differentiating G with respect to q_{ij} and set $\frac{\partial G}{\partial q_{ij}} = 0$, we get

$$\frac{1}{q_{ij}} - \lambda_1 - (n_1 + n_2)\lambda_2' u_{ij} = 0,$$

where $j = 1, \cdots, n_i$, $i = 1, 2$. This gives

$$1 - \lambda_1 q_{ij} - (n_1 + n_2)\lambda_2' q_{ij}u_{ij} = 0. \tag{2.5}$$

We sum the above equation over i and j and noting that $\sum_{i=1}^{2}\sum_{j=1}^{n_i} q_{ij}u_{ij} = 0$, we get

$$\lambda_1 = n_1 + n_2.$$

From (2.5), we have

$$q_{ij} = \frac{1}{n_1 + n_2}\frac{1}{1 + \lambda_2' u_{ij}}.$$

Substituting this into $\sum_{i=1}^{2}\sum_{j=1}^{n_i} q_{ij}u_{ij} = 0$, we obtain

$$g(\lambda_2) = \frac{1}{n_1 + n_2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\frac{u_{ij}}{1 + \lambda_2' u_{ij}} = 0. \tag{2.6}$$

Similar to the arguments used by Owen (2001), (2.6) has a unique solution asymptotically with probability one. However, the basic version of Newton's method is not reliable to solve this equation. A modified version of Newton's method developed by Wu (2004) can be used to solve (2.6). Let $\lambda_2 = \lambda_2(d)$ be the solution of (2.6) with a fixed d.

Since $l(p_1, p_2) = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\log(p_{ij})$, subject to $\sum_{j=1}^{n_i} p_{ij} = 1, i = 1, 2$ attains its maximum when $p_{ij} = \frac{1}{n_i}$. That is, $l = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\log(\frac{1}{n_i})$.

Besides,

$$l_1(d) = \sup\{l_1(q_1, q_2)|\sum_{i=1}^{2}\sum_{j=1}^{n_i} q_{ij} = 1, \sum_{i=1}^{2}\sum_{j=1}^{n_i} q_{ij}u_{ij} = 0\}$$

attains its maximum when $d = \bar{y}_{1.} - \bar{y}_{2.}$, so that $\hat{d}_{el} = \bar{y}_{1.} - \bar{y}_{2.}$.

Notice that

$$l_1(d) = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\log(\frac{1}{n_1 + n_2}\frac{1}{1 + \lambda_2' u_{ij}}),$$

so

$$l(d) = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\log(\frac{2}{n_1 + n_2}\frac{1}{1 + \lambda_2' u_{ij}}),$$

then we know the two-sample log-likelihood ratio is

$$-2r(d) = 2l - 2l(d) = -2\sum_{i=1}^{2}\sum_{j=1}^{n_i}\log(\frac{2n_i}{n_1 + n_2}\frac{1}{1 + \lambda_2' u_{ij}}).$$

Next our task is to find the confidence interval $\{d| - 2r(d) < \chi_1^2(\alpha)\}$. First we note that

$$y_{1(1)} - y_{2(n_2)} \le d \le y_{1(n_1)} - y_{2(1)},$$

where $y_{1(1)} = \min\{y_{11}, \cdots, y_{1n_1}\}$, $y_{2(n_2)} = \max\{y_{21}, \cdots, y_{2n_2}\}$, $y_{1(n_1)} = \max\{ y_{11}, \cdots, y_{1n_1}\}$ and $y_{2(1)} = \min\{ y_{21}, \cdots, y_{2n_2}\}$. Then according to Wu (2004) and Tsao and Wu (2006), $-2r(d)$ is convex for $d \in (y_{1(1)} - y_{2(n_2)}, y_{1(n_1)} - y_{2(1)})$, hence we get the monotone property of $-2r(d)$: it is monotone decreasing in $(y_{1(1)} - y_{2(n_2)}, \bar{y}_{1.} - \bar{y}_{2.})$ and monotone increasing in $(\bar{y}_{1.} - \bar{y}_{2.}, y_{1(n_1)} - y_{2(1)})$. As a result, we can use the bisection method to find the lower and upper bound of $\{d| - 2r(d) < \chi^2_1(\alpha)\}$ for any given α.

Next we briefly introduce the computational algorithms for weighted empirical likelihood.

The maximization procedure of WEL is similar to the naive EL, except that we need to deal with the adjusting factor in this section.

The major task is to maximize

$$l_w(p_1, p_2) = \frac{1}{2n_1}\sum_{j=1}^{n_1}\log(p_{1j}) + \frac{1}{2n_2}\sum_{j=1}^{n_2}\log(p_{2j})$$

subject to

$$\sum_{i=1}^{2}\frac{1}{2}\sum_{j=1}^{n_i} p_{ij} = 1,$$
$$\sum_{i=1}^{2}\frac{1}{2}\sum_{j=1}^{n_i} p_{ij}u_{ij} = 0.$$

Here $u_{1j} = (1/2, 2y_{1j})'$ and $u_{2j} = (-1/2, -2y_{2j} - d)'$. The adjusted weighted two-sample log-likelihood ratio is

$$-2r_w(d)/c_1 = 2\sum_{i=1}^{2}\frac{1}{2c_1 n_i}\sum_{j=1}^{n_i}\log(1 + \lambda' u_{ij})$$

where $c_1(d) = d^{(22)}\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)$ and λ is the solution of

$$g_w(\lambda) = \sum_{i=1}^{2}\frac{1}{2n_i}\sum_{j=1}^{n_i}\frac{u_{ij}}{1 + \lambda' u_{ij}} = 0.$$

Use similar arguments to Wu (2004), $-2r_w(d)/c_1$ still maintains similar monotone property. Hence, we can still use bi-section method to find the lower and upper bound of $\{d| - 2r_w(d)/c_1 < \chi^2_1(\alpha)\}$.

Chapter 3

Two-sample Empirical Likelihood in Linear Models with Missing Data

3.1 Missing data in two-sample problems

As shown by Owen (2001) and many other authors, the naive EL method and it's varieties have been successfully applied to many statistical areas. However, EL methods for multi-sample problems with missing data have not been developed.

For one-sample problems, Qin's University of Waterloo 1992 Ph.D. thesis and Qin (2000) studied a conditional parametric model, say, $f(y|x, \theta)$, while leaving the marginal distribution $G(x)$ of X totally unspecified. He assumed the auxiliary information is available and can be summarized through estimating equations. Suppose the observed values are (y_i, x_i), for $i = 1, \cdots, m$, and $x_{m+1}, \cdots, x_n$, and the last $n - m$ response data are missing. The major interest is the mean response. His key idea was to combine parametric and empirical likelihood to give an efficient point estimator and to construct confidence intervals. Under this setting, he showed that the point estimator of θ is more efficient than maximum likelihood estimator when the auxiliary information is available. Besides, Qin (2000) proved a Wilks' type theorem for the combined likelihood ratio statistic when making inference for the mean response. However, Qin (2000) simply omitted the missing data. That is, he dropped the observed x_i with the missing response from his analysis. This may cause a serious loss of efficiency.

Based on Qin (2000)'s work, Wang and Dai (2008) imputed the missing data in the same setting except that they assumed that the data are missing at random. Hence, they proposed a new point estimator of θ and demonstrated that it's more efficient than Qin (2000)'s estimator. Besides, they also proved that the imputation-based likelihood ratio statistic converges to a weighted sum of chi-squared variables.

However, Qin (2000) and Wang and Dai (2008) did not give point estimation of the mean response. Wang and Veraverbeke (2002) studied the same model by using a different imputation method and didn't combine empirical likelihood with parametric likelihood. Similar results were proved. Furthermore, they gave a point estimator of the mean response, while the efficiency was not studied and the point estimator of θ was not provided.

Wang and Rao (2001, 2002a, 2002b) and Wang, Linton and Hardle (2004)'s ideas are similar. They used several kinds of imputation, including linear regression imputation and kernel regression imputation, to handle the incomplete data, and then adjusted the empirical likelihood methods to make inference. The empirical likelihood method needs to be adjusted in these situations since the data become dependent after imputation. These adjustments are based on different model assumptions and can lead to efficient point estimators and Wilks' type theorems. Above all, the intuitive idea to handle missing data problems by using EL methods is originated from Qin's 1992 Ph.D. thesis.

Those papers we mentioned above were all dealing with single sample problems. In this section, we will extend their results to two-sample problems where our interest is the difference of two population means. Difficulties arise since the technique to deal with multi-sample problems is quite different from single-sample problems.

3.2 Main results

In this section, we consider the following situation. Suppose we obtain *i.i.d.* $(x_{1j}, y_{1j}, \delta_{1j})$, and *i.i.d.* $(x_{2j}, y_{2j}, \delta_{2j})$, where $j = 1, \cdots, n_i$, and $\delta_{ij} = 1$ if y_{ij} observed, and $\delta_{ij} = 0$ otherwise. Here all x_{ij} are t-dimensional vectors and y_{ij} are univariates. Our interest is still the difference of two population means. We assume that y_{ij}'s are missing at random (MAR) assumption. That is, $P(\delta_i = 1|y_i, x_i) = P(\delta_i = 1|x_i)$.

We further assume the following linear regression model:

$$y_{ij} = x_{ij}^{'}\beta_i + \epsilon_{ij},$$

where $j = 1, \cdots, n_i$, $i = 1, 2$, and β_i are regression parameters. Here ϵ_{ij}'s are i.i.d. random errors with mean 0 and variance σ_i^2. We should note that when responses are complete, then these covariates x_{ij} provides no additional information for the inference of the difference of two means. However, when the MAR assumption is made and there are missing responses, x_{ij} provides useful information on the missing responses.

First we estimate β_i by the least squares estimator:

$$\tilde{\beta}_i = (\sum_{j=1}^{n_i} \delta_{ij} x_{ij} x'_{ij})^{-1} \sum_{j=1}^{n_i} \delta_{ij} x_{ij} y_{ij},$$

where $i = 1, 2$. Then we define

$$\tilde{y}_{ij} = \delta_{ij} y_{ij} + (1 - \delta_{ij}) x'_{ij} \tilde{\beta}_i,$$

where $j = 1, \cdots, n_i$, $i = 1, 2$.

Next, the (profile) two-sample empirical likelihood ratio is defined as

$$r(d) = \sum_{j=1}^{n_1} \log(n_1 p_{1j}) + \sum_{j=1}^{n_2} \log(n_2 p_{2j}), \tag{3.1}$$

subject to $\sum_{j=1}^{n_i} p_{ij} = 1$, $i = 1, 2$, and $\sum_{j=1}^{n_1} p_{1j} \tilde{y}_{1j} - \sum_{j=1}^{n_2} p_{2j} \tilde{y}_{2j} = d$.

Now we establish the main result:

Theorem 3.1 Under the assumed linear regression model and suitable regularity conditions, we have

$$-2r(d_0)/c_2 \longrightarrow \chi_1^2, \tag{3.2}$$

where c_2 is an adjusted constant and is specified in (3.7).

Proof: let μ_{i0} be the true parameter and μ_i is in the neighborhood of μ_{i0} such that $\mu_i = \mu_{i0} + O(n_i^{-1/2})$. We define

$$r(\mu_2, d) = \sum_{j=1}^{n_1} \log(n_1 p_{1j}) + \sum_{j=1}^{n_2} \log(n_2 p_{2j})$$

where $\mathbf{p}_1 = (p_{11}, \cdots, p_{1n_1})'$ and $\mathbf{p}_2 = (p_{21}, \cdots, p_{2n_2})'$ maximize

$$l(\mathbf{p}_1, \mathbf{p}_2) = \sum_{j=1}^{n_1} \log(p_{1j}) + \sum_{j=1}^{n_2} \log(p_{2j}), \tag{3.3}$$

subject to

$$\sum_{j=1}^{n_i} p_{ij} = 1, \qquad i = 1, 2,$$
$$\sum_{j=1}^{n_1} p_{1j}(\tilde{y}_{1j} - \mu_2 - d) = 0,$$
$$\sum_{j=1}^{n_2} p_{2j}(\tilde{y}_{2j} - \mu_2) = 0.$$

By using the Lagrange multiplier method, we have

$$p_{1j} = \frac{1}{n_1[1+\tau_1^{'}(\tilde{y}_{1j}-\mu_2-d)]},$$
$$p_{2j} = \frac{1}{n_2[1+\tau_2^{'}(\tilde{y}_{2j}-\mu_2)]},$$

where τ_1, τ_2 are the solutions to

$$\frac{1}{n_1}\sum_{j=1}^{n_1}\frac{\tilde{y}_{1j}-\mu_2-d}{1+\tau_1^{'}(\tilde{y}_{1j}-\mu_2)} = 0$$

and

$$\frac{1}{n_2}\sum_{j=1}^{n_2}\frac{\tilde{y}_{2j}-\mu_2}{1+\tau_2^{'}(\tilde{y}_{2j}-\mu_2)} = 0.$$

Then,

$$r(\mu_2, d) = -\sum_{j=1}^{n_1}[1+\tau_1^{'}(\tilde{y}_{1j}-\mu_2-d)] - \sum_{j=1}^{n_2}[1+\tau_2^{'}(\tilde{y}_{2j}-\mu_2)].$$

Let

$$\frac{\partial r(\mu_2, d)}{\partial \mu_2} = 0,$$

then we get

$$n_1\tau_1 + n_2\tau_2 = 0. \tag{3.4}$$

Besides, from Wang and Rao (2002), we have

$$\frac{1}{\sqrt{n_i}}\sum_{j=1}^{n_i}(\tilde{y}_{ij}-\mu_{i0}) \to N(0, V_i),$$
$$\tau_i = O_p(n_i^{-1/2}),$$
$$\frac{1}{n_i}\sum_{j=1}^{n_i}(\tilde{y}_{ij}-\mu_{i0})^2 \to U_i,$$
$$\max_{1\le j\le n_i}|\tilde{y}_{ij}| = o_p(n_i^{1/2}).$$

Here

$$V_i = V_i(\mu_i) = S_{i1} + S_{i2}^{\prime}S_{i3}^{-1}S_{i2}\sigma_i^2 + \beta_i^{\prime}S_{i4}\beta_i - 2S_{i5}^{\prime}\beta_i\mu_i + \mu_i^2 + 2S_{i2}^{\prime}S_{i3}^{-1}S_{i6}\sigma_i^2,$$

and

$$U_i = U_i(\mu_i) = S_{i1} + \beta_i^{\prime}S_{i4}\beta_i - 2S_{i5}^{\prime}\beta_i\mu_i + \mu_i^2,$$

where $S_{i1} = E[\delta_i(Y_i - X_i'\beta_i)^2]$, $S_{i2} = E[(1-\delta_i)X_i]$, $S_{i3} = E[\delta_i X_i X_i']$, $S_{i4} = E[X_i X_i']$, $S_{i5} = E[X_i]$, $S_{i6} = E[\delta_i X_i]$. Hence we can get

$$\tau_1 = [\sum_{j=1}^{n_1} (\tilde{y}_{1j} - \mu_2 - d)^2]^{-1} \sum_{j=1}^{n_1} (\tilde{y}_{1j} - \mu_2 - d) + o_p(n_1^{-1/2}), \tag{3.5}$$

$$\tau_2 = [\sum_{j=1}^{n_2} (\tilde{y}_{2j} - \mu_2)^2]^{-1} \sum_{j=1}^{n_2} (\tilde{y}_{2j} - \mu_2) + o_p(n_2^{-1/2}), \tag{3.6}$$

then (3.5) and (3.6) lead to the solution to (3.4) as

$$\tilde{\mu_2} = \alpha(\tilde{y}_{1.} - d) + (1-\alpha)\tilde{y}_{2.} + o_p(n^{-1/2}),$$

where

$$\alpha = \left(\frac{n_1}{U_1}\right) / \left(\frac{n_1}{U_1} + \frac{n_2}{U_2}\right)$$

and $\tilde{y}_{i.} = n_i^{-1} \sum_{j=1}^{n_i} \tilde{y}_{ij}$. It follows that

$$\begin{aligned} \tilde{\mu_2} - \mu_{20} &= \alpha(\tilde{y}_{1.} - \mu_{20} - d_0) + (1-\alpha)(\tilde{y}_{2.} - \mu_{20}) + o_p(n^{-1/2}) \\ &= \alpha(\tilde{y}_{1.} - \mu_{10}) + (1-\alpha)(\tilde{y}_{2.} - \mu_{20}) + o_p(n^{-1/2}) \\ &= O_p(n^{-1/2}), \end{aligned}$$

where $n = n_1 + n_2$. As a result, $\tau_1(\tilde{y}_{1.} - \tilde{\mu}_{20} - d_0) = o_p(n^{-1/2})$ and $\tau_2(\tilde{y}_{2.} - \tilde{\mu_2}) = o_p(n^{-1/2})$.

Finally we get

$$\begin{aligned} -2r(d_0) &= -2r(\tilde{\mu_2}, d_0) \\ &= \frac{n_1}{U_1}(\tilde{y}_{1.} - \tilde{\mu_2} - d_0)^2 + \frac{n_2}{U_2}(\tilde{y}_{2.} - \tilde{\mu_2})^2 + o_p(1) \\ &= (\tilde{y}_{1.} - \tilde{y}_{2.} - d_0)^2 / \left(\frac{U_1}{n_1} + \frac{U_2}{n_2}\right) + o_p(1). \end{aligned}$$

If we define

$$c_2 = \left(\frac{V_1}{n_1} + \frac{V_2}{n_2}\right) / \left(\frac{U_1}{n_1} + \frac{U_2}{n_2}\right), \tag{3.7}$$

we can easily see that $-2r(d_0)/c_2$ converges in distribution to a χ_1^2 random variable.

3.3 Simulation studies

In this section, we report results from a simulation study in comparing our two-sample EL method with the student's t-test treating imputed data as the original data, under small or moderate sample-size settings.

We consider the simple linear regression model $Y_i = X_i\beta_i + \epsilon_i$, $i = 1, 2$. Here we set $\beta_1 = 1.5$ and $\beta_2 = 1$. The dimension of the X_i's is set to be 1. The X_i's and ϵ_i's are generated from different distributions.

Different variance structures and different sample sizes are also taken into account. We always set the null hypothesis as

$$H_0 : d = \mu_1 - \mu_2 = 0,$$

and the alternative hypothesis as

$$H_1 : d = 0.5.$$

We pre-specify the nominal level of the test as $\alpha = 0.05$. The total number of simulations is $B = 2000$. Tables 3.1, 3.2 and 3.3 present the empirical power of the tests under different settings.

Under each simulation run, the data were generated under different distributions, which have different means, variances and sample sizes, and parameters were set under the alternative hypothesis. The proportion of times of rejecting the null hypothesis is the so-call empirical power of a test. See Section 2.4 for detailed definitions.

Under the model $Y_i = X_i\beta_i + \epsilon_i$, $i = 1, 2$, the correlation between the responses Y_i and the covariates X_i

$$\rho_i = \frac{Cov(Y_i, X_i)}{\sqrt{Var(Y_i)Var(X_i)}} = \frac{1}{\sqrt{1 + \frac{\sigma^2_{\epsilon_i}}{\sigma^2_{X_i}}\frac{1}{\beta_i^2}}}.$$

Let

$$EP(X_i) = E[P(\delta_i = 1|X_i)],$$

which indicates the expectation of observed rate of the responses in sample i, where $i = 1, 2$.

In Table 3.1, the covariates X_1 and X_2 are both generated from the normal distribution with mean 1 and standard deviation 1.

We consider the following two missing cases under the MAR assumption in Table 3.1.

*Case*1:

$$P(\delta_i = 1|X_i = x_i) = \begin{cases} 0.2, & \text{if } X_i - 1 \le -1; \\ 0.84, & \text{if } -1 \le X_i - 1 \le 1; \\ 0.6, & \text{if } X_i - 1 \ge 1. \end{cases}$$

*Case*2:

$$P(\delta_i = 1|X_i = x_i) = \begin{cases} 0.2, & \text{if } X_i - 1 \le -1; \\ 0.638, & \text{if } -1 \le X_i - 1 \le 1; \\ 0.2, & \text{if } X_i - 1 \ge 1. \end{cases}$$

where Case 1 leads to $EP(X_i) \cong 0.7$ and Case 2 leads to $EP(X_i) \cong 0.5$.

The error terms ϵ_i are generated from normal distributions with mean 0 and different variances, that is, $\epsilon_i \sim N(0, \sigma_{\epsilon_i}^2)$. The choose of σ_{ϵ_i} is specified as follows:

In Table 3.1, we used $\sigma_{\epsilon_1} = 1.125$ and $\sigma_{\epsilon_2} = 0.75$. This leads to $\rho_2 = \rho_2 = 0.8$. Similarly we let $\sigma_{\epsilon_1} = 2.598$ and $\sigma_{\epsilon_2} = 1.732$ such that $\rho_1 = \rho_2 = 0.5$ and $\sigma_{\epsilon_1} = 4.770$ and $\sigma_{\epsilon_2} = 3.180$ such that $\rho_1 = \rho_2 = 0.3$.

In Table 3.1, t-test and EL have similar performance in terms of power. In some cases, EL performs a little better. When the correlation between the response and covariate decreases, the power of both tests tend to decrease too, which justifies that the information from covariates could help improve the power of the test. For example, when $(E[P(X_1)], E[P(X_2)]) = (0.7, 0.7)$, $(n_1, n_2) = (60, 60)$ and the correlations (ρ_1, ρ_2) decrease, the power of both tests also decreases while the power of the EL test is always slightly higher than the power of the t-test. When $(\rho_1, \rho_2) = (0.5, 0.5)$, $(E[P(X_1)], E[P(X_2)]) = (0.7, 0.7)$ and $(n_1, n_2) = (30, 60)$ then the power of the t-test is 0.143 while EL is 0.154.

When the sample sizes $(n_1, n_2) = (60, 30)$, the power of the t-test tends to be slightly higher than the power of the EL test. And in some other cases, t-test is a little better than the EL method. For example, when $(\rho_1, \rho_2) = (0.3, 0.3)$, $(E[P(X_1)], E[P(X_2)]) = (0.5, 0.5)$ and $(n_1, n_2) = (60, 60)$, the power of the t-test is 0.1515 while EL is 0.1380.

Table 3.1: Comparisons of powers when covariates $X_i \sim N(1,1)$ and $\epsilon_i \sim N(0,\sigma_i^2)$

$(\rho_1,\, \rho_2)$	$(EP(X_1),\, EP(X_2))$	(n_1,n_2)	t-test	EL
(0.8,0.8)	(0.7,0.7)	(60,60)	0.4160	0.4190
		(30,60)	0.2850	0.2880
		(60,30)	0.3655	0.3590
	(0.5,0.5)	(60,60)	0.4120	0.4080
		(30,60)	0.2880	0.2825
		(60,30)	0.3635	0.3590
	(0.5,0.7)	(60,60)	0.3950	0.3910
		(30,60)	0.2755	0.2770
		(60,30)	0.3510	0.3465
(0.5,0.5)	(0.7,0.7)	(60,60)	0.1895	0.1910
		(30,60)	0.1460	0.1570
		(60,30)	0.1970	0.1900
	(0.5,0.5)	(60,60)	0.2175	0.2135
		(30,60)	0.1795	0.1800
		(60,30)	0.2170	0.2135
	(0.5,0.7)	(60,60)	0.1940	0.1880
		(30,60)	0.1710	0.1705
		(60,30)	0.2010	0.1885
(0.3,0.3)	(0.7,0.7)	(60,60)	0.1060	0.1085
		(30,60)	0.0960	0.1030
		(60,30)	0.1035	0.1005
	(0.5,0.5)	(60,60)	0.1515	0.1380
		(30,60)	0.1430	0.1375
		(60,30)	0.1565	0.1455
	(0.5,0.7)	(60,60)	0.1265	0.1220
		(30,60)	0.1335	0.1345
		(60,30)	0.1370	0.1300

Tables 3.2 and 3.3 report results when data are generated from skewed distributions. In Table 3.2, the covariates X_1 and X_2 are generated from exponential distributions: $X_i \sim \exp(1)$, $i = 1, 2$, so that their means are both 1 and variances are both 1.

We consider the following two missing cases under the MAR assumption in Table 3.2.

*Case*1:

$$P(\delta_i = 1|X_i = x_i) = \begin{cases} 0.78, & \text{if } -1 \leq X_i - 1 \leq 1; \\ 0.2, & \text{if } X_i - 1 \geq 1. \end{cases}$$

*Case*2:

$$P(\delta_i = 1|X_i = x_i) = \begin{cases} 0.44, & \text{if } -1 \leq X_i - 1 \leq 1; \\ 0.9, & \text{if } X_i - 1 \geq 1. \end{cases}$$

where Case 1 leads to $EP(X_i) \cong 0.7$ and Case 2 leads to $EP(X_i) \cong 0.5$.

We set the error terms $\epsilon_i \sim \sigma_{\epsilon_i}(\exp(1)-1)$ so that they have standard deviations σ_i. As a result, $\sigma_{\epsilon_1} = 1.125$ and $\sigma_{\epsilon_2} = 0.75$ lead to $\rho_1 = \rho_2 = 0.8$, $\sigma_{\epsilon_1} = 2.598$ and $\sigma_{\epsilon_2} = 1.732$ lead to $\rho_1 = \rho_2 = 0.5$, $\sigma_{\epsilon_1} = 4.770$ and $\sigma_{\epsilon_2} = 3.180$ lead to $\rho_1 = \rho_2 = 0.3$.

Table 3.2: Comparisons of powers when covariates $X_i \sim \exp(1)$ and $\epsilon_i \sim \sigma_{\epsilon_i}(\exp(1) - 1)$

(ρ_1, ρ_2)	$(EP(X_1), EP(X_2))$	(n_1,n_2)	t-test	EL
(0.8,0.8)	(0.7,0.7)	(60,60)	0.3815	0.4050
		(30,60)	0.2140	0.2690
		(60,30)	0.3450	0.3370
	(0.5,0.5)	(60,60)	0.4615	0.4845
		(30,60)	0.2790	0.3585
		(60,30)	0.3965	0.3790
	(0.5,0.7)	(60,60)	0.4355	0.4545
		(30,60)	0.2840	0.3525
		(60,30)	0.3965	0.3875
(0.5,0.5)	(0.7,0.7)	(60,60)	0.1665	0.1970
		(30,60)	0.0980	0.1355
		(60,30)	0.1745	0.1780
	(0.5,0.5)	(60,60)	0.2285	0.2630
		(30,60)	0.1575	0.2160
		(60,30)	0.2400	0.2390
	(0.5,0.7)	(60,60)	0.2300	0.2625
		(30,60)	0.1545	0.2185
		(60,30)	0.2370	0.2370
(0.3,0.3)	(0.7,0.7)	(60,60)	0.0980	0.1125
		(30,60)	0.0765	0.1010
		(60,30)	0.1210	0.1275
	(0.5,0.5)	(60,60)	0.1310	0.1575
		(30,60)	0.1115	0.1560
		(60,30)	0.1565	0.1675
	(0.5,0.7)	(60,60)	0.1360	0.1570
		(30,60)	0.1080	0.1560
		(60,30)	0.1565	0.1605

In Table 3.3, the covariates X_1 and X_2 are still generated from exponential distributions: $X_i \sim \exp(1)$, $i = 1, 2$, so that their means are both 1 and variances are both 1.

We consider the same missing cases under the MAR assumption in Table 3.3 as the one in Table 3.2.

The error terms are generated from a re-scaled lognormal distribution:

$$\epsilon \sim \sigma_{\epsilon_i} \frac{(\mathrm{lognorm}(0,1) - \mathrm{e}^{1/2})}{\mathrm{e}^2 - \mathrm{e}}$$

so that they have standard deviations σ_i. Here definitions of σ_{ϵ_i} are the same as in Table 3.1 and 3.2.

From Table 3.3, we see that in 21 out of 27 cases, EL performs better or much better than the t-test. When sample sizes are $(30, 60)$, EL has powers approximately $30\% \sim 50\%$ higher than those by the t-test.

Table 3.3: Comparisons of powers when $X \sim \exp(1)$, $\epsilon \sim \sigma_{\epsilon_i}(\text{lognorm}(0,1) - \text{e}^{1/2})/(\text{e}^2 - \text{e})$

(ρ_1, ρ_2)	$(EP(X_1), EP(X_2))$	(n_1,n_2)	t-test	EL
(0.8,0.8)	(0.7,0.7)	(60,60)	0.5270	0.5480
		(30,60)	0.3195	0.3985
		(60,30)	0.4580	0.4355
	(0.5,0.5)	(60,60)	0.5450	0.5635
		(30,60)	0.3435	0.4220
		(60,30)	0.4695	0.4465
	(0.5,0.7)	(60,60)	0.5455	0.5620
		(30,60)	0.3370	0.4200
		(60,30)	0.4730	0.4480
(0.5,0.5)	(0.7,0.7)	(60,60)	0.3710	0.4050
		(30,60)	0.2100	0.2975
		(60,30)	0.3485	0.3450
	(0.5,0.5)	(60,60)	0.4620	0.4855
		(30,60)	0.2735	0.3515
		(60,30)	0.3945	0.3890
	(0.5,0.7)	(60,60)	0.4340	0.4530
		(30,60)	0.2755	0.3550
		(60,30)	0.3815	0.3795
(0.3,0.3)	(0.7,0.7)	(60,60)	0.2100	0.2495
		(30,60)	0.1250	0.1920
		(60,30)	0.2250	0.2440
	(0.5,0.5)	(60,60)	0.3025	0.3335
		(30,60)	0.1750	0.2575
		(60,30)	0.2880	0.3010
	(0.5,0.7)	(60,60)	0.2815	0.3100
		(30,60)	0.1835	0.2650
		(60,30)	0.2885	0.2910

Chapter 4

Bootstrap Procedures for the Two-Sample Empirical Likelihood Methods

4.1 Bootstrap procedure for the two-sample weighted empirical likelihood method

As we described in Chapter 2, we derived the so-called weighted two-sample empirical likelihood method for the complete data case, which also involves a scaling constant c_1 in Theorem 2.5. In order to improve the performance of weighted EL method when sample sizes are not large and to avoid calculating c_1, we now develop a bootstrap procedure to achieve that goal.

Suppose we have univariate data $y_{ij},\ i = 1, 2,\ j = 1, \cdots, n_i$, which are our original samples. Here i indicate the i-th sample. The sample sizes n_1 and n_2 are fixed. The two samples are independent.

The weighted EL function defined in Chapter 2 is given by (2.1):

$$l_w(\mathbf{p}_1, \mathbf{p}_2) = \frac{w_1}{n_1}\sum_{j=1}^{n_1}\log(p_{1j}) + \frac{w_2}{n_2}\sum_{j=1}^{n_2}\log(p_{2j}),$$

where $w_1 = w_2 = 1/2$ are the equal "weights" for these two samples, and $\mathbf{p}_1 = (p_{11}, \cdots, p_{1n_1})'$ and $\mathbf{p}_2 = (p_{21}, \cdots, p_{2n_2})'$ are the discrete probability measures over the two samples, respectively. Due to Kiefer and Wolfowitz (1956), it is easy to verify that $\hat{\mathbf{p}}_1 = (1/n_1, \cdots, 1/n_1)'$ and $\hat{\mathbf{p}}_2 = (1/n_2, \cdots, 1/n_2)'$ uniquely maximize $l_w(\mathbf{p}_1, \mathbf{p}_2)$ subject to $\sum_{j=1}^{n_1} p_{1j} = 1$ and $\sum_{j=1}^{n_2} p_{2j} = 1$. As a result, the maximum WEL estimator for the difference $d = \mu_1 - \mu_2$ is simply $\hat{d} = \bar{y}_{1.} - \bar{y}_{2.}$, where $\bar{y}_{1.} = n_1^{-1}\sum_{j=1}^{n_1} y_{1j}$ and $\bar{y}_{2.} = n_2^{-1}\sum_{j=1}^{n_2} y_{2j}$. Let $\tilde{\mathbf{p}}_1(d)$ and $\tilde{\mathbf{p}}_2(d)$ be the maximizer of $l_w(\mathbf{p}_1, \mathbf{p}_2)$ subject to

$$\sum_{j=1}^{n_1} p_{1j} = 1,$$
$$\sum_{j=1}^{n_2} p_{2j} = 1,$$
$$\sum_{j=1}^{n_1} p_{1j}y_{1j} - \sum_{j=1}^{n_2} p_{2j}y_{2j} = d$$

for a fixed d. Then the two-sample weighted empirical likelihood ratio function $r_w(d) = \{l_w(\tilde{\mathbf{p}}_1(d), \tilde{\mathbf{p}}_2(d)) - l_w(\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2)\}$, which has another expression as stated in (2.3). As we showed in Chapter 2, when d_0 is the true difference of two population means, $-2r_w(d_0)/c_1$ converges in distribution to a chi-squared distribution with one degree of freedom under some mild regularity conditions, where c_1 is defined in (2.4). Hence, the (1-α)-level WEL confidence interval on d_0 is constructed as

$$\mathcal{C}_1 = \{d \mid -2\ r_w(d)/c_1 < \chi_1^2(\alpha)\}$$

since $P(-2r_w(d_0) < \chi_1^2(\alpha)) \doteq 1 - \alpha$ in the asymptotical sense.

However, the two-sample WEL confidence interval has the low coverage probability problem, which can be seen from the simulation studies reported in the next section. Sometimes it performs worse than the usual two-sample t-test. In addition, c_1 has a quite complicated form and one may wish to avoid calculating it every time using the two-sample WEL. These problems motivate us to use bootstrap method, an effective calibration method, to improve the performance in terms of coverage probabilities. Our proposed two-sample bootstrap empirical likelihood method is as follows.

First select a bootstrap sample $\{y_{i1}^*, \cdots, y_{in_i}^*\}$ from the i-th original sample $\{y_{i1}, \cdots, y_{in_i}\}$, $i = 1, 2$ with replacement. Then the bootstrap analogy of $l_w(\mathbf{p}_1, \mathbf{p}_2)$ can be defined to be

$$l_w^*(\mathbf{p}_1^*, \mathbf{p}_2^*) = \frac{w_1}{n_1}\sum_{j=1}^{n_1} \log(p_{1j}^*) + \frac{w_2}{n_2}\sum_{j=1}^{n_2} \log(p_{2j}^*),$$

Maximizing $l_w^*(\mathbf{p}_1^*, \mathbf{p}_2^*)$ subject to

$$\sum_{j=1}^{n_1} p_{1j}^* = 1,$$
$$\sum_{j=1}^{n_2} p_{2j}^* = 1,$$
$$\sum_{j=1}^{n_1} p_{1j}^*y_{1j}^* - \sum_{j=1}^{n_2} p_{2j}^*y_{2j}^* = \hat{d}$$

gives us the bootstrap version of the WEL ratio function in (2.3), here $\hat{d}$ is the difference of two original sample means, that is $\hat{d} = \bar{y}_{1.} - \bar{y}_{2.}$.

$$r_w^*(\hat{d}) \;\; = \;\; -\sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \log(1 + \lambda^{*'} u_{ij}^*),$$

where $u_{1j}^* = (1 - w_1, y_{1j}^*/w_1 - \hat{d})'$ and $u_{2j}^* = (-w_1, -y_{2j}^*/w_2 - \hat{d})'$, and λ^* satisfies

$$\sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{u_{ij}^*}{1 + \lambda^{*'} u_{ij}^*} = 0.$$

Next we need to show this bootstrap version of WEL has the same asymptotic scaled chi-squared distribution as $r_w(d_0)$, that is, the asymptotic distribution of $r_w^*(\hat{d})$ approximates that of $r_w(d_0)$. Denote b_α as the upper α quantile of the asymptotic distribution of $r_w(d_0)$. That is $P(r_w(d_0) < b_\alpha) \doteq 1-\alpha$, where b_α can be determined by Theorem 2.5. So that the (1-α)-level approximate WEL confidence interval can be constructed as

$$\mathcal{C}_2 = \{d \mid r_w(d) < b_\alpha\}.$$

If this bootstrap version of WEL has the same asymptotic scaled chi-squared distribution as $r_w(d_0)$, for which we will give a proof in the following, then we can replace b_α by using the upper α quantile of the asymptotic distribution of $r_w^*(\hat{d})$, denoted as $\tilde{b}_\alpha$. All we left is to calculate $\tilde{b}_\alpha$, which generally is also not easy to figure out. To do this, we use the Monte Carlo Simulation to approximate b_α^2 by the upper 100α-th sample quantile b_α^* from a repeated sequence. Specifically, we repeat our procedure independently for many times, say $B = 1000$ times, to get a sequences $r_{w,1}^*(\hat{d}), \cdots, r_{w,B}^*(\hat{d})$ in the same way that we get $r_w^*(\hat{d})$. Then b_α^* is the 100α-th sample quantile of this sequences. Then the so-called bootstrap calibrated WEL confidence interval on d_0 can be constructed as

$$\mathcal{C}^* = \{d \mid r_w(d) < b_\alpha^*\}.$$

As we can see, calculation of c_1 is bypassed. We also expect this bootstrap calibrated WEL confidence interval $\mathcal{C}^*$ would have better performance than the WEL confidence interval $\mathcal{C}_1$ or equally $\mathcal{C}_2$. This is demonstrated by our simulation studies reported in the next section.

In order to prove that the confidence interval $\mathcal{C}^*$ has correct asymptotic coverage probability at the (1-α)- level, we only need to show that the bootstrap version of WEL ratio $r_w^*(\hat{d})$ has the same asymptotic scaled chi-squared distribution as $r_w(d_0)$. That is, the bootstrap procedure provides an approximation to the asymptotic distribution of $r_w(d_0)$:

Theorem 4.1 Assume σ_1^2, $\sigma_2^2 < \infty$ and that $n_1/n_2 \to c_0 \neq 0$ as $n = n_1 + n_2 \to \infty$. When d_0 is the true difference of two population means, then as $n \to \infty$

$$|P(r_w(d_0) < x) - P^*(r_w^*(\hat{d}) < x)| \to 0$$

with probability 1 for any $x \in R$ and P^* means the bootstrap probability. That is, $r_w(d_0)$ and $r_w^*(\hat{d})$ converge in distribution to the same scaled χ_1^2 random variable according to Theorem 2.5.

Proof: Since $\{y_{i1}^*, \cdots, y_{in_i}^*\}$ is selected from the i-th original sample $\{y_{i1}, \cdots, y_{in_i}\}$, $i = 1, 2$ with replacement, it follows that $y_{i1}^*, \cdots, y_{in_i}^*$ are $i.i.d.$ for $i = 1, 2$ and these two bootstrap samples are still independent and they are all uniformly distributed on the observations of the i-th sample. As a result, we have $E^*(y_{ij}^*) = \bar{y}_{i.}$ and $Var^*(y_{ij}^*) = n_i^{-1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2$. Here E^* and Var^* means the bootstrap expectation and bootstrap variance respectively.

Since $E^*(y_{ij}^{*2}) = \bar{y}_{i.}^2 + n_i^{-1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2$, we have $P(E^*(y_{ij}^{*2}) < \infty) = 1$ under the assumption σ_1^2, $\sigma_2^2 < \infty$. Then from Lemma 11.2 in Chapter 11 of Owen (2001), we know that $\max_{1 \le j \le n_1} |y_{ij}^*| = o_{p^*}(\sqrt{n})$ and hence $\max_{1 \le j \le n_1} |u_{ij}^*| = o_{p^*}(\sqrt{n})$. Here $|u_{ij}^*| = \sqrt{(u_{ij,1}^*)^2 + (u_{ij,2}^*)^2}$, where $u_{ij,1}^*$ and $u_{ij,2}^*$ are the first and second component of u_{ij}^*.

Noting that λ^* satisfies

$$\sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{u_{ij}^*}{1 + \lambda^{*\prime} u_{ij}^*} = 0,$$

we rewrite this equation to get

$$\begin{aligned} 0 &= \sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{u_{ij}^*}{1 + \lambda^{*\prime} u_{ij}^*} \\ &= \sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{u_{ij}^*[1 + \lambda^{*\prime} u_{ij}^* - \lambda^{*\prime} u_{ij}^*]}{1 + \lambda^{*\prime} u_{ij}^*} \\ &= \sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} u_{ij}^* - \lambda^* \sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{u_{ij}^{*\prime} u_{ij}^*}{1 + \lambda^{*\prime} u_{ij}^*}. \end{aligned}$$

After some straightforward algebra, we have

$$U^* = \sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} u_{ij}^* = (0, \bar{y}_{1.}^* - \bar{y}_{2.}^* - \hat{d})'.$$

and we obtain

$$|\lambda^*| \sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{u_{ij}^{*'} u_{ij}^*}{1 + |\lambda^*| \max_{1 \le j \le n_1} |u_{ij}^*|} \le |\bar{y}_{1.}^* - \bar{y}_{2.}^* - \hat{d}|.$$

Here $|\lambda^*| = \sqrt{(\lambda_1^*)^2 + (\lambda_2^*)^2}$ where λ_1^* and λ_2^* is the first and second component of λ^*. Since $\max_{1 \le j \le n_1} |u_{ij}^*| = o_{p^*}(\sqrt{n})$ and we know from the central limit theorem that $\bar{y}_{1.}^* - \bar{y}_{2.}^* - \hat{d} = O_{p*}(n^{-1/2})$ and the above inequality, we must have $\lambda^* = O_{p*}(n^{1/2})$. Another similar way of proving this important order property can be found in Chapter 11.2 of Owen (2001). Using similar argument as in Theorem 2.5, we get

$$\begin{aligned} -2r_w^*(\hat{d}) &= 2\sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \log(1 + \lambda^{*'} u_{ij}^*) \\ &= 2\sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} (\lambda^{*'} u_{ij}^* - \frac{1}{2} \lambda^{*'} u_{ij}^* u_{ij}^{*'} \lambda^*) + o_{p^*}(n^{-1}) \\ &= U^{*'} D^{*-1} U^* + o_{p^*}(n^{-1}) \\ &= d^{*(22)} (\bar{y}_{1.}^* - \bar{y}_{2.}^* - \hat{d})^2 + o_{p^*}(n^{-1}), \end{aligned}$$

where $d^{*(22)}$ is the second diagonal element of D^{*-1} and $D^* = \sum_{i=1}^{2} \frac{w_i}{n_i} \sum_{j=1}^{n_i} u_{ij}^*$ $u_{ij}^{*'}$. Given the original samples $\{y_{i1}, \cdots, y_{in_i}\}$, $i = 1, 2$, we have from the central limit theorem that $-2r_w^*(\hat{d})/c_1^*$ converges in distribution to a chi-squared random variable where $c_1^* = d^{*(22)}(\sigma_1^{*2}/n_1 + \sigma_2^{*2}/n_2)$.

Finally, $\sigma_i^{*2} = Var^*(y_{ij}^*) = n_i^{-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 = n_i^{-1}(n_i - 1) S_i^2$ since we select our bootstrap sample with replacement. We can replace $\sigma_1^{*2}/n_1 + \sigma_2^{*2}/n_2$ by $S_1^2/n_1 + S_2^2/n_2$ since they are asymptotically the same. Noting that $c_1 = d^{(22)}(S_1^2/n_1 + S_2^2/n_2)$, similar discussion leads to $d^{(*22)}/d^{(22)} \to 1$. Hence Theorem 4.1 is proved.

4.2 Simulation studies

In this section, we conducted simulation studies to evaluate the performance of the usual two-sample t-test, naive two-sample empirical likelihood method (EL), weighted two-sample empirical likelihood method (WEL), weighted two-sample empirical likelihood method with bootstrap calibration (BWEL), in terms of their coverage probabilities of confidence intervals. We generated data from different distributions, including normal distribution, lognormal distribution, exponential distribution and uniform distribution, similar to those considered in section 2.4. The total number of simulation runs was 2000. For each simulation run, $B = 1000$ bootstrap samples were selected by sampling with replacement.

Table 4.1 contains results for the 95% confidence intervals on the difference d_0 of two population means when both samples were generated from normal distributions. We assessed the performances of confidence intervals in terms of empirical coverage probabilities. The true means μ_{10} and μ_{20} are both set to be 1. Sample sizes (n_1, n_2) are chosen to be (30,30), (15,30) and (30,15). The standard deviations (σ_1, σ_2) are chosen is be (1,1) or (1.5,1). As we can see from Table 4.1, four methods gave similar results when standard deviations of two samples were set to be equal to 1 while the coverage probability by EL is a little lower than the nominal 95% when sample sizes are set to be different. WEL and BWEL showed improvement of EL in those cases. When standard deviations are set to be $(1.5, 1)$ and sample sizes are the same, the coverage probabilities of the usual two-sample t-interval performs better than EL and WEL, and similar to BWEL. However, When standard deviations are set to be $(1.5, 1)$ and sample sizes are (15,30) or (30,15), the coverage probabilities by the t-interval are quite far away from the nominal value, while EL, WEL and BWEL all perform much better in these cases. In all five cases we examined here, WEL showed some improvement of EL while BWEL showed slightly improvement of WEL. In total, BWEL performed the best and EL, WEL and BWEL generally would not be far away from the nominal level. The usual two-sample t-interval doesn't perform well in all cases. In 2 cases out of 5 cases, it performs badly.

For the results reported in Table 4.2, we generated our data from other distributions including lognormal distribution, exponential distribution and uniform distribution. In the exponential and uniform case, the true means μ_{10} and μ_{20} for sample 1 and 2 are both set to be 1, while in the lognormal case, they are both set to be $e^{1/2}$. We chose two different scenarios of sample sizes: (30,30) and (15,30). In all 6 cased we considered here, the coverage probabilities by the t-interval and BWEL perform better than EL, WEL. WEL still showed slight improvement of EL in terms of coverage probabilities. Both EL and WEL encountered low coverage probability problem in the following two cases: (i) data were generated from lognormal distributions, while sample sizes are either (30,30) or (15,30); (ii) data were generated from exponential distributions and sample sizes were (15,30). In both cases BWEL performs very well and don't have low coverage probability problem.

In summary, BWEL performs best among four methods considered, which shows that bootstrap calibration does help improve the performance of WEL. There are several cases where the coverage probabilities of the usual two-sample t-interval, EL and WEL are far away from the nominal 95% level. We also note that WEL performs a little better than EL generally in our simulation studies, which shows that the "weighted" approach does have some advantage.

Table 4.1: Comparisons of coverage probabilities of t-test, EL, WEL and BWEL when $Y \sim N(1, \sigma^2)$

(σ_1, σ_2)	(n_1,n_2)	t-test	EL	WEL	BWEL
(1,1)	(30,30)	0.9490	0.9445	0.9475	0.9485
	(15,30)	0.9465	0.9295	0.9360	0.9390
(1.5,1)	(30,30)	0.9505	0.9440	0.9490	0.9525
	(15,30)	0.9055	0.9280	0.9360	0.9405
	(30,15)	0.9780	0.9370	0.9445	0.9465

Table 4.2: Comparisons of coverage probabilities of t-test, EL, WEL and BWEL when $Y \sim$ lognorm$(1, 1)$, exp(1) and unif$(0, 2)$

Distribution	(n_1,n_2)	t-test	EL	WEL	BWEL
lognorm	(30,30)	0.9630	0.9110	0.9170	0.9505
	(15,30)	0.9655	0.8985	0.8970	0.9345
exp	(30,30)	0.9550	0.9360	0.9400	0.9505
	(15,30)	0.9500	0.9165	0.9185	0.9300
unif	(30,30)	0.9490	0.9450	0.9495	0.9520
	(15,30)	0.9395	0.9300	0.9345	0.9360

Chapter 5

Conclusion and Future Work

In this thesis, we proposed new proof for the naive empirical likelihood method for two-sample problems, developed a two-sample weighted empirical likelihood method for complete data case, an adjusted two-sample empirical likelihood method for linear regression models with missing data, and the two-sample weighted empirical likelihood method with bootstrap calibration for complete data case. Simulation results showed that our new WEL, adjusted EL and BWEL all have favorable properties in terms of coverage probabilities and/or power of tests.

However, there are a lot of future work we can do, especially in the missing data area. An interesting research topic is to develop bootstrap method for the two-sample empirical likelihood with imputation for missing responses. Also, we investigated a parametric conditional model on the response combined with the nonparametric marginal distribution of the auxiliary variables and attempted to develop a semiparametric empirical likelihood but encountered some technical problems.

Applications of the two-sample techniques we developed here to case-control studies will also be of great interest.

Bibliography

[1] Fu, Y. J., Wang, X. G. and Wu, C. B. (2008), Weighted empirical likelihood inference for multiple samples, *Journal of Statistical Planning and Inference*, 139, 1462-1473.

[2] Kiefer, J. and Wolfowitz, J. (1956), Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Annals of Mathematical Statistics*, 27, 887-906.

[3] Owen, A. (1988), Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75, 237-249.

[4] Owen, A. (2001), *Empirical likelihood*, Chapman and Hall/CRC.

[5] Qin, J. (1992), Empirical likelihood and semiparametric models, Ph.D. thesis, University of Waterloo.

[6] Qin, J. and Lawless, J. F. (1994), Empirical likelihood and general estimating equations, *The Annals of Statistics*, 22, 300-325.

[7] Qin, J. (2000), Combing parametric and empirical likelihood, *Biometrika*, 87, 484-490.

[8] Tsao, M. and Wu, C. B. (2006). Empirical likelihood inference for a common mean in the presence of heteroscedasticity, *The Canadian Journal of Statistics*, 34, 45-59.

[9] Wang, L. C., and Veraverbeke, N. (2002), Empirical likelihood in a semi-parametric model for missing response data, *Communications in Statistics - Theory and Methods*, 35, 625-639.

[10] Wang, Q. H., Linton, O. and Hardle, W. (2004), Semiparametric regression analysis with missing response at random, *Journal of the American Statistical Association*, 99, 334-345.

[11] Wang, Q. H. and Rao, J. N. K. (2001), Empirical likelihood for linear regression models under imputation for missing responses, *The Canadian Journal of Statistics*, 29, 597-608.

[12] Wang, Q. H. and Rao, J. N. K. (2002a), Empirical likelihood-based inference under imputation with missing response, *The Annals of Statistics*, 30, 896-924.

[13] Wang, Q. H. and Rao, J. N. K. (2002b), Empirical likelihood-based inference in linear models with missing data, *Scandinavian Journal of Statistics*, 29, 563-576.

[14] Welch, B. L. (1938), The significance of the difference between two means when the population variances are unequal, *Biometrika*, 29, 350-62.

[15] Wu, C. B. (2004), Some algorithmic aspects of the empirical likelihood method in survey sampling, *Statistica Sinica*, 14, 1057-1067.

[16] Wu, C. B. (2009), Empirical likelihood inference for two populations, Working Paper, University of Waterloo.

[17] Zhou, X. H., Gao, S. and Hui, S. L. (1997), Methods for comparing the means of two independent log-normal samples, *Biometrics*, 53, 1129-1135.