Empirical likelihood for non-degenerate U-statistics

Bing-Yi Jing[1], Junqing Yuan[1], and Wang Zhou[2]

[1] Department of Mathematics
Hong Kong Univ. of Sci. and Tech.
Clear Water Bay, Kowloon
Hong Kong
(e-mail: majing@ust.hk, yuanjq@ust.hk)

[2] Dept. of Stat. and Applied Prob.
National University of Singapore
Singapore 117543
(e-mail: stazw@nus.edu.sg)

Abstract. Standard empirical likelihood for U-statistics is too computationally expensive. To overcome this computational difficulty, we reformulate the empirical likelihood for non-degenerate U-statistics in terms of "pseudo" mean in this paper, and show that the empirical log-likelihood ratio has an asymptotic chi-squared distribution under second moment condition. The method is extremely simple to use, and yet provide better coverage accuracy in general than other alternative methods from our simulation studies.
Keywords: U-statistics, empirical likelihood, confidence interval.

1 Introduction

The empirical likelihood method was first introduced by [Owen, 1988] for constructing confidence intervals and [Owen, 1990] for confidence regions. [Hall and LaScala, 1990] has summarized its advantages over the bootstrap: the empirical likelihood regions are shaped "automatically" by the sample, Bartlett correctable, range preserving and transformation respecting. For these reasons, the empirical likelihood has found lots of applications such as in smooth functions of means [DiCiccio *et al.*, 1989], in nonparametric density [Chen, 1996], in regression function estimation [Owen, 1991] [Chen and Qin, 2000] and so on. For a more thorough review of the empirical likelihood method and its applications, the reader is referred to the recent monograph by [Owen, 2001].

In this paper, we are interested in applying the empirical likelihood method to U-statistics. Let $X, X_1, \cdots, X_n, n \geq 2$, be independent and identically distributed (i.i.d) random variables with common distribution function $F(x)$. A U-statistic of degree $m \geq 2$ with a symmetric kernel h is defined to be

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \ldots < i_m \leq n} h(X_{i_1}, ..., X_{i_m}) \tag{1}$$

where $\theta = Eh(X_1, ..., X_m)$ is a parameter of interest. Under very weak conditions, U_n is a Minimum Variance Unbiased Estimator of θ. On the other hand, U-statistics have many applications in hypothesis testing. For further details on U-statistic see [Lee, 1990]. Define

$$g(x) = Eh(x, X_1, ..., X_{m-1}) - \theta, \qquad \sigma_g^2 = var(g(X)). \tag{2}$$

Throughout this paper, we shall assume that $\sigma_g^2 > 0$.

The straightforward application of Owen's empirical likelihood in this context can be described as follows. Denote F_q to be the empirical distribution function which assigns probability q_i to observation X_i. Then, the empirical likelihood, evaluated at the true parameter value θ, can be defined by

$$\widetilde{L}(\theta) = \max_{\widetilde{\theta}(F_q)=\theta, \sum q_i = 1} \prod_{i=1}^{n} q_i, \tag{3}$$

where

$$\widetilde{\theta}(F_q) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < ..., < i_m \le n} n^m q_{i_1} ... q_{i_m} h(X_{i_1}, ..., X_{i_m}).$$

Note that $\prod_{i=1}^n q_i$, subject to $\sum_{i=1}^n q_i = 1$, attains its maximum n^{-n} at $q_i = n^{-1}$. Then, the empirical likelihood ratio at θ is given by

$$\widetilde{R}(\theta) = \widetilde{L}(\theta)/n^{-n} = \max_{\widetilde{\theta}(F_q)=\theta, \sum q_i = 1} \prod_{i=1}^{n} (nq_i). \tag{4}$$

As mentioned in [Wood *et al.*, 1996], Wilks's theorem holds under mild conditions in this case, i.e., $-2 \log \widetilde{R}(\theta) \xrightarrow{d} \chi_1^2$, where $\xrightarrow{d}$ means converges in distribution as $n \to \infty$, and χ_1^2 denotes the chi square distribution with one degree of freedom. This can be used to construct confidence intervals for the parameter θ. We shall refer to this procedure as Owen's *direct or "exact" empirical likelihood method* to U-statistics.

The major drawback of Owen's direct empirical likelihood method is its computational difficulty due to the presence of nonlinear constraints in the underlying optimization problem. [Wood *et al.*, 1996] proposed a so-called *sequential linerization method* for empirical likelihood methods with nonlinear constraints, and applied it to U-statistics. They found in their simulation studies that a single iteration of the linearization procedure may not be enough to achieve reliable coverage probabilities, and suggested to employ multiple (three to ten) iterations of the linearization procedure or bootstrap calibration in practice in order to improve coverage probabilities.

In this paper, we propose a new empirical likelihood method to U-statistics. The key idea of our method is to turn the U-statistic into a "sample mean" based on some "pseudo" observations, and then simply apply Owen's empirical likelihood to that "sample mean". As will be seen from the next section, those "pseudo" observations are in fact dependent random variables. Wilks's theorem will be established under mild conditions, which can then be used to construct confidence intervals for the parameter θ. The most attractive feature of our approach is its simplicity. Furthermore, our simulations results show that the coverage probabilities of our approach are in general better than alternative methods.

The paper is organized as follows. In Section 2, we introduce a new empirical likelihood method for U-statistics, and presents some theoretical results. Some simulation studies are conducted in Section 3 to compare the performances of the empirical likelihood and other methods. Proofs are deferred to Section 4.

2 Methodology and main results

First we rewrite U_n as

$$U_n = \frac{1}{n}\sum_{i=1}^{n} V_i,$$

where the "components" of U_n, defined by [Sen, 1960]

$$V_i = \binom{n-1}{m-1}^{-1} \sum_{\substack{1\le j_1<\ldots<j_{m-1}\le n \\ j_r\neq i, 1\le r\le m-1}} h(X_i, X_{j_1}, ..., X_{j_{m-1}}) \tag{5}$$

are treated as "pseudo" observations. Note that V_i's are dependent.

To employ empirical likelihood, let $p = (p_1, \cdots, p_n)$ be a probability vector, i.e., $\sum_{i=1}^n p_i = 1$ and $p_i \ge 0$ for $1 \le i \le n$. Let G_p be the distribution function which assigns probability p_i at the ith pseudo observation V_i, and hence $\theta(G_p) = \sum_{i=1}^n p_i V_i$. Then, the empirical likelihood ratio, evaluated at θ, is given by

$$L(\theta) = \max_{\theta(G_p)=\theta, \sum p_i=1} \prod_{i=1}^{n} p_i. \tag{6}$$

Note that $\prod_{i=1}^n p_i$, subject to $\sum_{i=1}^n p_i = 1$, attains its maximum n^{-n} at $p_i = n^{-1}$. So we define the empirical likelihood ratio at θ by

$$R(\theta) = L(\theta)/n^{-n} = \max_{\theta(G_p)=\theta, \sum p_i=1} \prod_{i=1}^{n}(np_i). \tag{7}$$

Using Lagrange multipliers, we have

$$p_i = \frac{1}{n}\frac{1}{1+\lambda(V_i-\theta)}, \tag{8}$$

where λ satisfies

$$g(\lambda) := \frac{1}{n} \sum_{i=1}^{n} \frac{V_i - \theta}{1 + \lambda(V_i - \theta)} = 0. \tag{9}$$

After plugging the p_i's back into (7) and taking the logarithm of $R(\theta)$, we get the nonparametric log-likelihood ratio

$$\log R(\theta) = -\sum_{i=1}^{n} \log\{1 + \lambda(V_i - \theta)\}.$$

The next theorem shows that Wilks's theorem holds here under a mild condition.

Theorem 1 *Assume that* $Eh^2(X_1, ..., X_m) < \infty$ *and* $\sigma_g^2 > 0$*, then*

$$-\frac{2}{m^2} \log R(\theta) \xrightarrow{d} \chi_1^2.$$

The proof of Theorem 1 will be given in Section 4.

Remark 1 *Wilks's theorem, stated in Theorem 1, is slightly different from the ones we normally encounter. For example, for the Owen's direct empirical likelihood method, one has*

$$-2 \log \widetilde{R}(\theta) \xrightarrow{d} \chi_1^2.$$

However, in our case here, we have

$$-\frac{2}{m^2} \log R(\theta) \xrightarrow{d} \chi_1^2.$$

Remark 2 *An approximate* $1-\alpha$ *level confidence interval for* θ *can be defined as*

$$\Re_c = \{\theta : -\frac{2}{m^2} \log R(\theta) \le c\},$$

where c *is chosen to satisfy* $P(\chi_1^2 \ge c) = \alpha$*. From Theorem 1, we have*

$$\lim_{n\to\infty} P\{\theta \in \Re_c\} = P(\chi_1^2 \le c) = 1 - \alpha.$$

In other words, the interval $\Re_c$ *gives asymptotic correct coverage probability.*

3 Simulation results

In this section, we shall conduct some simulation studies to investigate the coverage accuracy of the empirical likelihood method proposed in this paper. Comparisons will be made with some alternative methods such as the normal approximation method, Owen's direct or "exact" empirical likelihood method, and the sequential linerization method proposed by [Wood *et al.*, 1996]. Three examples will be used for illustration: population variance, probability weighted moments, and Gini's mean difference as special cases of U-statistics. All the simulation results are based on 1,000 repetitions.

Example 1: population variance $\sigma^2 = var(X)$. In this case, the sample variance is a U-statistic with the kernel $h(x, y) = (x-y)^2/2$. For this example, it is also rather easy to apply Owen's empirical likelihood method directly by placing probability weight p_i on X_i and maximizing the empirical likelihood subject to

$$\sum_{i=1}^{n} p_i (X_i - \mu_X)^2 = \sigma^2 \qquad \text{with } \mu_X = \sum_{i=1}^{n} p_i X_i.$$

Therefore, it would be interesting to compare this direct approach with the one proposed in this paper. For illustrative purposes, we shall include the normal approximation method as well for comparison. The underlying population is selected as standard Normal, then the actual value $\theta = 1$. The results are summarized in Table 1.

Table 1. Coverage accuracy for the variance

	nominal level	0.80	0.90	0.95
n=15	Normal Appr.	0.655	0.751	0.816
	Owen's EL	0.668	0.782	0.847
	Our EL	0.708	0.806	0.868
n=40	Normal Appr.	0.723	0.828	0.878
	Owen's EL	0.758	0.855	0.918
	Our EL	0.748	0.845	0.898
n=100	Normal Appr.	0.772	0.872	0.917
	Owen's EL	0.804	0.906	0.949
	Our EL	0.789	0.884	0.931

Example 2: probability weighted moment $E[XF(X)]$. In this case, the sample probability weighted moment is a U-statistic with the kernel $h(x, y) = \max\{x, y\}/2$. Coverage probabilities of the "exact" empirical likelihood method, described in the Introduction, were given in table 4 of [Wood *et al.*, 1996], which will be used for comparison with our own approach in this paper. Two underlying distributions are considered: the standard Normal and the exponential with mean 1. For these distributions, the population values are 0.282 and 0.75 respectively. Table 2 records the simulation results, with those in parentheses for the latter distribution.

Example 3: Gini's mean difference $E|X_1 - X_2|$. Gini's mean difference is an attractive measure for describing the population concentration. Its sample version is a U-statistics with the kernel $h(x, y) = |x - y|$. This

Table 2. Coverage accuracy for the probability weighted moment

	nominal level	0.80	0.90	0.95
n=15	"Exact" EL	0.745 (0.705)	0.844 (0.801)	0.896 (0.882)
	Our EL	0.746 (0.740)	0.845 (0.830)	0.912 (0.888)
n=40	"Exact" EL	0.742 (0.741)	0.849 (0.844)	0.922 (0.899)
	Our EL	0.768 (0.761)	0.866 (0.857)	0.923 (0.910)
n=100	"Exact" EL	0.787 (0.783)	0.895 (0.864)	0.944 (0.929)
	Our EL	0.821 (0.771)	0.904 (0.873)	0.941 (0.924)

example was also studied by [Wood *et al.*, 1996], who used their sequential linearization approach in this case. The comparisons with our method is presented in Table 3, where Wood *et al.*(r) denotes the sequential linearization approach with r iterations. For the underlying distribution, we use a standard Normal, so $\theta = 1.1284$.

Table 3. Coverage accuracy for Gini's mean difference

	nominal level	0.80	0.90	0.95
n=15	Wood *et al.*(1)	0.693	0.799	0.859
	Wood *et al.*(3)	0.737	0.864	0.932
	Our EL	0.741	0.846	0.889
n=40	Wood *et al.*(1)	0.756	0.862	0.919
	Wood *et al.*(3)	0.751	0.862	0.924
	Our EL	0.772	0.864	0.917
n=100	Wood *et al.*(1)	0.782	0.884	0.935
	Wood *et al.*(3)	0.780	0.887	0.939
	Our EL	0.787	0.889	0.936

The following observations can be made from our simulation studies:

(1) As expected, all methods improve as the sample size n increases.
(2) From Table 1, we see that, our method outperforms Normal Approximation method. Comparing with Owen's empirical likelihood method, our's looks better for small sample size.
(3) From Table 2, our method seems to perform slightly better than the "exact" empirical likelihood, mentioned in the Introduction. But our method is much simpler to use.
(4) From Table 3, we see that, overall, our method performs equally well as Wood *et al*'s sequential linearization approach with 3 iterations, and both are better than Wood *et al*'s approach with only 1 iteration. However, our method is the simplest amongst the three.

In summary, our empirical likelihood method for U-statistics in general performs better or as well as all other alternative methods such as normal approximation, exact empirical likelihood and sequential linearization procedure. Furthermore, our approach is the simplest one to use. For these reasons, our method should always be preferred.

4 Proof of main results

For notational simplicity, we shall prove our main results for U-statistics of order $m = 2$ only. The case for the general order $m \geq 2$ can be done similarly. But first we shall list several simple lemmas for easy reference later in the section.

Lemma 1 *[Hoeffding, 1948] Suppose* $Eh^2(X_1, X_2) < \infty$, *then*

$$\frac{\sqrt{n}(U_n - \theta)}{2\sigma_g} \xrightarrow{d} N(0,1).$$

Corollary 1 *Assuming* $Eh^2(X_1, X_2) < \infty$, *then* $U_n - \theta = O_p(n^{-1/2})$.

Proof. This is a direct consequence of Lemma 1.

Lemma 2 *Let* $S = n^{-1}\sum_{i=1}^{n}(V_i - \theta)^2$, *if* $Eh^2(X_1, X_2) < \infty$, *then*

$$S = \sigma_g^2 + o(1) \qquad a.s.$$

Proof. Note that

$$S = \frac{1}{n}\sum_{i=1}^{n}(V_i - \theta)^2 = \frac{1}{n}\sum_{i=1}^{n}(V_i - U_n)^2 + (U_n - \theta)^2.$$

Let $\sigma^2 = var\{h(X_1, X_2)\} < \infty$, since $Eh^2(X_1, X_2) < \infty$, thus

$$var(U_n) = \frac{4(n-2)}{n(n-1)}\sigma_g^2 + \frac{2}{n(n-1)}\sigma^2.$$

Denote the jackknife estimate of $var(U_n)$ by $\widehat{var}(JACK)$, Lee (1990) identified that (page 223-4)

$$\frac{1}{n}\sum_{i=1}^{n}(V_i - U_n)^2 = \frac{(n-2)^2}{4(n-1)}\widehat{var}(JACK).$$

Since $\widehat{var}(JACK)$ is a consistent estimator of $var(U_n)$ in the sense that

$$n\{\widehat{var}(JACK) - var(U_n)\} \to 0, \qquad a.s.$$

then as $n \to \infty$, we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(V_i - U_n)^2 &= \frac{(n-2)^2}{4(n-1)}\left(var(U_n) + o(n^{-1})\right)\\ &= \frac{(n-2)^2}{4(n-1)}\left(\frac{4(n-2)}{n(n-1)}\sigma_g^2 + \frac{2}{n(n-1)}\sigma^2 + o(n^{-1})\right)\\ &= \sigma_g^2 + o(1), \qquad a.s.\end{aligned}$$

In addition, the strong law of large number for U-statistics results in $U_n = \theta + o(1)$ $a.s.$ Therefore, $S = \sigma_g^2 + o(1)$ $a.s.$, which ends the proof.

Lemma 3 *Let* $Y_n = \max_{1 \le i \ne j \le n} |h(X_i, X_j)|$, *if* $Eh^2(X_1, X_2) < \infty$, *then*

$$Y_n = o(n^{1/2}) \qquad a.s.$$

Proof. Since $Eh^2(X_1, X_2) < \infty$, we have

$$\sum_{n=1}^{\infty} P\left(h^2(X_1, X_2) > n\right) < \infty,$$

which implies that

$$\sum_{n=1}^{\infty} P\left(h^2(X_i, X_j) > n\right) < \infty, \qquad \text{for any } 1 \le i \ne j \le n.$$

And hence by the Borel-Cantelli Lemma, with probability 1,

$$|h(X_i, X_j)| > n^{1/2}, \qquad \text{for any } 1 \le i \ne j \le n$$

finitely often. Thus with probability 1, $Y_n > n^{1/2}$ occurs finitely often. By the same argument $Y_n > An^{1/2}$ finitely often with probability 1 for any $A > 0$. Consequently,

$$\limsup_{n\to\infty} \frac{Y_n}{n^{1/2}} \le A \qquad a.s. \tag{10}$$

Inequality (10) holds simultaneously with probability 1 for any countable set of values for A. Therefore $Y_n = o(n^{1/2})$ $a.s.$

Corollary 2 *Let* $Z_n = \max_{1 \le i \le n} |V_i - \theta|$, *if* $Eh^2(X_1, X_2) < \infty$, *then*

$$Z_n = o(n^{1/2}) \qquad a.s., \tag{11}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} |V_i - \theta|^3 = o(n^{1/2}) \qquad a.s. \tag{12}$$

Proof. Note that

$$|V_i - \theta| \le \frac{1}{n-1}\sum_{\substack{j=1\\ j\ne i}}^{n} |h(X_i, X_j)| + |\theta| \le Y_n + |\theta|$$

for any $1 \le i \le n$. By Lemma 3, $Z_n = o(n^{1/2})$ *a.s.*

For the second assertion, by (11) and Lemma 2, with probability 1

$$\frac{1}{n}\sum_{i=1}^{n} |V_i - \theta|^3 \le Z_n \times \frac{1}{n}\sum_{i=1}^{n}(V_i - \theta)^2 = o(n^{1/2})$$

as has to be shown.

Proof of Theorem 1. We first show that the root of (9) satisfies $|\lambda| = O_p(n^{-1/2})$. Note that

$$\begin{aligned} 0 = |g(\lambda)| &= \frac{1}{n}\left|\sum_{i=1}^{n}(V_i - \theta) - \lambda\sum_{i=1}^{n}\frac{(V_i-\theta)^2}{1+\lambda(V_i-\theta)}\right| \\ &\ge \frac{|\lambda|}{n}\sum_{i=1}^{n}\frac{(V_i-\theta)^2}{1+\lambda(V_i-\theta)} - \frac{1}{n}\left|\sum_{i=1}^{n}(V_i-\theta)\right| \\ &\ge \frac{|\lambda| S}{1+|\lambda| Z_n} - \left|\frac{1}{n}\sum_{i=1}^{n}(V_i-\theta)\right|. \end{aligned}$$

By Corollary 1, the second term is $O_p(n^{-1/2})$. Recalling Lemma 2, $S = \sigma_g^2 + o(1)$ *a.s.*, it follows that $\frac{|\lambda|}{1+|\lambda| Z_n} = O_p(n^{-1/2})$, and hence by (11),

$$|\lambda| = O_p(n^{-1/2}). \tag{13}$$

For convenience, let $\gamma_i = \lambda(V_i - \theta)$ where λ is the root of (9). Then by (11) and (13),

$$\max_{1\le i\le n} |\gamma_i| = O_p(n^{-1/2}) o(n^{1/2}) = o_p(1). \tag{14}$$

Expanding (9),

$$\begin{aligned} 0 = g(\lambda) &= \frac{1}{n}\sum_{i=1}^{n}(V_i-\theta)\left(1 - \gamma_i + \gamma_i^2/(1+\gamma_i)\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} V_i - \theta - S\lambda + \frac{1}{n}\sum_{i=1}^{n}(V_i-\theta)\gamma_i^2/(1+\gamma_i), \end{aligned} \tag{15}$$

The final term in (15) is bounded by

$$\frac{1}{n}\sum_{i=1}^{n}|V_i-\theta|^3\lambda^2|1+\gamma_i|^{-1} = o(n^{1/2})O_p(n^{-1})O_p(1) = o_p(n^{-1/2}) \tag{16}$$

using (12), (13) and (14). Therefore, we may write

$$\lambda = S^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} V_i - \theta\right) + \beta = S^{-1}(U_n - \theta) + \beta, \tag{17}$$

where $|\beta| = o_p(n^{-1/2})$. By Taylor's expansion,

$$\begin{aligned}
-\frac{1}{2}\log R(\theta) &= \frac{1}{2}\sum_{i=1}^{n}\gamma_i - \frac{1}{4}\sum_{i=1}^{n}\gamma_i^2 + \frac{1}{2}\sum_{i=1}^{n}\eta_i \\
&= \frac{1}{2}n\lambda(U_n - \theta) - \frac{1}{4}nS\lambda^2 + \frac{1}{2}\sum_{i=1}^{n}\eta_i \\
&= \frac{n(U_n - \theta)^2}{4S} - \frac{1}{4}nS\beta^2 + \frac{1}{2}\sum_{i=1}^{n}\eta_i,
\end{aligned}$$

where $\eta_i = O(|\gamma_i|^3)$ *a.s.*. The first term has an asymptotic distribution χ_1^2 by Lemma 1 and 2. By Lemma 2 and (17), the second term is bounded by

$$\left|-\frac{1}{4}nS\beta^2\right| = n(\sigma_g^2 + o(1))o_p(n^{-1}) = o_p(1).$$

From (12) and (13), the final term is bounded by $o_p(1)$. Therefore applying Slutsky theorem completes the proof.

References

[Chen and Qin, 2000]S. X. Chen and Y. S. Qin. Empirical likelihood confidence intervals for local linear smoothers. *Biometrika*, pages 946–953, 2000.

[Chen, 1996]S. X. Chen. Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika*, pages 329–341, 1996.

[DiCiccio *et al.*, 1989]T. S. DiCiccio, P. Hall, and J. Romono. Comparison of parametric and empirical likelihood functions. *Biometrika*, pages 465–476, 1989.

[Hall and LaScala, 1990]P. Hall and B. LaScala. Methodology and algorithms of empirical likelihood. *International Statistical Review*, pages 109–127, 1990.

[Hoeffding, 1948]W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.*, pages 293 – 325, 1948.

[Lee, 1990]A. J. Lee. *U-statistics, Theory and Practice.* Marcel Dekker, Inc., New York, 1990.

[Owen, 1988]A. B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, pages 237–249, 1988.

[Owen, 1990]A. B. Owen. Empirical likelihood ratio confidence regions. *The Annals of Statistics*, pages 90–120, 1990.

[Owen, 1991]A. B. Owen. Empirical likelihood for linear models. *The Annals of Statistics*, pages 1725–1747, 1991.

[Owen, 2001]A. B. Owen. *Empirical Likelihood.* Chapman and Hall, London, 2001.

[Sen, 1960]P. K. Sen. On some convergence properties of u-statistics. *Calcutta Statistical Association Bulletin*, pages 1–18, 1960.

[Wood *et al.*, 1996]A. T. A. Wood, K. A. Do, and N. M. Broom. Sequential linearization of empirical likelihood constraints with application to u-statistics. *Journal of Computational and Graphical Statistics*, pages 365–385, 1996.