Envelope empirical likelihood ratio for the difference of two location parameters with constraints of symmetry

Kyoungmi Kim [1], Mai Zhou [2]

University of Kentucky

Abstract

Empirical likelihood ratio method is a new technique in nonparametric inference developed by A. Owen (1988, 2001). Sometimes empirical likelihood has difficulties to define itself. As such a case in point, we discuss the way to define a modified empirical likelihood for the location of symmetry using well-known points of symmetry as a side conditions. The side condition of symmetry is defined through a finite subset of the infinite set of constraints.

The modified empirical likelihood under symmetry studied in this paper is to construct a constrained parameter space Θ^+ of distributions imposing known symmetry as side information. We show that the usual asymptotic theory (Wilks theorem) still hold for the empirical likelihood ratio on the constrained parameter space and the asymptotic distribution of the empirical NPMLE of difference of two symmetric points is obtained.

AMS 1991 Subject Classification: Primary 62G10; secondary 62G05.
Key Words and Phrases: Two-sample Problem, Many Constraints of symmetry, Asymptotic Chi-square distribution.

0. Introduction

In parametric likelihood inference, we need a parametric family of distributions, but we might have no idea which parametric family to use. In this case, one often opting for one of the well-known parametric families for convenience. If the decision was correct, then this parametric likelihood method is powerful and effective. Otherwise, misspecification can cause estimates to be inefficient and the corresponding confidence intervals and tests can fail completely. There is not often enough power to suggest that the corresponding tests and confidence intervals give correct results. In such cases, the parametric likelihood methods may not be valid. Under the pressure that the correct assumption of parametric family of distribution for the data has to be made, many statisticians turn to nonparametric inferences. Nonparametric methods give valid tests or confidence intervals without having to make strong distributional assumption. Each nonparametric method has its advantage. Empirical likelihood, newly developed by A. Owen

[1]Doctoral student, Department of Statistics, University of Kentucky, Lexington, KY40506, E-mail:kkim@ms.uky.edu

[2]Associate Professor, Department of Statistics, University of Kentucky, Lexington, KY40506, E-mail:mai@ms.uky.edu

(2001), is a nonparametric method that provides the flexibility and effectiveness without having to assume that the data come from a known family of distributions. In empirical likelihood, adapted knowledge from the data plays a central role. The knowledge pooling out from the data can be formed as constraints that restrict the domain of the empirical likelihood function. As such a case in point, we no longer ignore any information obtained from the data. That means that we shall use all information arising from the data without discarding any of it. Therefore, we can expect that empirical likelihood offers a distinct improvement over ordinary nonparametric likelihood methods.

Definition *(Owen (2001)) For given n i.i.d observations $X_1, \cdots, X_n$ with common distribution function $F_X(t)$, the nonparametric or empirical likelihood of the distribution function $F_X(t)$ is*

$$L(F) = \prod_{i=1}^{n} w_i \tag{1}$$

where $w_i = \Delta F_X(x_i)$ is the probability of getting the value x_i in a sample and $\sum_i w_i = 1$.

Definition *(Owen (2001)) For given n i.i.d observations $X_1, \cdots, X_n$, the empirical distribution function of $X_1, \cdots, X_n$ is*

$$F_n = \frac{\sum_{i=1}^{n} \delta_{x_i}}{n}$$

where δ_x denotes the distribution under which $X = x$ with probability 1. Thus $\delta_x(A) = 1_{x \in A}$. The empirical distribution function maximizes L(F).

However, empirical likelihood in some nonparametric settings has difficulties. Defining empirical likelihood for symmetric distributions is such an example. The cause of this difficulty is that the condition on symmetry is equivalent to an infinite number of conditional constraints. In this case, the NPMLE does not exist or there are many NPMLE's having the same empirical likelihood value.

For example, suppose X_1, X_2, X_3 are ordered i.i.d observations from a continuous symmetric distribution $F_0(t)$ around c, which is unknown. It is easy to show that $P(F_n \textit{ is symmetric}) = 0$ where F_n is the empirical distribution function. By adding a number of jump points to F_n, we can find $F_n^*(t)$ which is symmetric and maximize the empirical likelihood among all symmetric distributions. There are a lot of candidates of $F_n^*(t)$'s that have the same maximum empirical likelihood value. If we believe that c is located in the middle of X_2 and X_3, one more artificial observation X on the right side of X_3 is needed to make $F_n^*(t)$ symmetric about c with the maximum likelihood value $L(F) = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4}$. Otherwise, if we believe that c is possibly located in the middle of X_1 and X_2, one more artificial observations X is needed to add on the left side of X_1 so that $F_n^*(t)$ is symmetric about c with the maximum likelihood value $L(F) = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4}$. Both are NPMLE's having the same empirical likelihood value $L(F) = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4}$. That implies that $F_n^*(t)$ is not unique. There are a lot of candidates that achieve the same maximum likelihood value.

We propose to use the method of "envelope empirical likelihood" (Zhou, 2000) to get a sequence of parameter spaces by first enlarging the parameter space so that the NPMLE uniquely exists. Then we shrink those enlarged parameter spaces by imposing a number of distinct known points of symmetry as side information. This constrained or reduced parameter space is called the envelope parameter space. The NPMLE is now well defined on the constrained parameter space. The property of the so defined likelihood ratio test and the resulting NPMLE is the focus of this paper. Asymptotic distributional properties are studied in sections 2 and 3. Examples and case study are presented in section 4.

To illustrate the possible use of the envelope empirical likelihood and difficulty/inefficiency of other available nonparametric method, think of two independent random samples both symmetrically distributed; one with a Cauchy distribution, the other a normal distribution. The goal is to estimate the difference of the two point of symmetry.

Finally we point out that the method proposed in this paper can easily (at least in principle) be generalized to handle higher dimensional data.

1. Envelope empirical likelihood ratio for symmetric distributions

1.1 Envelope empirical likelihood in a two sample problem

Suppose $X_1, \cdots, X_{n_1}$ are n_1 i.i.d. observations from a symmetric distribution F_1 with an arbitrary shift location parameter θ_1 (i.e. the center of symmetry of F_1 is θ_1) and $Y_1, \cdots, Y_{n_2}$ are n_2 i.i.d observations from a symmetric distribution F_2 with an arbitrary shift location parameter θ_2.

Define $w_i = \Delta F_1(t_{1i})$, the probability of getting the value t_{1i} in a sample from the distribution function $F_1(t)$. Define $p_j = \Delta F_2(t_{2j})$, the probability of getting the value t_{2j} in a sample from the distribution function $F_2(t)$. As discussed in the introduction, among all symmetric distributions based on the samples always yields many candidates of both F_1, F_2 and θ_1, θ_2 with the same likelihood value, when the true distribution is continuous, maximizing the log empirical likelihood defined by

$$\log_H Lik = \sum_{i=1}^{n_1} \log w_i + \sum_{j=1}^{n_2} \log p_j \ .$$

The NPMLE is not well defined.

We first enlarge the parameter space so that just one possible NPMLE can exist in this space. The enlarged parameter space is the set of all distributions, symmetric or not. Define those enlarged parameter spaces

$$\Theta_1 = \{F_1 : all\ distribution\ without\ any\ restriction\}$$

and

$$\Theta_2 = \{F_2 : all\ distribution\ without\ any\ restriction\} \, .$$

But these parameter spaces are too large, without restrictions of symmetry. We then shrink the enlarged parameter space by imposing many (but fixed number) constraints that the distributions are symmetric. As more constraints are imposed, the sequence of the parameter space shrinks. We show that the NPMLE is now well defined on the shrank spaces. We call this NPMLE, the envelope nonparametric likelihood estimator.

In this paper, we are interested to study the property of the NPMLE for the difference of two parameters, $\theta_1 - \theta_2$ in the constrained parameter space and also the envelope empirical likelihood ratio test.

The constraints are formed as:

For given t_{1i},

$$for\ some\ \theta_1, \quad F_1(\theta_1 - t_{1i}) = 1 - F_1(\theta_1 + t_{1i}), \qquad i = 1, 2, \ldots, m_1.$$

on the parameter space $\Theta_1 = \{F_1 : all\ \ distributions\}$. Now we put constraints of symmetry on the parameter space as an integration:

$$for\ some\ \theta_1, \int_{-\infty}^{\theta_1 - t_{1i}} dF_1(t) = \int_{\theta_1 + t_{1i}}^{\infty} dF_1(t), \qquad i = 1, 2, \ldots, m_1. \tag{2}$$

If we take the functions $g_{1i}(t) = I_{[t \leq \theta_1 - t_{1i}]}$ and $g^*_{1i}(t) = I_{[t \geq \theta_1 + t_{1i}]}$ in the above, the integration can take the form of

$$\int_{-\infty}^{\infty} g_{1i}(t) dF_1(t) = \int_{-\infty}^{\infty} g^*_{1i}(t) dF_1(t), \qquad i = 1, 2, \ldots, m_1.$$

Similarly, for given t_{2j},

$$for\ some\ \theta_2, \quad F_2(\theta_2 - t_{2j}) = 1 - F_2(\theta_2 + t_{2j}), \qquad j = 1, 2, \ldots, m_2.$$

on the parameter space $\Theta_2 = \{F_2 : all\ \ distributions\}$. Now we put constraints of symmetry on the parameter space as an integration:

$$for\ some\ \theta_2, \int_{-\infty}^{\theta_2 - t_{2j}} dF_2(t) = \int_{\theta_2 + t_{2j}}^{\infty} dF_2(t), \quad j = 1, 2, \ldots, m_2. \tag{3}$$

If we take the functions $g_{2j}(t) = I_{[t \leq \theta_2 - t_{2j}]}$ and $g^*_{2j}(t) = I_{[t \geq \theta_2 + t_{2j}]}$ in the above, the integration can take the form of

$$\int_{-\infty}^{\infty} g_{2j}(t) dF_2(t) = \int_{\infty}^{\infty} g^*_{2j}(t) dF_2(t), \quad j = 1, 2, \ldots, m_2.$$

We can in fact use g and g* that are smooth and define the symmetry similarly.

Maximizing the log empirical likelihood

$$\log_H Lik = \sum_{i=1}^{n_1} \log w_i + \sum_{j=1}^{n_2} \log p_j \tag{4}$$

defined above among distributions in the constrained parameter space

$$\begin{aligned} \Theta^+ \;&=\; \{F_1 : distributions\ \ satisfying\ \ the\ \ equation (2)\} \\ &\bigcap\; \{F_2 : distributions\ \ satisfying\ \ the\ \ equation (3)\} \end{aligned}$$

yields both the envelope empirical NPMLE $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{F}_1(t)$, $\hat{F}_2(t)$. The envelope empirical likelihood ratio can be used to obtain an asymptotic hypothesis test for the difference of two location parameters. Confidence intervals can also be constructed by inverting hypothesis tests.

1.2 Envelope empirical likelihood ratio

Consider testing hypothesis of the difference of two location parameters. We take

$$H_0 : \theta_1 = \theta_2 (\equiv \theta); \qquad vs. \qquad H_A : \theta_1 \neq \theta_2$$

The test statistic we propose is the likelihood ratio statistics

$$T = -2\{ \max_{\theta_1 = \theta_2 = \theta \in \Theta^+} \log_H Lik - \max_{all \theta_1, \theta_2 \in \Theta^+} \log_H Lik \} \quad . \tag{5}$$

We show below that this empirical test statistics will have an approximate chi-square distribution with one degrees of freedom under the null hypothesis. We reject H_0 for larger values of T. Confidence intervals for $(\theta_1 \ - \ \theta_2)$ can be obtained by inverting the chi square test.

2. Asymptotic Results of envelope ELR test statistic and NPMLE

We shall prove in this section that under the null hypothesis our proposed envelope empirical likelihood ratio test statistic T defined in section 1.2 (5) has asymptotically a chi-square distribution with one degree of freedom and obtain the asymptotic distribution of the envelope empirical NPMLE $\hat{\theta}_1$ and $\hat{\theta}_2$.

Denote the column vectors

$$\begin{aligned} g_k(t) &= \{g_{k1}(t), \cdots, g_{km_k}(t)\}^T \ , \quad g_k^*(t) = \{g_{k1}^*(t), \cdots, g_{km_k}^*(t)\}^T \ ; \\ \lambda_k &= \{\lambda_{k1}, \cdots, \lambda_{km_k}\}^T \ , k = 1, 2. \end{aligned}$$

Lemma 1 *The probabilities of getting the value x_i and y_i in a sample from the distribution function $F_1(t)$ and $F_2(t)$ respectively that maximize the log likelihood function (4) satisfying the constraint (2), (3) with any fixed θ_1 and θ_2 are given by*

$$w_i(\lambda_1(\theta_1), \theta_1) = \frac{1}{n_1 - n_1\lambda_1(\theta_1)^T \cdot (g_1(\theta_1 - t; x_i) - g_1^*(\theta_1 + t; x_i))} \ ; \tag{6}$$

$$p_j(\lambda_2(\theta_2), \theta_2) = \frac{1}{n_2 - n_2\lambda_2(\theta_2)^T \cdot (g_2(\theta_2 - t; y_j) - g_2^*(\theta_2 + t; y_j))} \ . \tag{7}$$

where $\lambda_k(\theta_k)^T \cdot g_k(\theta_k - t; \cdot)$ denote the inner product $\sum_{m_k} \lambda_{km_k}(\theta_k) \cdot g_{km_k}(\theta_k - t; \cdot)$. The $\lambda_k(\theta_k)$ value in the equation (6), (7) above is obtained as the solution of the following m_k equations respectively

$$h_k(\lambda_k(\theta_k), \theta_k) = \{h_{k1}(\lambda_k(\theta_k), \theta_k), \cdots, h_{km_k}(\lambda_k(\theta_k), \theta_k)\}^T \ k = 1, 2,$$

by

$$\begin{aligned} h_{1r}(\lambda_1(\theta_1), \theta_1) &= \sum_i [g_{1r}(\theta_1 - t; x_i) - g_{1r}^*(\theta_1 + t; x_i)] w_i(\lambda_1(\theta_1), \theta_1)) \\ &= \sum_i \frac{[g_{1r}(\theta_1 - t; x_i) - g_{1r}^*(\theta_1 + t; x_i)]}{n_1 - n_1\lambda_1^T(\theta_1) \cdot (g_1(\theta_1 - t; x_i) - g_1^*(\theta_1 + t; x_i))} = 0 \ , \\ h_{2l}(\lambda_2(\theta_2), \theta_2) &= \sum_j [g_{2l}(\theta_2 - t; y_j) - g_{2l}^*(\theta_2 + t; y_j)] p_j(\lambda_2(\theta_2), \theta_2)) \\ &= \sum_j \frac{[g_{2r}(\theta_2 - t; y_j) - g_{2r}^*(\theta_2 + t; y_j)]}{n_2 - n_2\lambda_2^T(\theta_2) \cdot (g_2(\theta_2 - t; y_j) - g_2^*(\theta_2 + t; y_j))} = 0 \ . \end{aligned}$$

PROOF: The result follows from a standard Lagrange multiplier argument applied to (2), (3) and (4). Using Lagrange multipliers $n_k\lambda_k$ and η_k, let

$$\begin{aligned} G &= \sum_{i=1}^{n_1}\log w_i + \sum_{r=1}^{m_1} n_1\lambda_1 \sum_i [g_{1r}(\theta_1 - t; x_i) - g^*_{1r}(\theta_1 + t; x_i)]w_i + \eta_1[1 - \sum_i w_i] \\ &\quad + \sum_{j=1}^{n_2}\log p_j + \sum_{l=1}^{m_2} n_2\lambda_2 \sum_j [g_{2l}(\theta_2 - t; y_j) - g^*_{2l}(\theta_2 + t; y_j)]p_j + \eta_2[1 - \sum_j p_j] \end{aligned}$$

Taking derivatives with respect to w_i and p_j and setting $\dfrac{\partial G}{\partial w_i}$ and $\dfrac{\partial G}{\partial p_j}$ to zero, we obtain

$$\begin{aligned} w_i &= \frac{1}{\eta_1 - n_1\lambda_1^T \cdot (g_1(\theta_1 - t; x_i) - g_1^*(\theta_1 + t; x_i))} \ ; \\ p_j &= \frac{1}{\eta_2 - n_2\lambda_2^T \cdot (g_2(\theta_2 - t; y_j) - g_2^*(\theta_2 + t; y_j))} \ , \end{aligned}$$

And the calculations show $\eta_1 = \mathrm{n}_1 \;\; and \;\; \eta_2 = \mathrm{n}_2$.

Thus we get

$$\begin{aligned} w_i &= \frac{1}{n_1 - n_1\lambda_1^T \cdot (g_1(\theta_1 - t; x_i) - g_1^*(\theta_1 + t; x_i))}; \\ p_j &= \frac{1}{n_2 - n_2\lambda_2^T \cdot (g_2(\theta_2 - t; y_j) - g_2^*(\theta_2 + t; y_j))} \end{aligned}$$

where λ_k must satisfy the following equations

$$0 = h_k(t) = \{h_{k1}(t), \cdots, h_{km_k}(t)\}^T \ ,$$

by

$$\begin{aligned} h_{1r}(t) &= \sum_i [g_{1r}(\theta_1 - t; x_i) - g^*_{1r}(\theta_1 + t; x_i)] \times w_i \\ &= \sum_i \frac{[g_{1r}(\theta_1 - t; x_i) - g^*_{1r}(\theta_1 + t; x_i)]}{n_1 - n_1\lambda_1^T(\theta_1) \cdot (g_1(\theta_1 - t; x_i) - g_1^*(\theta_1 + t; x_i))} = 0 \ ; \qquad (8) \\ h_{2l}(t) &= \sum_j [g_{2l}(\theta_2 - t; y_j) - g^*_{2l}(\theta_2 + t; y_j)] \times p_j \\ &= \sum_j \frac{[g_{2l}(\theta_2 - t; y_j) - g^*_{2l}(\theta_2 + t; y_j)]}{n_2 - n_2\lambda_2^T(\theta_2) \cdot (g_2(\theta_2 - t; y_j) - g_2^*(\theta_2 + t; y_j))} = 0 \ . \qquad (9) \end{aligned}$$

Since the solution w_i and p_j of G depend on (θ_1, λ_1) and (θ_2,λ_2) respectively and also the λ_k, k=1,2 does depend on θ_k as well, we denote that w_i is of the function of θ_1 and $\lambda_1(\theta_1)$ and p_j is of the function of θ_2 and $\lambda_2(\theta_2)$.

$\diamondsuit$

The concern coming after Lemma 1 is the asymptotic behavior of the solution λ_k, k=1,2 we have found in Lemma 1.

Lemma 2 *Under some standard conditions, the solution λ_k of the constraint equations in (8) and (9) under the null hypothesis have the following asymptotic representations:*

(i)

$$\sqrt{N}\lambda_k(\theta_0) \xrightarrow{\mathcal{D}} N(0,\Sigma_k)\ ;\quad \text{as}\ \ N\to\infty, \quad N=\min(n_1,n_2)$$

where θ_0 is the true parameter and Σ_k is defined by (11), k=1,2.

(ii) In addition, assume $g_k(\cdot)$ and $g_k^*(\cdot)$ are smooth and $h_k'(0,\theta_0)$ (a $m_k\times m_k$ matrix defined in (10)) is invertible. For $|\theta_k| = O(1/\sqrt{N})$, we have

$$\lambda_k(\theta_k) = \lambda_k(\theta_0) - h_k'(0,\theta_0)^{-1}G_k(\theta_k-\theta_0)^T + o_p(1/\sqrt{N})$$

where G_k is a $m_k\times 1$ matrix with its columns defined as

$$G_{k\cdot} = \left\{\sum_{i(j)} \frac{g_{k1}'(\theta_0-t) - {g_{k1}'}^*(\theta_0+t)}{n_k}, \cdots, \sum_{i(j)} \frac{g_{km_k}'(\theta_0-t) - {g_{km_k}'}^*(\theta_0+t)}{n_k}\right\}^T .$$

(iii) Similarly under the same conditions of part (ii), we have the solutions $\lambda_k(\theta)$ under the null hypothesis$H_0: \theta_1=\theta_2\equiv\theta$

$$\lambda_k(\theta) = \lambda_k(\theta_0) - h_k'(0,\theta_0)^{-1}G_k(\theta-\theta_0)^T + o_p(1/\sqrt{N})$$

where G_k is a $m_k\times 1$ matrix with its columns defined as

$$G_{k\cdot} = \left\{\sum_{i(j)} \frac{g_{k1}'(\theta_0-t) - {g_{k1}'}^*(\theta_0+t)}{n_k}, \cdots, \sum_{i(j)} \frac{g_{km_k}'(\theta_0-t) - {g_{km_k}'}^*(\theta_0+t)}{n_k}\right\}^T .$$

Proof : We show the asymptotic distribution of $\lambda_k(\theta_0)$.

Define a vector function $h_k(\lambda_k(s),s) = (h_{k1}(\lambda_k(s),s),\cdots,h_{km_k}(\lambda_k(s),s))$ by

$$\begin{aligned}
h_{11}(\lambda_1(s),s) &= \sum_i (g_{11}(s-t) - g_{11}^*(s+t))(w_i(\lambda_1(s),s))\ , \\
&\cdots \quad \cdots \\
h_{1m_1}(\lambda_1(s),s) &= \sum_i (g_{1m_1}(s-t) - g_{1m_1}^*(s+t))(w_i(\lambda_1(s),s))\ )\ ; \\
h_{21}(\lambda_2(s),s) &= \sum_j (g_{21}(s-t) - g_{21}^*(s+t))(p_j(\lambda_2(s),s))\ , \\
&\cdots \quad \cdots \\
h_{2m_2}(\lambda_2(s),s) &= \sum_j (g_{2m_2}(s-t) - g_{2m_2}^*(s+t))(p_j(\lambda_2(s),s))\ )\ .
\end{aligned}$$

Then $\lambda_k(\theta_0)$ is the solution of $h_k(\lambda_k(s), s) = 0$ where θ_0 is the true parameter of θ so that $\lambda_k(\theta_0)$ is a consistent statistic of $\lambda_k(\theta_k)$ and $\lambda_k(\theta_0)$ is sufficiently very small so that the expansion below is valid. Thus we have $0 = \mathrm{h}_k(\lambda_k(\theta_0), \theta_0) = h_k(0, \theta_0) + (\lambda_k(\theta_0) - 0\ ,\ \theta_0 - \theta_0) \times [h'_k(0, \theta_0)\ ,\ G_{k\cdot}]^T + o_p(1/\sqrt{N})$, where $h'_k(0, \theta_0)$ is a $m_k \times m_k$ matrix. Rearranging the above equation, we obtain the following equation

$$\sqrt{N}\lambda_k(\theta_0) = -h'_k(0, \theta_0)^{-1}(\sqrt{N}h_k(0, \theta_0)) + o_p(1)$$

where the vector function $h_k(0, \theta_0) = (h_{k1}(0, \theta_0), \cdots, h_{km_k}(0, \theta_0))$. The elements of $h'_k(0, \theta_0)$ are explicitly computed as

$$\begin{aligned} h'_{1rl}(0, \theta_0) &= \sum_i \frac{(g'_{1r}(\theta_0 - t) - {g'_{1r}}^*(\theta_0 + t))(g'_{1l}(\theta_0 - t) - {g'_{1l}}^*(\theta_0 + t))}{n_1}\ ; \\ h'_{2rl}(0, \theta_0) &= \sum_j \frac{(g'_{2r}(\theta_0 - t) - {g'_{2r}}^*(\theta_0 + t))(g'_{2l}(\theta_0 - t) - {g'_{2l}}^*(\theta_0 + t))}{n_2}\ . \end{aligned} \tag{10}$$

Notice we have $n_k h'_{krl} = -D_{krl}$ where

$$-\mathbf{D_k} = \begin{pmatrix} n_k h'_k & G_k \\ G_k^T & 0 \end{pmatrix}.$$

By WLLN,

$$\begin{aligned} h_{1r}(0, \theta_0) = \sum_i \frac{g_{1r}(\theta_0 - t) - g^*_{1r}(\theta_0 + t)}{n_1} \xrightarrow{\mathcal{P}} \quad & E_{F_1}(g_1(\theta_0 - t) - g^*_1(\theta_0 + t)) \\ &= E(g_1(\theta_0 - t)) - E(g^*_1(\theta_0 + t)) \\ &= F_1(\theta_0 - t) - [1 - F_1(\theta_0 + t)] = 0; \\ h_{2l}(0, \theta_0) = \sum_j \frac{g_{2l}(\theta_0 - t) - g^*_{2l}(\theta_0 + t)}{n_2} \xrightarrow{\mathcal{P}} \quad & E_{F_2}(g_2(\theta_0 - t) - g^*_2(\theta_0 + t)) \\ &= E(g_2(\theta_0 - t)) - E(g^*_2(\theta_0 + t)) \\ &= F_2(\theta_0 - t) - [1 - F_2(\theta_0 + t)] = 0 \end{aligned}$$

by the assumption of symmetric distributions and

$$cov(h_{krl}(0, \theta_0)) = E(h_{krr}(0, \theta_0) \times h_{kll}(0, \theta_0)).$$

Applying the Cramer-Wold Device, we can show that

$$\sqrt{N}h_k(0, \theta_0) \xrightarrow{\mathcal{D}} N(0, \Sigma_{h_k})$$

with $\Sigma_{h_k} = \lim h'_k(0, \theta_0)$.

Finally, we have put all together so that

$$\sqrt{N}\lambda_k(\theta_0) = [h_k'(0,\theta_0)]^{-1}(-\sqrt{N}h_k(0,\theta_0)) + o_p(1) \xrightarrow{\mathcal{D}} N(0,\Sigma_k)$$

with

$$\Sigma_k = \lim[h'(0,\theta_0)]^{-1} \ . \tag{11}$$

Recall $n_k h_{krl}' = -D_{krl}$, we see that $\Sigma_k^{-1} = \lim[h_k'(0,\theta_0)] = \Sigma_{h_k} = D_k^*$.

This completes the proof of (i).

Let's look at the second part of Lemma.

Define a vector function $h_k(\lambda_k(s), s) = (h_{k1}(\lambda_k(s), s), \cdots, h_{km_k}(\lambda_k(s), s))$ by

$$\begin{aligned}
h_{11}(\lambda_1(s), s) &= \sum_i (g_{11}(t) - g_{11}^*(t))(w_i(\lambda_1(s), s)) \ , \\
&\cdots \quad \cdots \\
h_{1m_1}(\lambda_1(s), s) &= \sum_i (g_{1m_1}(t) - g_{1m_1}^*(t))(w_i(\lambda_1(s), s))) \ ; \\
h_{21}(\lambda_2(s), s) &= \sum_j (g_{21}(t) - g_{21}^*(t))(p_j(\lambda_2(s), s)) \ , \\
&\cdots \quad \cdots \\
h_{2m_2}(\lambda_2(s), s) &= \sum_j (g_{2m_2}(t) - g_{2m_2}^*(t))(p_j(\lambda_2(s), s))) \ .
\end{aligned}$$

The $\lambda_k(\theta_k)$ and $\lambda_k(\theta_0)$ are the solution of $h_k(\lambda_k(\theta_k), \theta_k) = 0$ and $h_k(\lambda(\theta_0), \theta_0) = 0$ respectively where $\lambda_k(\theta_0)$ is small, so that the expansion below is valid. Thus we have expanded the above equations at $\lambda_k(\theta_0) = 0$ as follows $0 = \mathrm{h}_k(\lambda_k(\theta_k), \theta_k) = h(0,\theta_0) + (\lambda_k(\theta_k) - 0 \ , \ \ \theta_k - \theta_0) \times [h_k'(0,\theta_0) \ , \ \ G_k]^T + o_p(1/\sqrt{N})$, and $0 = \mathrm{h}_k(\lambda_k(\theta_0), \theta_0) = h(0,\theta_0) + (\lambda_k(\theta_0) - 0 \ , \ \ \theta_0 - \theta_0) \times [h_k'(0,\theta_0) \ , \ \ G_k]^T + o_p(1/\sqrt{N})$, where $h_k'(0,\theta_0)$ is a $m_k \times m_k$ matrix and G_k is a $m_k \times 1$ matrix with its columns defined as

$$\begin{aligned}
G_1\cdot &= \left\{ \sum_i \frac{g_{11}'(\theta_0 - t) - g_{11}'{}^*(\theta_0 + t)}{n_1}, \cdots, \sum_i \frac{g_{1m_1}'(\theta_0 - t) - g_{1m_1}'{}^*(\theta_0 + t)}{n_1} \right\}^T , \\
G_2\cdot &= \left\{ \sum_j \frac{g_{21}'(\theta_0 - t) - g_{21}'{}^*(\theta_0 + t)}{n_2}, \cdots, \sum_j \frac{g_{2m_2}'(\theta_0 - t) - g_{2m_2}'{}^*(\theta_0 + t)}{n_2} \right\}^T .
\end{aligned}$$

Then it finally turns out that

$$\begin{aligned}
0 &= h_k(\lambda_k(\theta_k), \theta_k) - h_k(\lambda(\theta_0), \theta_0) \\
&= h_k'(0,\theta_0)[\lambda_k(\theta_k) - \lambda_k(\theta_0)] + G_k^T(\theta_k - \theta_0) + o_p(1/\sqrt{N}) \ ,
\end{aligned}$$

and therefore

$$\lambda_k(\theta_k) - \lambda_k(\theta_0) = -[h'_k(0,\theta_0)]^{-1} G_k^T(\theta_k - \theta_0) + o_p(1/\sqrt{N}) \ .$$

We have

$$\sqrt{N}\lambda_k(\theta_k) = \sqrt{N}\lambda_k(\theta_0) - [h'_k(0,\theta_0)]^{-1}(\sqrt{N} G_k^T(\theta_k - \theta_0)) + o_p(1) \ .$$

The second part of lemma is completely shown.

The third part of lemma can be proved similar to the second part of lemma. ♢

Theorem 1 *Under the same conditions in Lemma 2, the test statistics T has asymptotically a chi-square distribution with one degrees of freedom.*

PROOF: Let's define

$$\begin{aligned} f_1(\lambda_1(\theta_1), \theta_1) &= \sum_i \log w_i(\lambda_1(\theta_1), \theta_1), \\ f_2(\lambda_2(\theta_2), \theta_2) &= \sum_j \log p_j(\lambda_2(\theta_2), \theta_2) \end{aligned}$$

Then the log empirical likelihood ratio statistic is the form of

$$T = -2 \min_{\theta \in \Theta^+} \{ f_1(\lambda_1(\theta), \theta) + f_2(\lambda_2(\theta), \theta) \} + 2 \min_{\forall \theta_1, \theta_2 \in \Theta^+} \{ f_1(\lambda_1(\theta_1), \theta_1) + f_2(\lambda_2(\theta_2), \theta_2) \} \ .$$

By Taylor expansion at $\theta_k = \theta_0$ and $\lambda_k(\theta_k)$=$\lambda_k(\theta_0)$, the test statistic T can be expended as follows:

$$\begin{aligned} T = -2 \min_{\theta \in \Theta^+} \{ \quad & f_1(0,\theta_0) + (\lambda_1(\theta) \ , \ \theta - \theta_0)(\frac{\partial}{\partial \lambda_1(\theta)} f_1(\lambda_1(\theta), \theta) \ , \ \frac{\partial}{\partial \theta} f_1(\lambda_1(\theta), \theta))^T|_{\lambda_1(\theta)=0, \theta=\theta_0} \\ & +1/2(\lambda_1(\theta) \ , \ \theta - \theta_0) D_1(\lambda_1(\theta) \ , \ \theta - \theta_0)^T + o_p(1) \\ & +f_2(0,\theta_0) + (\lambda_2(\theta) \ , \ \theta - \theta_0)(\frac{\partial}{\partial \lambda_2(\theta)} f_2(\lambda_2(\theta), \theta) \ , \ \frac{\partial}{\partial \theta} f_2(\lambda_2(\theta), \theta))^T|_{\lambda_2(\theta)=0, \theta=\theta_0} \\ & +1/2(\lambda_2(\theta) \ , \ \theta - \theta_0) D_2(\lambda_2(\theta) \ , \ \theta - \theta_0)^T + o_p(1) \} \\ +\, 2 \min_{\theta_1 \in \Theta^+} \{ \quad & f_1(0,\theta_0) + (\lambda_1(\theta_1) \ , \ \theta_1 - \theta_0)(\frac{\partial}{\partial \lambda_1(\theta_1)} f_1(\lambda_1(\theta_1), \theta_1) \ , \ \frac{\partial}{\partial \theta_1} f_1(\lambda_1(\theta_1), \theta_1))^T|_{\lambda_1(\theta_1)=0, \theta_1=\theta_0} \\ & +1/2(\lambda_1(\theta_1) \ , \ \theta_1 - \theta_0) D_1(\lambda_1(\theta_1) \ , \ \theta_1 - \theta_0)^T + o_p(1) \} \\ +\, 2 \min_{\theta_2 \in \Theta^+} \{ \quad & f_2(0,\theta_0) + (\lambda_2(\theta_2) \ , \ \theta_2 - \theta_0)(\frac{\partial}{\partial \lambda_2(\theta_2)} f_2(\lambda_2(\theta_2), \theta_2) \ , \ \frac{\partial}{\partial \theta_2} f_2(\lambda_2(\theta_2), \theta_2))^T|_{\lambda_2(\theta_2)=0, \theta_2=\theta_0} \\ & +1/2(\lambda_2(\theta_2) \ , \ \theta_2 - \theta_0) D_2(\lambda_2(\theta_2) \ , \ \theta_2 - \theta_0)^T + o_p(1) \} \end{aligned} \tag{12}$$

where D_k denotes the $(m_k+1) \times (m_k+1)$ matrix of second derivatives of $f_k(\lambda_k(\theta_k), \theta_k)$ with respect to $\lambda_k(\theta_k)$ and θ_k, i.e.

$$\mathbf{D_k} = \begin{pmatrix} \frac{\partial^2}{\partial^2 \lambda_k(\theta_k)} f_k(\lambda_k(\theta_k), \theta_k)|_{\lambda_k(\theta_k)=0, \theta_k=\theta_0} & \frac{\partial^2}{\partial \lambda_k(\theta_k) \partial \theta_k} f_k(\lambda_k(\theta_k), \theta_k)|_{\lambda_k(\theta_k)=0, \theta_k=\theta_0} \\ \frac{\partial^2}{\partial \lambda_k(\theta_k) \partial \theta_k} f_k(\lambda_k(\theta_k), \theta_k)|_{\lambda_k(\theta_k)=0, \theta_k=\theta_0} & \frac{\partial^2}{\partial^2 \theta_k} f_k(\lambda_k(\theta_k), \theta)|_{\lambda_k(\theta_k)=0, \theta_k=\theta_0} \end{pmatrix} .$$

The expansion are valid in view of Lemma 2.

Notice we have

$$\lambda_k(\theta_k) = \lambda_k(\theta_0) - [h'_k(0,\theta_0)]^{-1} G_k^T(\theta_k - \theta_0) + o_p(1) \ .$$

and

$$\lambda_k(\theta) = \lambda_k(\theta_0) - [h'_k(0,\theta_0)]^{-1} G_k^T(\theta - \theta_0) + o_p(1) \ .$$

(see Lemma 2).

We substitute the Taylor expansions of $\lambda_k(\theta_k)$ to the last form of the test statistics T. Define

$$\begin{aligned} (\lambda_k(\theta_k)\ ,\ \theta_k - \theta_0) &= (\lambda_k(\theta_0) - [h'_k(0,\theta_0)]^{-1} G_k^T(\theta_k - \theta_0)\ ,\ \theta - \theta_0) \\ &= (\lambda_k(\theta_0)\ ,\ 0) - ([h'_k(0,\theta_0)]^{-1} G_k^T\ ,\ -1)(\theta_k - \theta_0) \\ &= V_{k1} - V_{k2}(\theta_k - \theta_0) \end{aligned} \tag{13}$$

and

$$\left(\frac{\partial}{\partial \lambda_k(\theta_k)} f_k(\lambda_k(\theta_k),\theta_k)|_{\lambda_k(\theta_k)=0,\theta_k=\theta_0}\ ,\ \frac{\partial}{\partial \theta_k} f_k(\lambda_k(\theta_k),\theta_k)|_{\lambda_k(\theta_k)=0,\theta_k=\theta_0}\right)^T = V_{k3}. \tag{14}$$

By the equation (13) and (14) above, the test statistic T can be rewritten to

$$\begin{aligned} T = \quad & -2 \min_{\theta\in\Theta^+} \{(V_{11} - V_{12}(\theta-\theta_0))^T V_{13} + (V_{11} - V_{12}(\theta-\theta_0))^T D_1 (V_{11} - V_{12}(\theta-\theta_0)) \\ & + (V_{21} - V_{22}(\theta-\theta_0))^T V_{23} + (V_{21} - V_{22}(\theta-\theta_0))^T D_2 (V_{21} - V_{22}(\theta-\theta_0))\} \\ & +2 \min_{\theta_1\in\Theta^+} \{(V_{11} - V_{12}(\theta_1-\theta_0))^T V_{13} + (V_{11} - V_{12}(\theta_1-\theta_0))^T D_1 (V_{11} - V_{12}(\theta_1-\theta_0))\} \\ & +2 \min_{\theta_2\in\Theta^+} \{(V_{21} - V_{22}(\theta_2-\theta_0))^T V_{23} + (V_{21} - V_{22}(\theta_2-\theta_0))^T D_2 (V_{21} - V_{22}(\theta_2-\theta_0))\} \\ & + o_p(1) \ . \end{aligned}$$

Denote $\hat\theta$ the NPMLE that minimizes $f_1(\lambda_1(\theta),\theta) + f_2(\lambda_2(\theta),\theta)$. Similarly denote $\hat\theta_k$ the NPMLE that minimizes $f_k(\lambda_k(\theta_k),\theta_k)$, k=1,2.

Considering that the $o_p(1)$ term can be ignored, we find the minimization over θ and θ_k and obtain the minimum value (see Lemma 4 on the appendix). The minimum value of the test statistic T is the quadratic form of

$$\begin{aligned} T &= \{(V_{12}^T(-D_1)V_{12})(\hat\theta - \hat\theta_1)^2\} + \{(V_{22}^T(-D_2)V_{22})(\hat\theta - \hat\theta_2)^2\} + o_p(1) & (15) \\ &= \{A_1(\hat\theta - \hat\theta_1)^2\} + \{A_2(\hat\theta - \hat\theta_2)^2\} + o_p(1)\ , & (16) \end{aligned}$$

where $A_k = (V_{k2}^T(-D_k)V_{k2})$ for the convenience. With the very simple form (15) or (16) of the test statistic T above, we are to show the asymptotic distribution of the statistic T.

Now, recall the distribution of $\lambda_k(\theta_0)$ in Lemma 2 (ii). You can forward to have the asymptotic distribution of $\hat{\theta} - \hat{\theta}_k$ in Theorem 2.

Define

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\mathbf{\Gamma} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^T & \Gamma_{22} \end{pmatrix}$$

$$\begin{aligned} \Gamma_{kk} &= \sqrt{N} cov(\sqrt{V_{k2}^T(-D_k)V_{k2}}(\hat{\theta} - \hat{\theta}_k)), \quad k = 1, 2 \ ; \\ \Gamma_{12} &= \sqrt{N} cov(\sqrt{V_{12}^T(-D_1)V_{12}}(\hat{\theta} - \hat{\theta}_1) \ , \ \sqrt{V_{22}^T(-D_2)V_{22}}(\hat{\theta} - \hat{\theta}_2)) \ , \end{aligned}$$

with $rank(M\Gamma) = trace(M\Gamma) = \Gamma_{11} + \Gamma_{22}$ due to $M\Gamma$ being idempotent (see Lemma 5 on the appendix) and $Y^T = (A_1^{1/2}(\hat{\theta} - \hat{\theta}_1) \ , \ A_2^{1/2}(\hat{\theta} - \hat{\theta}_2))$ where

$$\sqrt{N} Y \xrightarrow{\mathcal{D}} N(0, \Gamma) \ .$$

That shows that the log empirical likelihood ratio statistic T is

$$T = Y^T M Y \xrightarrow{\mathcal{D}} \chi^2_{(\Gamma_{11})} + \chi^2_{(\Gamma_{22})} = \chi^2_1 \ .$$

(see Theorem 3 on the appendix for more details)

The proof is completely done. $\diamondsuit$

Theorem 2 *Under the same conditions of Theorem 1 except we assume the alternative hypothesis is true with $\hat{\theta}$ that is the NPMLE of θ under the null hypothesis:*

(i) The asymptotic distribution of the envelope estimator $(\hat{\theta} - \hat{\theta}_k)$ is given by

$$\sqrt{N}\sqrt{V_{k2}^T(-D_k)V_{k2}}(\hat{\theta} - \hat{\theta}_k)^T \xrightarrow{\mathcal{D}} N(0, \Gamma_{kk}) \ ,$$

where

$$\Gamma_{kk} = \sqrt{N} cov(\sqrt{V_{k2}^T(-D_k)V_{k2}}(\hat{\theta} - \hat{\theta}_k)), k = 1, 2 \ .$$

(ii) The asymptotic joint distribution of the envelope estimator $(\hat{\theta} - \hat{\theta}_1 \ , \ \hat{\theta} - \hat{\theta}_2)^T$ is given by

$$\sqrt{N}(\sqrt{V_{12}^T(-D_1)V_{12}}(\hat{\theta} - \hat{\theta}_1) \ , \ \sqrt{V_{22}^T(-D_2)V_{22}}(\hat{\theta} - \hat{\theta}_2))^T \xrightarrow{\mathcal{D}} N(0, \Gamma) \quad ,$$

where

$$\mathbf{\Gamma} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^T & \Gamma_{22} \end{pmatrix}$$

$$\begin{aligned} \Gamma_{kk} &= \sqrt{N} cov(\sqrt{V_{k2}^T(-D_k)V_{k2}}(\hat{\theta} - \hat{\theta}_k)), \quad k = 1, 2 \quad ; \\ \Gamma_{12} &= \sqrt{N} cov(\sqrt{V_{12}^T(-D_1)V_{12}}(\hat{\theta} - \hat{\theta}_1) \ , \ \sqrt{V_{22}^T(-D_2)V_{22}}(\hat{\theta} - \hat{\theta}_2)) \quad . \end{aligned}$$

Proof: Since $\hat{\theta}$ and $\hat{\theta}_k$, k=1,2 are correlated each other, we need to find the joint distribution of $\hat{\theta}$ and $\hat{\theta}_k$, k=1,2.

Aside from the $o_p(1/\sqrt{N})$ term, the $\hat{\theta}$ and $\hat{\theta}_k$ that achieves the minimum are computed as

$$\hat{\theta} = \frac{\sum_k (V_{k2}^T(D_k)V_{k1} + V_{k2}^T V_{k3})}{\sum_k V_{k2}^T(D_k)V_{k2}} \ ,$$

and

$$\hat{\theta}_k = \frac{V_{k2}^T(D_k)V_{k1} + V_{k2}^T V_{k3}}{V_{k2}^T(D_k)V_{k2}} \ .$$

(see Lemma 4 on the appendix)

Denote $A_k = V_{k2}^T(-D_k)V_{k2}$ and $B_k = V_{k2}^T(-D_k)V_{k1} - V_{k2}^T V_{k3}$, k=1,2.

Recall that

$$\begin{aligned} G_1\cdot &= \left\{ \sum_i \frac{g_{11}'(\theta_0 - t; x_i) - {g_{11}'}^*(\theta_0 + t; x_i)}{n_1}, \ \cdots \ , \sum_i \frac{g_{1m_1}'(\theta_0 - t; x_i) - {g_{1m_1}'}^*(\theta_0 + t; x_i)}{n_1} \right\}^T \ , \\ G_2\cdot &= \left\{ \sum_j \frac{g_{21}'(\theta_0 - t; y_j) - {g_{21}'}^*(\theta_0 + t; y_j)}{n_2}, \ \cdots \ , \sum_j \frac{g_{2m_2}'(\theta_0 - t; y_j) - {g_{2m_2}'}^*(\theta_0 + t; y_j)}{n_2} \right\}^T \ . \end{aligned}$$

By SLLN,

$$G_k \xrightarrow{\mathcal{P}} E(g_k'(\theta_0 - t; \cdot) - {g_k'}^*(\theta_0 + t; \cdot)) = C_k$$

where C_k is a constant in R. Therefore,

$$\begin{aligned} A_k = V_{k2}^T(-D_k)V_{k2} &= (n_k - 2) G_k^T h_k'(0, \theta_0)^{-1} G_k \\ &= (\frac{n_k - 2}{n_k}) G_k^T (n_k h_k'(0, \theta_0)^{-1}) G_k \\ &\xrightarrow{\mathcal{P}} C_k^T \Sigma_k C_k \ , \quad k = 1, 2 \end{aligned}$$

and in view of Lemma 2(i)

$$\begin{aligned}
\sqrt{N}B_k &= \sqrt{N}(V_{k2}^T(-D_k)V_{k1} - V_{k2}^T V_{k3}) \\
&= -\sqrt{N}((n_k)G_k^T h_k'(0,\theta_0)^{-1}h_k'(0,\theta_0)\lambda_k(\theta_0) + G_k^T\lambda_k(\theta_0) + n_k G_k^T\lambda_k(\theta_0)) \\
&= \sqrt{N}(G_k^T\lambda_k(\theta_0)) \xrightarrow{\mathcal{D}} N(0, C_k^T\Sigma_k C_k)\ , \quad k=1,2
\end{aligned}$$

Thus we have

$$\sqrt{N}\sqrt{A_k}(\hat{\theta}-\hat{\theta}_k)^T \xrightarrow{\mathcal{D}} N(0,\Gamma_{kk})\ ,$$

where

$$\begin{aligned}
\Gamma_{11} &= \sqrt{N}cov(\sqrt{V_{12}^T(-D_1)V_{12}}(\hat{\theta}-\hat{\theta}_1)) \xrightarrow{\mathcal{P}} \frac{C_2^T\Sigma_2C_2}{\sum_{k=1}^2(C_k^T\Sigma_kC_k)}\ ; \\
\Gamma_{22} &= \sqrt{N}cov(\sqrt{V_{22}^T(-D_2)V_{22}}(\hat{\theta}-\hat{\theta}_2)) \xrightarrow{\mathcal{P}} \frac{C_1^T\Sigma_1C_1}{\sum_{k=1}^2(C_k^T\Sigma_kC_k)}\ .
\end{aligned}$$

This completes the part(i).

We know that $(\hat{\theta}-\hat{\theta}_1)$ and $(\hat{\theta}-\hat{\theta}_2)$ are not independent. In order to find the joint distribution of those two variables, we compute the variance-covariance matrix of those variables

$$\begin{aligned}
\Gamma_{12} &= \sqrt{N}cov(\sqrt{V_{12}^T(-D_1)V_{12}}(\hat{\theta}-\hat{\theta}_1)\ ,\ \sqrt{V_{22}^T(-D_2)V_{22}}(\hat{\theta}-\hat{\theta}_2)) \\
&= \sqrt{N}cov(\sqrt{A_1}(\hat{\theta}-\hat{\theta}_1)\ ,\ \sqrt{A_2}(\hat{\theta}-\hat{\theta}_2)) \\
&= \frac{\sqrt{C_1^T\Sigma_1C_1C_2^T\Sigma_2C_2}}{\sum_k(C_k^T\Sigma_kC_k)}\ .
\end{aligned}$$

Therefore the asymptotic joint distribution of the estimator $(\hat{\theta}-\hat{\theta}_1\ ,\ \hat{\theta}-\hat{\theta}_2)^T$ is given by

$$\sqrt{N}(\sqrt{A_1}(\hat{\theta}-\hat{\theta}_1)\ ,\ \ \sqrt{A_2}(\hat{\theta}-\hat{\theta}_2))^T \xrightarrow{\mathcal{D}} N(0,\Gamma)\quad ,$$

where

$$\mathbf{\Gamma} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^T & \Gamma_{22} \end{pmatrix}.$$

The theorem is completely proved. $\diamondsuit$

3. An illustrative example

3.1 An illustration with small samples

Consider the data that appear in section 6.9 of Snedecor and Cochran (1989), which shows the weight gains (in grams) for two groups of female rats under the two diets. The 12 rats

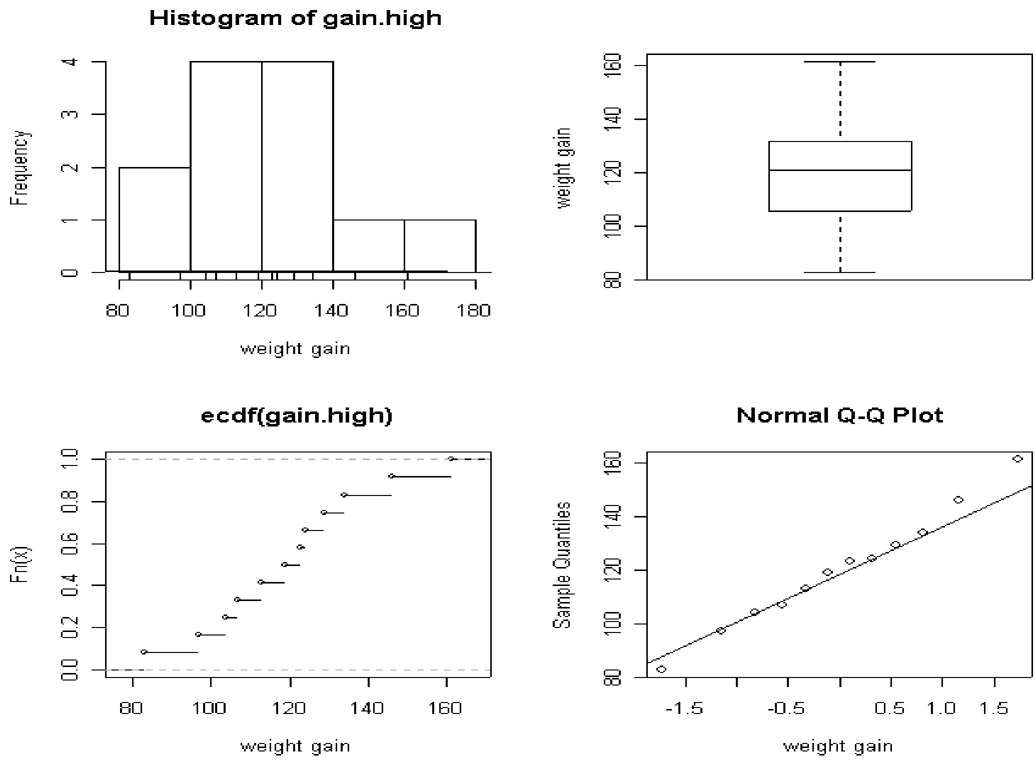

Figure 3.1.1: Descriptive Statistics for weight gain in high protein diet

were given the high protein diet, and the 7 rats were given the low protein diet. Suppose a nutritionist is interested in the relative merits of two diets, one featuring high protein, the other low protein. Do the two diets lead to differences in mean weight gain? The high protein and low protein samples are presumed to have mean-value location parameters, θ_1 and θ_2, and standard deviation scale parameters, σ_1 and σ_2, respectively. We are primarily interested in whether there is any difference in the mean weight gains.

For each sample, we make a set of exploratory data analysis plots, consisting of a histogram, a box plot, an empirical likelihood function, and a normal qq-plot, all displayed in a two-by-two layout in Figure 3.1.1. The resulting plots for the high protein group indicate that the data come from a symmetric distribution, and there is no indication of outliers. As shown in Figure 3.1.2, the plots for the low protein group support the same conclusion.

As a result, we have good reason to believe that the envelope empirical likelihood ratio test for a difference of the two mean weight gains proposed in this paper will provide a valid test for

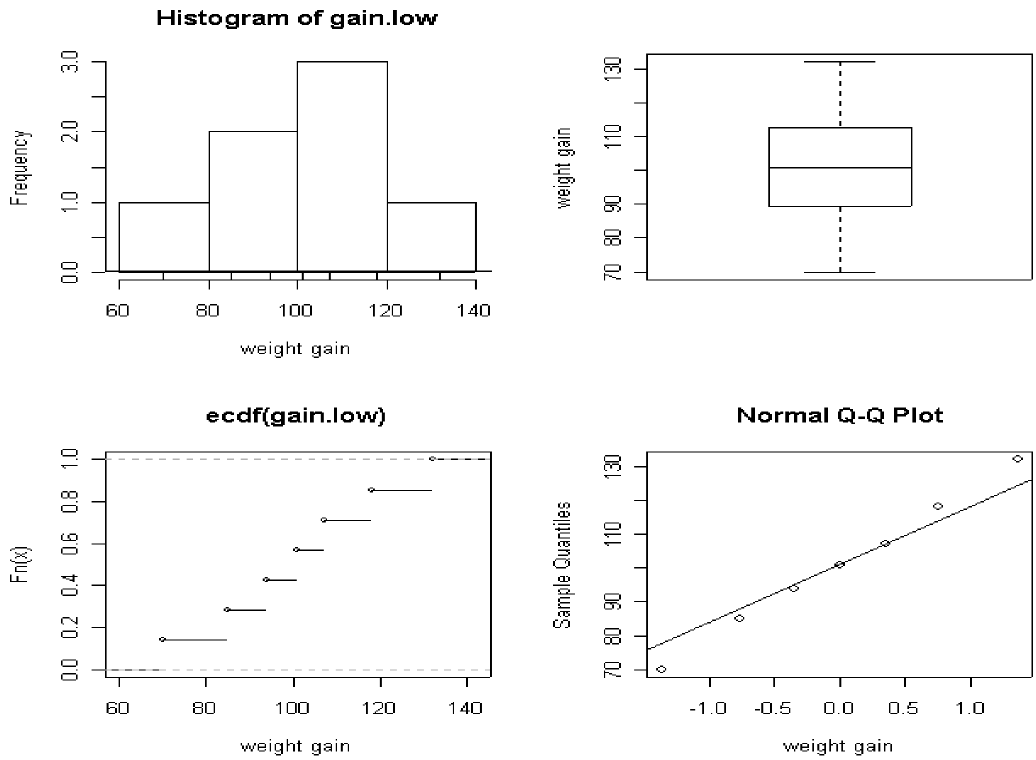

Figure 3.1.2: Descriptive Statistics for weight gain in low protein diet

the hypotheses.

The plots in Figure 3.1.1 and Figure 3.1.2 suggest a set of points of symmetry. We use those points as side constraints. We apply our algorithm to the data set with the side constraints.

Figure 3.1.3 shows the likelihood values of the envelope empirical likelihood function under the corresponding hypothesis. The global maximum value of the envelope empirical likelihood function is chosen and served to the envelope empirical likelihood ratio for testing. With this ratio the corresponding P-value with one degree of freedom is estimated.

According to the corresponding p-value, the null hypothesis of no difference is not rejected. We conclude that there is no difference in the mean weight gains. Also the estimate of NPMLE of θ that maximizes the envelope empirical likelihood function under the alternative hypothesis can be found by choosing the $\hat{\theta}$ corresponding to the maximum likelihood function value.

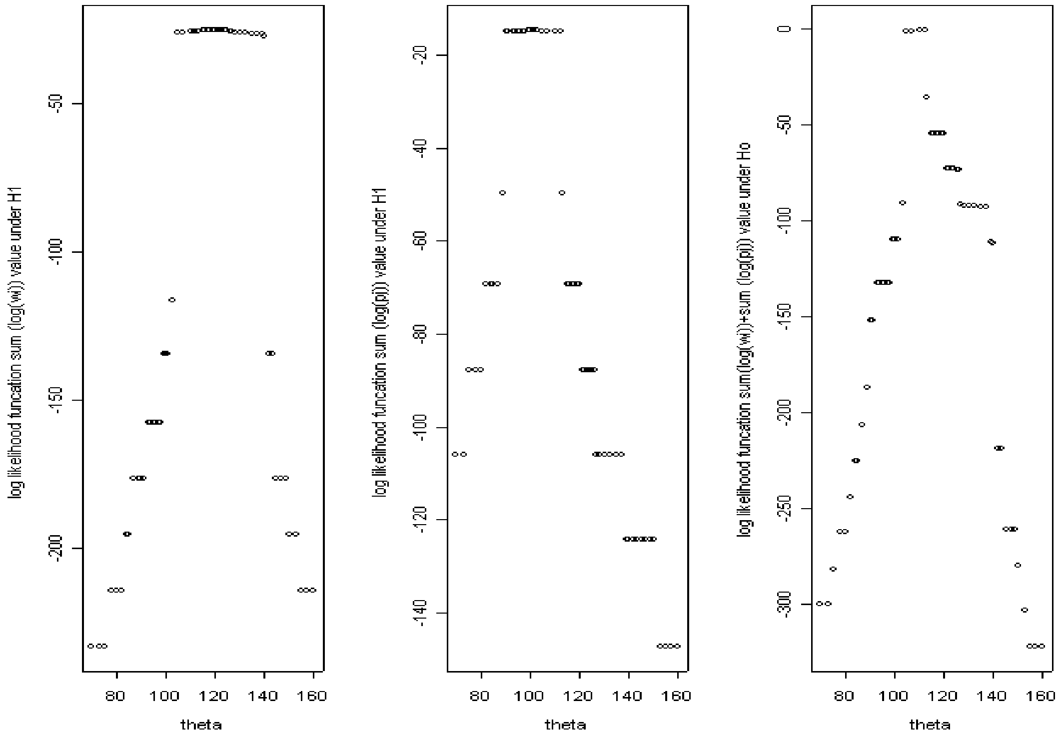

Figure 3.1.3: Log empirical likelihood function value for θ

3.2 An illustration with simulated large samples

We consider an application to the simulation data set, which is generated from Cauchy distribution with location parameter 0 and scale parameter 1. The aim is to identify whether there is a difference of two location parameters. The set of exploratory data analysis plots for each simulated data set is described in Figure 3.2.1 and Figure 3.2.2. The plots show that they are symmetric around 0.

As a result, we have good reason to believe that the envelope empirical likelihood ratio test for a difference of the two location parameters proposed in this paper will provide a valid test for the hypotheses.

In the same manner of section 3.1, the descriptive statistics suggest a set of points of symmetry so that we can use those points as side constraints. We apply our algorithm to the data set with the side constraints. We maximize the log empirical likelihood function under those constraints.

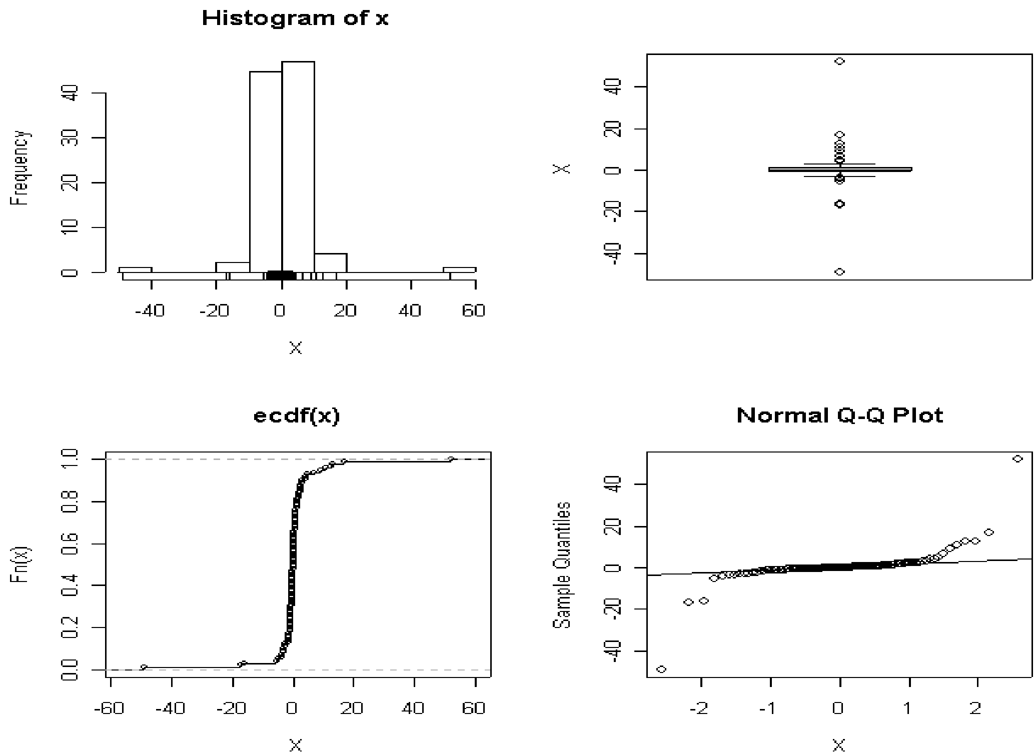

Figure 3.2.1: Descriptive Statistics for simulated data set X

As shown in Figure 3.2.3, there are the global maximum values of the envelope empirical likelihood function under the corresponding hypothesis. Those maximum values serve to estimate the envelope empirical likelihood ratio for testing. With this ratio the corresponding P-value with one degree of freedom is estimated.

According to the corresponding p-value, the null hypothesis of no difference is not rejected. We conclude that there is no difference in the location parameters.

Remark: We have discovered that the smoothness of the log empirical likelihood is subject to a point of symmetry used as side constraint. As the points of symmetry that are far from the center are chosen as side constraints, a lot of observation can be disregarded. Therefore, we recommend opting the points of symmetry that are close to the center as side constraints.

Appendix

Lemma3 Show $f_1'(0, \theta_0) = n_1 h_1(0, \theta_0)$.

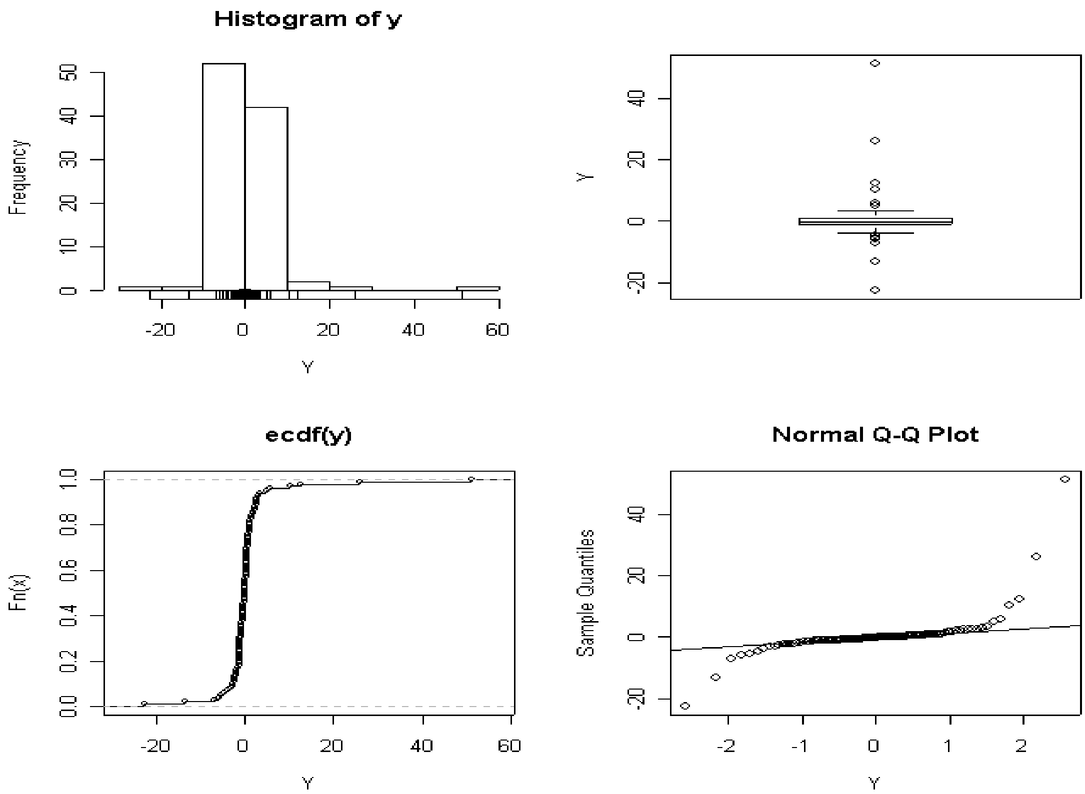

Figure 3.2.2: Descriptive Statistics for simulated data set Y

We compute

$$\begin{aligned}\frac{\partial}{\partial\lambda_1(\theta)} f_1(\lambda_1(\theta),\theta) &= \frac{\partial}{\partial\lambda_1(\theta)} \sum \log(w_i(\lambda_1(\theta),\theta)) \\ &= \sum \frac{n_1(g_1(\theta-t;x_i)-g_1^*(\theta+t;x_i))}{n_1 - n_1\lambda_1^T(\theta)\cdot(g_1(\theta-t;x_i)-g_1^*(\theta+t;x_i))}\end{aligned}$$

So

$$f_1'(0,\theta_0) = \sum g_1(\theta-t;x_i) - g_1^*(\theta+t;x_i) = n_1 h_1(0,\theta_0)$$

Similarly, we can show that $f_2'(0,\theta_0) = h_2(0,\theta_0)$.

By SLLN,

$$\frac{1}{n_1} f_1'(0,\theta_0) \xrightarrow{\mathcal{P}} E(g_1(\theta_0-t;x_i) - g_1^*(\theta_0+t;x_i)) = 0$$

and in view of Lemma 2(i)

$$\sqrt{N}\lambda_1(\theta_0) \xrightarrow{\mathcal{D}} N(0,\Sigma_1)\ , \quad \text{as} \quad N\to\infty.$$

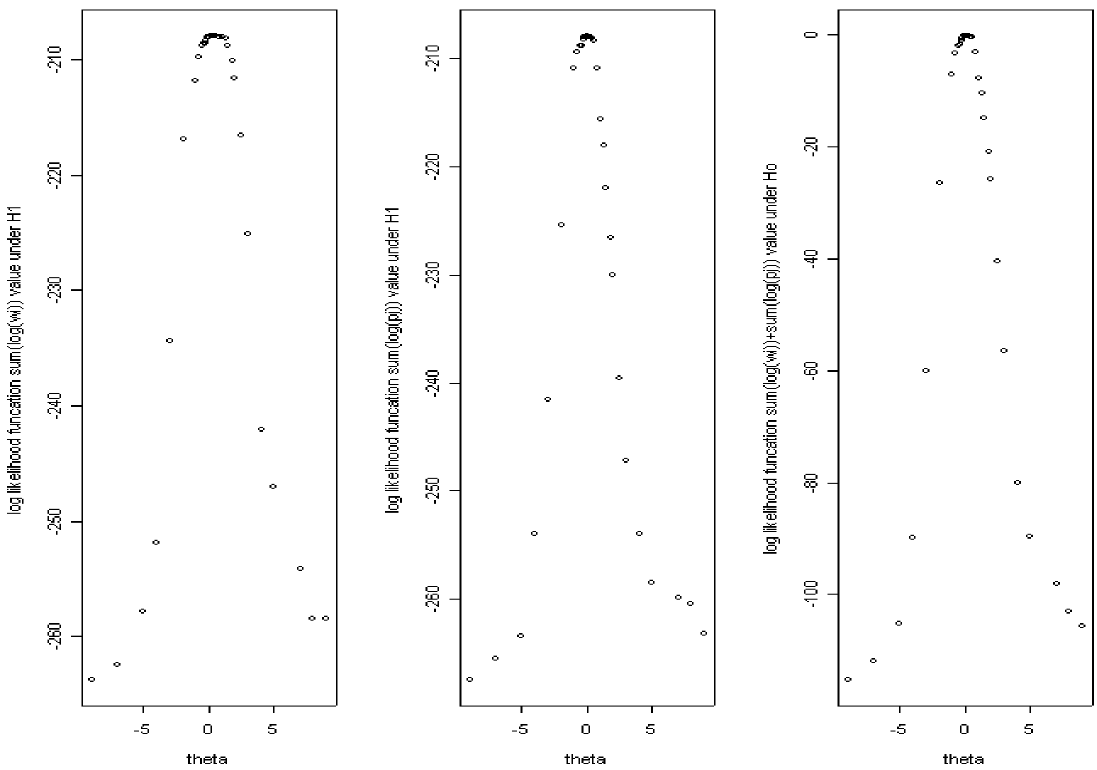

Figure 3.2.3: Log empirical likelihood function value for θ

By Slutsky's theorem,

$$\sqrt{N}\lambda_1^T(\theta_0)\frac{f_1'(0,\theta_0)}{n_1} \xrightarrow{\mathcal{P}} 0 \ , \quad \text{as} \quad N \to \infty.$$

Also taking the second derivatives with respect to $\lambda_1(\theta)$, we now compute $f_1''(0,\theta_0) = n_1 h_1'(0,\theta_0) = D_{1rl}$. The rl^{th} element of the $(m_1+1)\times(m_1+1)$ matrix D_1 is

$$D_{1rl} = \frac{\partial^2}{\partial\lambda_{1r}(\theta)\partial\lambda_{1l}(\theta)} f_1(\lambda_1(\theta),\theta)|_{\lambda_1(\theta)=0} \ .$$

$\diamondsuit$

Lemma4 Suppose D_k is a positive definite matrix of $s \times s$, V_{ki}, i=1,2,3 is a vector of sx1, and θ is a scalar in R.

The minimization of the following equation

$$\begin{aligned} \min_\theta\{ \quad & 2(V_{11}-V_{12}\theta)^T V_{13} + (V_{11}-V_{12}\theta)^T D_1 (V_{11}-V_{12}\theta) \\ & +2(V_{21}-V_{22}\theta)^T V_{23} + (V_{21}-V_{22}\theta)^T D_2 (V_{21}-V_{22}\theta)\} \end{aligned} \tag{17}$$

occurs when θ is $\hat{\theta}$ that is the solution of the following equation $f_1'(\theta) + f_2'(\theta) = 0$ where

$$f_k(\theta) = \{(V_{k1} - V_{k2}\theta)^T V_{k3} + (V_{k1} - V_{k2}\theta)^T D_k (V_{k1} - V_{k2}\theta)\}$$

and

$$f_k'(\theta) = \frac{\partial}{\partial \theta}\{(V_{k1} - V_{k2}\theta)^T V_{k3} + (V_{k1} - V_{k2}\theta)^T D_k (V_{k1} - V_{k2}\theta)\} \ .$$

The value $\hat{\theta}$ that achieves the minimum value is

$$\hat{\theta} = \frac{\sum_k (V_{k2}^T D_k V_{k1} + V_{k2}^T V_{k3})}{\sum_k (V_{k2}^T D_k V_{k2})} \ . \tag{18}$$

The minimum value achieved is $f_1(\hat{\theta}) + f_2(\hat{\theta})$.

Similarly, the minimization of the following equation

$$\min_{\theta_k}\{2(V_{k1} - V_{k2}\theta_k)^T V_{k3} + (V_{k1} - V_{k2}\theta_k)^T D_k (V_{k1} - V_{k2}\theta_k)\} \tag{19}$$

occurs when θ_k is $\hat{\theta}_k$ that is the solution of the following equation $f_k'(\theta_k) = 0$ where

$$f_k(\theta_k) = \{(V_{k1} - V_{k2}\theta_k)^T V_{k3} + (V_{k1} - V_{k2}\theta_k)^T D_k (V_{k1} - V_{k2}\theta_k)\}$$

and

$$f_k'(\theta_k) = \frac{\partial}{\partial \theta_k}\{(V_{k1} - V_{k2}\theta_k)^T V_{k3} + (V_{k1} - V_{k2}\theta_k)^T D_k (V_{k1} - V_{k2}\theta_k)\} \ .$$

The value $\hat{\theta}_k$ that achieves the minimum value is

$$\hat{\theta}_k = \frac{V_{k2}^T D_k V_{k1} + V_{k2}^T V_{k3}}{V_{k2}^T D_k V_{k2}} \tag{20}$$

and the minimum value achieved is $f_k(\hat{\theta}_k)$.

Theorem 3 (F. Graybill (1976)) Suppose $Y \xrightarrow{\mathcal{D}} N(\mu, \Gamma)$, Γ is a positive definite matrix with rank n. Suppose M is a positive symmetric matrix. Then the quadratic form $Y^T M Y \xrightarrow{\mathcal{D}} \chi^2(df = \gamma, ncp = 1/2\mu^T M \mu)$ if and only if rank of MΓ is γ and MΓ is idempotent.

Lemma 5 Show that MΓ is idempotent.

Proof: The claim we need to show is that (MΓ)(MΓ)=MΓ.

In Theorem 1, M is the identity matrix so that MΓ=Γ and (MΓ)(MΓ)=Γ^2. Since each element of Γ^2 is the same as that of Γ, it proves that Γ is idempotent. That implies that MΓ is also idempotent as well with rank=trace(MΓ)=1. $\diamondsuit$

References

Andersen, P.K., Borgan, O., Gill, R. and Keiding, N. (1993), *Statistical Models Based on Counting Processes.* Springer, New York.

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

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90-120.

Owen, A. (2001). *Empirical Likelihood. Chapman and Hall.*

Zhou, M. (2000). Empirical envelope MLE and LR test. *Univ. of Kentucky, Dept. of Statist. Tech Report* .

Zhou, M. (2001). Censored empirical likelihood with many constraints. *Univ. of Kentucky, Dept. of Statist. Tech Report* .

Pan, X.R. and Zhou, M. (1999). Empirical likelihood in terms of cumulative hazard function for censored data. *J. Multivariate Analysis.*

Fang, H. and Zhou, M. (2000). On empirical likelihood ratio method with k-sample problems. *Univ. of Kentucky, Dept. of Statist. Tech Report* .

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *J. Amer. Statist. Assoc.*

Graybill, F. (1976), *Theory and Application of the Linear Model.* Wadsworth Publishing Company,Inc., Belmont, California.

Snedecor and Cochran. (1989), *Statistical Methods.* Iowa State University Press, Ames, Iowa.

Department of Statistics
University of Kentucky
Lexington, KY 40506-0027
kkim@ms.uky.edu