Symmetric Location Estimation/Testing by Empirical Likelihood

Kyoungmi Kim and Mai Zhou

Author Query

AQ1 Au: Please provide street name & fax. no. if any in CAF.

200026602_LSTA33_09_R1_072704

COMMUNICATIONS IN STATISTICS
Theory and Methods
Vol. 33, No. 9, pp. 1–11, 2004

Symmetric Location Estimation/Testing by Empirical Likelihood

Kyoungmi Kim and Mai Zhou*

Department of Statistics, University of Kentucky, Lexington, Kentucky, USA

ABSTRACT

The problem of estimating the center of a symmetric distribution is well studied and many nonparametric procedures are available. It often serves as the test problem for many nonparametric estimation procedures, and stimulated the development of efficient nonparametric estimation theory. We use this familiar setting to illustrate a novel use of empirical likelihood method for estimation and testing. Empirical likelihood is a general nonparametric inference method, see Owen [Owen, A. (2001). *Empirical Likelihood*. London: Chapman and Hall]. However, for symmetric location problem (and some other problems) empirical likelihood has difficulties. Owen (2001) call them ‘‘challenges for the empirical likelihood’’. We propose and study a way to use the empirical likelihood with such problems by modifying the parameter space. We illustrate this approach by

*Correspondence: Mai Zhou, Department of Statistics, University of Kentucky, Lexington, KY 40506-0027, USA; E-mail: mai@ms.uky.edu. AQ1

DOI: 10.1081/STA-200026602 0361-0926 (Print); 1532-415X (Online)

applying it to the symmetric location problem. We show that the usual asymptotic theory of empirical likelihood still holds and the asymptotic efficiency of the so obtained empirical NPMLE of location is studied.

Key Words: Many constraints of symmetry; Asymptotic chi-square distribution; Nonparametric information bound.

AMS 1991 Subject Classification: Primary 62G10; Secondary 62G05.

1. INTRODUCTION

Empirical likelihood, see Owen (2001), is a general nonparametric method that provides the advantages of a (parametric) likelihood ratio based inference without having to assume a parametric family of distributions. The advantages are more profound for censored/truncated data analysis where the traditional (Wald) approach becomes more complicated due to the difficulty in the estimation of variance. While the procedure discussed in this paper can clearly be used in the case of censored data, we shall focus on the non-censored data setting for clarity.

Definition (Owen, 2001). For given n i.i.d. observations $X_1, \ldots, X_n$ with common distribution function $F_X(t)$, the nonparametric or empirical likelihood for 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 $P_F(X = x_i)$. The empirical distribution function, $\widehat{F}_n$, maximizes $L(F)$ among all the distribution functions.

However, when we restrict the parameter space to be all the symmetric distributions,

$$\Theta = \{F \,|\, F \text{ is symmetric wrt } \theta, \text{ some } \theta \in R^1\},$$

the maximization of the above empirical likelihood has problems: the NPMLE does not exist or there are many NPMLE's having the same (empirical) likelihood value. When the true F is continuous, it is easy to see that $P(\widehat{F}_n \textit{ is symmetric}) = 0$ where $\widehat{F}_n$ is the empirical distribution function.

Example. Suppose $X_1 < X_2 < X_3$ are three ordered observations from a continuous symmetric distribution $F_0(t)$, with an unknown location of symmetry, θ. Without loss of generality, assume $X_3 - X_2 \neq X_2 - X_1$.

By adding one extra jump point to $\widehat{F}_n$, we can make $\widehat{F}_n$ into $F_n^*(t)$ which is symmetric. Unfortunately there are more than one candidate of $F_n^*(t)$'s that can have the same empirical likelihood value. If we believe that θ is located in the middle of X_2 and X_3, one more jump at the location X_4 on the right side of X_3 is needed to produce an $F_n^*(t)$ which is symmetric about θ with the empirical likelihood value $L(F) = w_1 \times w_2 \times w_3$, and $w_1 + w_2 + w_3 + w_4 = 1$. Otherwise, if we believe that θ is located in the middle of X_1 and X_2, then an extra jump at location X_0 to the left of X_1 is needed so that $F_n^*(t)$ is symmetric about θ with the empirical likelihood value $L(F) = w_1 \times w_2 \times w_3$ with $w_0 + w_1 + w_2 + w_3 = 1$.

It is clear the two different $F_n^*(t)$'s can achieve the same likelihood value. Indeed the local maximum is achieved for the first case at $w_2 = w_3 = 1/3$ and $w_1 = w_4 = 1/6$ and for the second case at $w_0 = w_3 = 1/6$ and $w_1 = w_2 = 1/3$.

Both $F_n^*(t)$ are NPMLE's having the same empirical likelihood value but with a very different θ and $F_n^*(t)$. This implies that NPMLE of θ and $F_n^*(t)$ is not unique. For larger samples there are even more candidates, F_n^*, that are symmetric and can achieve the same maximum empirical likelihood value.

There are many other setups that have the same difficulty, including two-sample location-shift problem. See Zhou (2001) for a two-sample location problem, and Kim (2003) for other cases and more discussions. Roughly speaking, when the knowledge/restriction on the distributions cannot be achieved by adjusting the jump sizes of the empirical distribution function, then there often is difficulty. (Here we have to add an extra jump to the empirical distribution to make it symmetric.)

One of the purpose of this paper is to illustrate the use of the "envelope empirical likelihood" method of Zhou (2001) to overcome this difficulty. The idea is to apply the empirical likelihood on a carefully constructed sequence of shrinking parameter spaces Θ_k, that converge to the $\Theta = \{all\ symmetric\ CDF\}$.

On each of the Θ_k, the NPMLE uniquely exists and the (regular) empirical likelihood theory works beautifully. This sequence of shrinking parameter spaces is called the envelope parameter space.

This approach is quite general. First, it can easily work with censored data. Second, this approach also works for other challenging situations like location-scale problems where the parameter space is not all the possible distributions but is still infinite dimensional and requires the CDF to have jump points other than the observed data points. Finally we point out that the method proposed can easily be generalized to handle higher dimensional data. We, however, will stick to the one sample symmetric location problem with uncensored data in this paper for the clarity of the presentation.

The semi-parametric problem of estimating symmetric location has been studied by many people, see many examples in the book by Bickel et al. (1993).

Our approach is closer to the Empirical Process Approach of Hsieh (1996). The advantage of the method proposed here is the simplicity of the procedure, we have a chi-square null distribution to set the P-value and there is no need to estimate the variance–covariance matrix when construct confidence interval/region. In the empirical process approach of Hsieh (and many other adaptive estimation procedures) you need to first estimate a variance–covariance matrix of the empirical process involved and then use the estimated matrix to do a weighted least squares to produce the estimator. For doubly censored data, the variance–covariance can be difficult to estimate. The advantages of our procedure are of course inherited from the (empirical) likelihood ratio method.

2. ENVELOPE EMPIRICAL LIKELIHOOD FOR SYMMETRIC DISTRIBUTIONS

Suppose $X_1, \ldots, X_n$ are i.i.d. observations from a symmetric distribution F with an arbitrary location parameter θ (i.e., the center of symmetry of F is θ).

Maximizing the log empirical likelihood,

$$\log L = \sum_{i=1}^{n} \log w_i = \sum_{i=1}^{n} \log \Delta F(X_i), \tag{2}$$

over all symmetric distributions is not well defined as seen in the example of section one.

We enlarge the parameter space to Θ_1. It will be shown later that the NPMLE is well defined on this space. The enlarged parameter space is defined as

$$\Theta_1 = \{F : \text{all distributions satisfy (4)}\}. \tag{3}$$

For given t_i, $i = 1, 2, \ldots, k$

$$\text{for some } \theta, \quad F(\theta - t_i) = 1 - F(\theta + t_i). \tag{4}$$

We may rewrite the above as integrations:

$$\text{for some } \theta, \quad \int_{-\infty}^{\theta - t_i} dF(t) = \int_{\theta + t_i}^{\infty} dF(t), \quad i = 1, 2, \ldots, k. \tag{5}$$

If we take the functions

$$g_i(\theta - t) = I_{[0 \le \theta - t - t_i]} = I_{[t \le \theta - t_i]} \quad \text{and} \quad g_i^*(\theta - t) = I_{[0 \ge \theta - t + t_i]} = I_{[t \ge \theta + t_i]}$$

in the above, the integration equations can take the form of

$$\int_{-\infty}^{\infty} g_i(\theta - t) dF(t) = \int_{-\infty}^{\infty} g_i^*(\theta - t) dF(t), \quad i = 1, 2, \ldots, k. \tag{6}$$

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

It turns out that maximizing the log empirical likelihood $\log L$ defined above among distributions in the parameter space Θ_1 is well defined and thus yields both the (envelope) empirical NPMLE, $\hat{\theta}$ and $\widehat{F}(t)$. The estimate $\widehat{F}(t)$ is symmetric at least on k points: t_i. We shall only focus on the study of the estimator $\hat{\theta}$ in this paper.

We note that the newly defined parameter space actually is dependent on the choice and number of the functions g_i and g_i^*, or the t_i points. When the points t_i in (4) becomes dense then the space Θ_1 becomes the space of all symmetric distributions. The many choices of the space Θ_1 is similar to the choices of bandwidth in the histogram estimation of a density function. However, as sample size grows, we do not have to adaptively chose the t_i points like choosing the bandwidth in density estimation. If we use a fixed choice of Θ_1, (not changing with sample size) we still obtain a root n consistent estimator of θ and a Wilks theorem in likelihood ratio test. If we do adaptively change the space Θ_1, we may improve efficiency. In the following we only work with a fixed Θ_1.

3. ENVELOPE EMPIRICAL LIKELIHOOD RATIO TEST

Suppose $F(\cdot)$ is symmetric about θ. Consider testing the hypothesis:

$$H_0 : \theta = \theta_0; \quad \text{vs.} \quad H_A : \theta \neq \theta_0.$$

The test statistic we propose is the likelihood ratio statistics

$$T = -2\left\{ \max_{\Theta_1 \text{with } \theta = \theta_0} \log L - \max_{\Theta_1 \text{with } \theta \in R^1} \log L \right\}. \tag{7}$$

We show below that this empirical likelihood ratio test statistics will have an approximate chi-square distribution with one degree of freedom under the null hypothesis. We reject H_0 for larger values of T. Confidence intervals for θ can be obtained by inverting the chi-square test.

The θ value that achieve the maximum in the second term of (7) will be our (envelope) NPMLE of the location, $\hat{\theta}$.

Denote the column vectors

$$\begin{aligned} g(\theta - t) &= \{g_1(\theta - t), \ldots, g_k(\theta - t)\}^T, \\ g^*(\theta - t) &= \{g_1^*(\theta - t), \ldots, g_k^*(\theta - t)\}^T, \quad \text{and} \\ \lambda &= \{\lambda_1, \ldots, \lambda_k\}^T. \end{aligned}$$

Lemma 1. *Suppose $X_1, \ldots, X_n$ are n i.i.d. observations from a symmetric distribution F with an arbitrary location parameter θ_0.*

Then for any fixed θ, the probability w_i that maximizes the log likelihood function (2) *satisfying the constraints of* (6) *is given by*

$$w_i(\lambda, \theta) = \frac{1}{n - n\lambda^T \cdot (g(\theta - x_i) - g^*(\theta - x_i))}, \tag{8}$$

where $\lambda^T \cdot g(\theta - x_i)$ denotes the inner product $\sum_{j=1}^{k} \lambda_j g_j(\theta - x_i)$. The λ value in (8) *is obtained as the solution of the following k equations, respectively,*

$$h_r(\lambda, \theta) = \sum_{i=1}^{n} \frac{[g_r(\theta - x_i) - g_r^*(\theta - x_i)]}{n - n\lambda^T \cdot (g(\theta - x_i) - g^*(\theta - x_i))} = 0, \quad r = 1, \ldots, k. \tag{9}$$

Clearly the so determined λ value depends on the θ given, so in the subsequent discussions we shall write λ as $\lambda(\theta)$.

Proof. Standard Lagrange multiplier calculation similar to Owen (2001). □

4. LARGE SAMPLE RESULTS

Lemma 2. *Under mild regularity conditions, the solution λ of the constraint equation in* (9) *under the null hypothesis has the following* *asymptotic representations*:

(i) *Let* θ_0 *be the true parameter, and assume*

$$h'(0,\theta_0) = \left[\frac{\partial h_r(\lambda,\theta_0)}{\partial\lambda_s}\Big|_{\lambda=0}\right]$$

is an invertible $k \times k$ *matrix, then we have*

$$\sqrt{n}\lambda(\theta_0) \xrightarrow{\mathscr{D}} N(0,\Sigma); \quad \textit{as } n \to \infty$$

where the variance–covariance matrix is

$$\Sigma = \lim_{n\to\infty}[h'(0,\theta_0)]^{-1}.$$

(ii) *In addition, assume that* $g(\cdot)$ *and* $g^*(\cdot)$ *are smooth and* $|\theta - \theta_0| = O(1/\sqrt{n})$, *we have*

$$\lambda(\theta) = \lambda(\theta_0) - h'(0,\theta_0)^{-1}G(\theta-\theta_0) + o_p(|\theta-\theta_0|)$$

where G *is a* $k \times 1$ *matrix with its column defined as*

$$G = \left\{\sum_{i=1}^{n}\frac{g_1'(\theta_0-x_i) - g_1'^*(\theta_0-x_i)}{n}, \ldots, \sum_{i=1}^{n}\frac{g_k'(\theta_0-x_i) - g_k'^*(\theta_0-x_i)}{n}\right\}^T.$$

Proof. Use Taylor expansion on h with respect to λ in Eq. (9). □

Remark 1. The $k \times k$ matrix

$$h'(0,\theta_0) = \left[\frac{\partial h_r}{\partial\lambda_s}\right]_{\lambda=0} = \left[\sum_{i=1}^{n}\frac{[g_r - g_r^*][g_s - g_s^*]}{n}\right]$$

is easy to verify to be symmetric and at least non-negative definite. Proper choice of the g, g^* function will guarantee it to be positive definite.

Remark 2. As $n \to \infty$, the limit of G is easily seen to be

$$\left\{E[g_1'(\theta_0 - X) - g_1'^*(\theta_0 - X)], \ldots, E[g_k'(\theta_0 - X) - g_k'^*(\theta_0 - X)]\right\}.$$

Notice that we have assumed the smoothness of the g function in the above. However, even if g is an indicator function as in (6), we may use Dirac's delta function to obtain that

$$E[g_1'(\theta_0 - X)] = \int g_1'(\theta_0 - x)f(x)dx = f(\theta_0 - t_1)$$

and

$$E[g_1'^{*}(\theta_0 - X)] = \int g_1'^{*}(\theta_0 - x)f(x)dx = -f(\theta_0 + t_1)$$

and the limit of G in this case will be

$$\{[f(\theta_0 - t_1) + f(\theta_0 + t_1)], \ldots, [f(\theta_0 - t_k) + f(\theta_0 + t_k)]\}.$$

Now the main theorems.

Theorem 1. *Under the same conditions as in Lemma 2, the empirical likelihood ratio statistic T defined in (7) has asymptotically a chi-square distribution with one degree of freedom:*

$$T \xrightarrow{\mathscr{D}} \chi^2_{(1)}, \quad \textit{as } n \to \infty.$$

Proof. See Appendix.

Theorem 2. *Let $\hat{\theta}$ be the location parameter that maximizes the second term in (7).*

The asymptotic distribution of the (envelope) maximum empirical likelihood estimator, $\hat{\theta}$, is given by

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathscr{D}} N(0, \sigma^2),$$

where

$$\sigma^2 \lim_{n\to\infty} \{G^T[h'(0, \theta_0)]^{-1} G\}^{-1}.$$

Proof. See Appendix.

5. EFFICIENCY

We now take a closer look at the asymptotic variance of the (envelope) NPMLE obtained in Theorem 2 above. By noticing the special structure of the matrix, $h'(0, \theta_0)$, we can show (see Kim, 2003 for details), that the quadratic form appeared in the variance above, $G^T[h'(0, \theta_0]^{-1}G$, can be explicitly computed as

$$G^T[h'(0, \theta_0)]^{-1}G = \sum_r \frac{(G_r - G_{r-1})^2}{h_r - h_{r-1}}$$

where G_r is the rth component of the vector G, and h_r is the rth diagonal element of the matrix $h'(0, \theta_0)$. ($G_0 = 0$ and $h_0 = 0$ by convention.)

In view of Remark 2, the numerators in each term of the above summation has limit as $n \to \infty$. Similar calculation is available for the denominator. Putting them together we see that the above summation have a limit

$$\sum_{r=1}^{k} \frac{([f(\theta_0 - t_r) + f(\theta_0 + t_r)] - [f(\theta_0 - t_{r-1}) + f(\theta_0 + t_{r-1})])^2}{[F(\theta_0 - t_r) + 1 - F(\theta_0 + t_r)] - [F(\theta_0 - t_{r-1}) + 1 - F(\theta_0 + t_{r-1})]}. \tag{10}$$

It is worth pointing out that (under mild regularity conditions on f) (10) is a finite sum approximation (from below) to the integral

$$\int_{-\infty}^{\theta_0} 2\frac{[f'(t)]^2}{f(t)}\,dt = \int_{-\infty}^{\infty} \frac{[f'(t)]^2}{f(t)}\,dt, \tag{11}$$

which is the information bound of the semiparametric model for estimating θ, see Bickel et al. (1993).

This shows that the (envelope) NPMLE obtained in Theorem 2 on the space Θ_1 is close to be fully asymptotically efficient, because the asymptotic variance obtained in Theorem 2 equals to the inverse of the finite sum, (10), which in turn is approximately the semiparametric information bound, (11).

APPENDIX

Proof of Theorem 1. Define a function of θ by

$$f(\lambda(\theta), \theta) = \sum_{i=1}^{n} \log w_i(\lambda(\theta), \theta).$$

The log empirical likelihood ratio statistic, T, is then

$$T = -2\{f(\lambda(\theta_0), \theta_0)\} + 2\min_{\theta}\{f(\lambda(\theta), \theta)\}.$$

By the Taylor expansion on the second term above, we have

$$\begin{aligned} T &= -2\{f(\lambda(\theta_0), \theta_0)\} \\ &\quad + 2\min_{\theta}\left\{f(\lambda(\theta_0), \theta_0) + (\theta - \theta_0)\frac{\partial f}{\partial \theta} + \frac{1}{2}(\theta - \theta_0)^2\frac{\partial^2 f}{\partial \theta^2} + o_p(1)\right\} \qquad (12) \\ &= \min_{\theta}\{2(\theta - \theta_0)A + (\theta - \theta_0)^2 B + o_p(1)\} \quad \text{(say)}. \qquad (13) \end{aligned}$$

Aside from the small term, we immediately get that the minimum value achieved by T is $-A^2/B$, and the $\hat{\theta}$ that achieves this minimum satisfy $\hat{\theta} - \theta = -A/B$.

The rest of the proof is just calculating the derivatives and checking that $-A^2/B$ has the correct asymptotic distribution. Notice that the derivative $\partial f/\partial \lambda$ at $(\lambda = \lambda(\theta_0), \theta = \theta_0)$ is zero. And from Lemma 2(ii), we have the derivative of $\lambda(\theta)$:

$$\lambda'(\theta)|_{\theta=\theta_0} = -[h'(0, \theta_0)]^{-1} G.$$

After tedious but straight forward calculations, we have

$$\begin{aligned} T &= -\frac{A^2}{B} + o_p(1) \\ &= -\frac{(\sqrt{n}\lambda(\theta_0)G)^2}{B/n} + o_p(1). \end{aligned}$$

On the other hand, we can show that as $n \to \infty$

$$\mathrm{P}\lim B/n = \mathrm{P}\lim G^T [h']^{-1} G$$

where P lim denote the limit in probability.

Applying the asymptotic distribution of $\lambda(\theta_0)$ shown in Lemma 2(i):

$$\sqrt{n}\lambda(\theta_0) \xrightarrow{\mathscr{D}} N(0, \Sigma); \quad \text{as } n \to \infty,$$

and the Slutsky theorem, the log empirical likelihood ratio statistic T is seen to have the limit of chi-square with one degree of freedom. □

Proof of Theorem 2. From the proof of Theorem 1 we already noticed that (aside from a negligible term) the θ value that achieves the minimum is

$$\hat{\theta} - \theta_0 = -\frac{A}{B} + o_p(1) = -\frac{n\lambda(\theta_0)G}{B} + o_p(1)$$

and, therefore, by Lemma 2

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\frac{\sqrt{n}\lambda(\theta_0)G}{B/n} + o_p(1)$$
$$\xrightarrow{\mathscr{D}} N(0, (G^T h'(0, \theta_0)^{-1} G)^{-1}).$$

REFERENCES

Bickel, P., Klaassen, C., Ritov, Y., Wellner, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: The Johns Hopkins University Press.

Hsieh, F. (1996). Empirical process approach in two-sample location-scale model with censored data. *Ann. Statist.* 24:2705–2719.

Kim, K. (2003). Empirical Likelihood Ratio Method When Additional Information is Known. Ph.D. dissertation, Department of Statistics, University of Kentucky.

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

Zhou, M. (2001). Empirical Envelope MLE and LR Test. Technical Report, University of Kentucky, Department of Statistics.