Journal of the Nigerian Association of Mathematical Physics
Volume 7(2003)

Empirical Bootstrap Likelihood for Contrasts of Parameters

S. M. Ogbonmwan
Department of Mathematics
University of Benin, Benin City, Nigeria.

Abstract

The Bootstrap resampling technique is used to obtain simulated likelihoods for $k \geq 2$ contrasts of parameters in this work. We obtained the exact bootstrap empirical likelihoods and compared the results obtained with those of the bootstrap kernel density estimates. The bootstrap likelihoods give the distribution functions for the unknown contrasts of parameters. The comparisons show that good results can be obtained from both the bootstrap exact empirical and the smoothed bootstrap kernel likelihoods.

Key words: *Bootstrap, Likelihoods, Contrasts, Kernel Density likelihood, exact empirical likelihood, confidence interval.*

pp 97 – 104

1.0 Introduction

Ogbonmwan and Wynn (1987) gave an expository introduction to the bootstrap generated likelihood. The paper focused on the use of the Bootstrap resampling and restricted reference set to obtain simulated likelihoods for complex statistical models. The paper sketched a theory of likelihood generation in semi-parametric models which showed itself applicable to the two samples, regression and autoregressive models. Ogbonmwan and Wynn (1992) gave further methods in which alternative sample paths may be generated. In both Ogbonmwan and Wynn (1987, 1992) papers, the examples considered centred around the two samples problem, the simple linear regression and the first order autoregressive models. In an earlier work, Ogbonmwan (1986) suggested that for the $k \geq 2$ problems, a calculation of the maximum of the absolute differences of pairs of means for each of the bootstrap samples got after transformation of the data set should be used as the generating seeds of the likelihood. However, to the best of our knowledge, this suggestion regarding the $k \geq 2$ problems has never been implemented. In this work, we extend the simulated Bootstrap generated likelihood to the $k \geq 2$ problems of contrasts of parameters.

A related technique is the empirical likelihood which was introduced by Owen (1988, 1990) as an alternative method to the bootstrap for the construction of confidence regions in nonparametric problems. Diciccio, Hall and Romano (1991) showed that empirical likelihood has a major advantage over the bootstrap and related techniques. One of such advantages is that empirical likelihood admits a Bartlett correction while Bootstrap does not. The works of Chen (1993, 1994, 1996) used empirical likelihood in alliance with the kernel density approach to construct confidence interval for the value of a probability density f at a given point x (say). Chen and Qin (2002) considered empirical likelihood in conjunction with the local linear smoother to construct confidence intervals for nonparametric regression function with bounded support. The paper provided significant improvement over confidence intervals based directly on the asymptotic normal distribution of the local linear estimator. Recently, a likelihood based inference was used to explore classification models in the nearest – neighbourhood by Holmes and Adams (2003). The method developed in Holmes and Adams (2003) incorporates a conventional generalized linear model and a conventional k nearest neighbour algorithm. In this paper and in the earlier works of Ogbonmwan (1986) and Ogbonmwan and Wynn (1987, 1992), we combine the ideas of empirical distribution and the Bootstrap to generate our likelihoods. The theoretical demands are conceptually not rigorous yet computationally intense. We simply make use of very simple bootstrap (of Efron (1982)) and the kernel density smoothing to produce accurate results in a variety of examples. Davison and Hinkley (1988) used saddlepoint approximations in resampling methods for a variety of standard bootstrap and randomization applications.

In the generation of the likelihoods, two methods are developed. We generate the exact bootstrap empirical likelihoods and also generate the bootstrap kernel density estimates of the likelihoods. While the bootstrap kernel density estimation approach has been used in the earlier work in Ogbonmwan (1986) and Ogbonmwan and Wynn (1987, 1992), the exact bootstrap empirical method is being used for the first time for the actual generation of likelihood. The main difference with respect to the different approaches is that while exact empirical method gives the exact estimates of the theoretical likelihood, the kernel density method rather estimates the densities of the approximate likelihood and thereby gives an estimate of the exact likelihood. We compare via confidence intervals, the behaviour of the two methods as well as the classical t–test, the Tukeys and Schffe's methods by using numerical 2, 3, and 4 different one –dimensional samples. The results obtained allow us to make comments on the different methods with respect to the confidence intervals and the distributions obtained.

Therefore, the plan of this paper is to extend the simulated Bootstrap generated likelihood, to the $k \geq 2$ samples problems, and develop (and of course) generate the exact empirical likelihoods. The paper rest of the is organized as follows: In section 2, we briefly describe the main steps of the Bootstrap method for generating likelihood. Two procedures are adopted to generate the Bootstraps likelihood. The extension into $k \geq 2$ samples problems is discussed in section 3. In section 4, the results of simulations in numerical examples are compared. The study allows us to make general remarks and draw some conclusions. This is done in section 5.

2.0 The Main Steps

Consider that the parameter being estimated by a statistic T emanates from a family of transformations of the data set $X = (X_1, X_2, ..., X_n)$ to Y_θ such that

$$Y_\theta = (y_1, y_2, y_3, ..., y_n)_\theta = g_\theta(x) \quad (2.1)$$

We assume that the transformed data set $y_1, y_2, ..., y_n$ are independent and identically distributed with distribution function F, not depending on θ . For convenience we also assume that $y_{0,i} = X_i, i = 1, 2, ..., n$. If the statistic of interest on X is defined by T = $t(x)$, then we define $T_\theta = t(y_\theta)$ and take the partial likelihood for θ to be the density of T_θ and at the observed value of t_θ. The bootstrap likelihood is obtained by first considering the Bootstrap values of the statistic $T(y_\theta)$ and then let the bootstrap values be listed in some order T_1^*, T_2^*, T_B^* which will have the empirical cumulative density function (cdf) of the form

$$P_T(t|\theta) = \frac{1}{B} \# \left\{ T_i^* \middle| T_i^* \leq t \right\} \quad (2.2)$$

This empirical cdf in (2.2) is smoothed to obtain a continuous density $\hat{f}_T(t|\theta)$ by using the kernel density estimator. Thus the bootstrap likelihood is the density of the bootstrap value T_θ^* at t_θ, where $T_\theta^* = t(y_\theta)$ sampled randomly from the transformed data set $(y_1^*, y_2^*, ..., y_n^*)_\theta$.

A simplified version that yields the exact empirical likelihood is to do a count of the proportion of T_j^* values that lie in some intervals around j. Thus for some $\varepsilon > 0$, the exact empirical likelihood is defined as:

$$L(\theta) = \frac{1}{B} \# \left(T_j^* \middle| \tilde{t} - \varepsilon \leq T_j^* \leq \tilde{t} + \varepsilon \right) \quad (2.3)$$

3.0 The $k \geq 2$ Samples Problems:

Let us consider the design in which every individual belongs to one and only one of k distinct samples (levels) with means $\mu_1, \mu_2, ..., \mu_k$. Suppose we wish to estimate or test hypothesis about a linear function of the means of the form:

$$\sum a_i \mu_i = a_1\mu_1 + a_2\mu_2 + ... + a_k\ \mu_k \tag{3.1}$$

where a_1, a_2, ...,a_k are pre-selected constants. Thus in this case, $\sum a_i \mu_i$ would be estimated by $\sum a_i \overline{X}_i$, where $\overline{X}_i$ is the mean of the sample of n_i individuals taken from the i^{th} sample. Generally, linear functions of $\mu_1, \mu_2, ..., \mu_k$ can have any desired coefficients a_1, a_2,...,a_k, and in particular for a contrast of the parameters, $\sum a_i = 0$ and for convenience, we restrict the a_is, $i = (1\,(k))$, so that the sum of the positive valued $a_i s$ is 1 and the sum of the negative valued $a_i s$ must also equal 1.

3.1 Two Samples

Let $X_1, X_2, ..., X_m$ and $Z_1, Z_2, ..., Z_n$ be two samples drawn from two distributions whose means differ by the shift parameters $\theta = E(Z) - E(X)$ which can be estimated by $\overline{Z} - \overline{X}$. Our interest will be to generate the likelihood for θ. Let us for simplicity define $X_{m+i} = Z_i$, i = 1, 2, 3, ., ., ., n and let N = $m + n$.
Then the transformed data set

$$y_\theta = (X_1, X_2, ..., X_m, Z_1 - \theta, Z_2 - \theta, ..., Z_n - \theta) \tag{3.2}$$

could be expressed as:

$$y_{\theta}, i = \begin{cases} X_i, & (i = 1, 2, ..., m) \\ X_i - \theta & (i = m+1, m+2, ..., N) \end{cases} \tag{3.3}$$

since $X_{m+i} = Z_i$ and hence the estimate T_θ is defined by

$$T_\theta = \frac{1}{n}\sum_{i=m+1}^{N} Y_{\theta 1 i} - \frac{1}{m}\sum_{i=1}^{m} Y_{\theta 1 i} \tag{3.4}$$

The Bootstrap likelihood of θ is the density approximation

$$L^*(\theta) = L(\theta | x, z) = \hat{f}_{T^*_\theta}(t_\theta) \tag{3.5}$$

where $T_\theta = \frac{1}{n}\sum_{i=m+1}^{N} Y_{\theta 1 i} - \frac{1}{m}\sum_{i=1}^{m} Y_{\theta 1 i}$ with $y^*_{\theta,i}$'s as the standard bootstrap samples from $(y_{\theta 1}, y_{\theta 2}, ..., y_{\theta N})$

and t_θ is the observed value of T_θ. (i.e $t_\theta = t - \theta$). By adopting the kernel density estimator, we have

$$L^*(\theta) = \hat{f}_{T^*_\theta}(t) = \hat{f}(t) = \frac{1}{Bh}\sum_{i=1}^{B} k\left(\frac{t - T^*_i}{h}\right) \tag{3.6}$$

where $k(.) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ and h is the approximate optimal window width (of order 2). The exact empirical likelihood is defined by the distribution

$$L^{**}(\theta)=\frac{1}{B}\ \#\left(T_j^*\middle|\ \tilde{t}-\varepsilon\le T_j^*\le\tilde{t}+\varepsilon\right)\ \text{for some } \varepsilon>0. \tag{3.7}$$

3.2 k – Samples, k > 2

First, let us now consider the comparison of two samples mean with the mean of any third sample in a 3 – sample situation. Intuitively, we see that we are considering the contrast of the form: $\frac{1}{2}(\mu_X+\mu_Z)-\mu_Y$. Analogously, let $X_1, X_2, \ldots, X_m$, $Z_1, Z_2, \ldots, Z_n$, $\gamma_1, \gamma_2, \ldots, \gamma_\ell$ be three samples. To generate likelihood for a parameter θ (say) we simply would consider:

$$y_\theta=(x_1, x_2, \ldots, x_m,\ Z_1, Z_2, \ldots, Z_n,\ \gamma_1-\theta,\ \gamma_2-\theta,\ \gamma_3-\theta, \ldots, \gamma_\ell-\theta) \tag{3.8}$$

If $X_{m+i}=Z_i, i=1, 2, \ldots, n$ and $X_{m+n+i}=\gamma_{i,}=m+n+1,\ m+n+2, \ldots, m+n+\ell$ Then we could express (3.8) as:

$$y_{\theta,i}=\begin{cases}X_i, & i=1, 2, 3, \ldots, m\\ X_i, & i=m+1,\ m+2, \ldots, m+n\\ X_i-\theta, & i=m+n+1,\ m+n+2, \ldots m+n+\ell\end{cases}$$

The estimate of the statistic of interest T_θ is defined by

$$T_\theta=\frac{1}{2}\left(\frac{1}{m}\sum_{i=1}^{m}y_{\theta,i}+\frac{1}{n}\sum_{i=m+1}^{m+n}y_{\theta,i}\right)-\frac{1}{\ell}\sum_{i=m+n+1}^{m+n+\ell}y_{\theta,i} \tag{3.9}$$

We then consider the simple with replacement sampling taken from the elements of y_θ to get the bootstrap sample of the form: $y_\theta^*=\left(y_1^*, y_2^*, \ldots, y_{m+n+\ell}^*\right)_\theta$. Then consider the bootstrap sample statistic:

$$T\left(y_\theta^*\right)=\frac{1}{2}\left(\frac{1}{m}\sum_{i=1}^{m}y_i^*+\frac{1}{n}\sum_{i=1}^{m+n}y_i^*\right)-\frac{1}{\ell}\sum_{i=m+n+1}^{m+n+\ell}y_i^* \tag{3.10}$$

and hence get

$$t=T(y_\theta)=\frac{1}{2}\left(\bar{X}+\bar{Z}\right)-\bar{\gamma}+\theta. \tag{3.11}$$

with which to obtain the Bootstrap likelihood $L(\theta)=\overset{*}{f}_T\left(\tilde{t}\middle|\theta\right)$ by using the kernel density estimator and hence express L(θ) as:

$$L^*(\theta)=\overset{*}{f}_{T_\theta^*}(t)=\hat{f}(t)=\frac{1}{Bh}\sum_{i=1}^{B}k\left(\frac{t-T_i^*}{h}\right) \tag{3.12}$$

where k(.) and h take their usual meanings. Again, the exact empirical likelihood is defined by the distribution

$$L^{**}(\theta)=\frac{1}{B}\ \#\left(T_i^*\middle|\ \tilde{t}-\varepsilon\le T_i^*\le\tilde{t}+\varepsilon\right)\quad \text{for some } \in>0 \tag{3.13}$$

The analytical ideas used for the three samples problems can be extended to four, five or more samples problem by following the contrasts rules. This can therefore lead us into lots of solutions of the one–way analysis of variance problem especially for the cases where significant difference is suspected. We shall consider the four samples problems of a very practical nature. For instance, if among four soil treatments (fertilizers) two are known to contain high quality Nitrogen compound at two different levels, while the remaining two do not contain Nitrogen compound but also appear at two different levels. We may for this experiment want to compare the two types of fertilizers, viz: one with high quality Nitrogen compound and the other without Nitrogen compound. Thus, we will be interested in considering the contrast of the form: $\frac{1}{2}(\mu_1+\mu_2)-\frac{1}{2}(\mu_3+\mu_4)$, where μ_1 and μ_2 are the means of those that contain the high quality Nitrogen compound and μ_3 and μ_4 are the means of those that do not contain Nitrogen. Let $X_1, X_2, \ldots, X_m, Z_1, Z_2, \ldots, Z_n, \gamma_1, \gamma_2, \ldots, \gamma_\ell, S_1, S_2, \ldots, S_u$ be four samples. Let θ be the parameter of interest and let us consider the transformed data set $y_\theta = (x_1, x_2, \ldots, x_m, Z_1, Z_2, \ldots, Z_n, \gamma_1-\theta, \gamma_2-\theta, \gamma_3-\theta, \ldots, \gamma_\ell-\theta, S_1-\theta, S_2-\theta, \ldots, S_u-\theta)$.

Consider the bootstrap version of y_θ to get $y_\theta^* = \left(y_1^*, y_2^*, \ldots, y_{m+n+\ell+n}^*\right)_\theta$ and finally consider the statistic

$$T\left(y_\theta^*\right)=\frac{1}{2}\left(\frac{1}{m}\sum_{i=1}^{m}y_i^*+\frac{1}{n}\sum_{i=m+1}^{m+n}y_i^*\right)-\frac{1}{2}\left(\frac{1}{\ell}\sum_{i=m+n+1}^{m+n+\ell}y_i^*+\frac{1}{u}\sum_{i=m+n+\ell+1}^{m+n+\ell+u}y_i^*\right) \tag{3.14}$$

and hence get the Bootstrap likelihood $L(\theta)=\hat{f}_{T^*}\left(\tilde{t}\,\middle|\,\theta\right)$ by using the kernel density estimator and take t to be

$$t=T(y_\theta)=\frac{1}{2}\left(\overline{X}+\overline{Y}\right)-\frac{1}{2}\left(\overline{\gamma}-+\overline{S}\right)+\theta \tag{3.15}$$

The exact empirical likelihood distribution is defined by

$$L^{**}(\theta)=\frac{1}{B}\,\#\left(T_i^*\,\middle|\,t-\varepsilon\le T_i^*\le t+\varepsilon\right).$$

4.0 Examples and Results

Three illustrative examples are considered, one each for the 2, 3, and 4 samples problems. The differences of means/contrasts of the original data set were calculated by following the procedure described in section 3. Various values of θ (i.e. the transforming parameter) were assumed to be the "true" difference of means/contrasts. Subsequently, the original data set were transformed with each chosen value of θ and each transformed data set was bootstrapped and eventually the exact empirical likelihood as well as the kernel density estimate (normalized) of the bootstrap estimate of the differences of means/contrast (which we simply call the normalized kernel density likelihood) were estimated. In each of the examples, the Gaussian kernel function was used. The realization of the exact empirical likelihood was made by first making a choice of an epsilon ε value and counting the proportion of the number of differences of means/contrasts that lie within the epsilon neighbourhood of the "true" differences of means/contrasts, that is, do a count of the number of bootstrap differences of means/contrasts that lie within the interval $[t_0-\varepsilon, t_0+\varepsilon]$ where t_0 is the value of the true differences of means/contrasts for the transformed observed data set. ε is chosen subjectively but as a form of "histogram bin width" we propose that its value should be close to that of the window width used for the kernel density estimation. Thus, it is the choice of the epsilon ε that primarily controls the amount of smoothing inherent in the procedure. We illustrate these calculations with the following dataset in Tables 1 – 3.

Table 1: Percentages of iron in ore samples determined by two methods

Method A:	38.25, 31.68, 26.24, 41.29, 44.81, 46.37, 35.42, 38.41, 42.68, 46.71, 29.20, 30.76
Method B:	38.27, 31.71, 26.22, 41.33, 44.80, 46.39, 35.46, 38.39, 42.72, 46.76, 28.18, 30.79

Source: *Scheffer, R. L. and McClave, J. T. (1990), Probability and Statistics for Engineers 3rd Ed. p.331*

Table 2

	Test 1: 35, 42, 42, 30, 15, 31, 29, 29, 17, 21
Experimental Group	Test 2: 34, 38, 26, 17, 42, 28, 35, 33, 16, 40
Control Group	Test 3: 17, 29, 30, 36, 41, 30, 31, 23, 38, 30

Source: *Conover, W. J. ( ), Practical Nonparametric Statistics, 2nd Ed. p.*

Table 3: Pre and Post experimental Pd T scores following the administration of MMPI

Pre – Test 1: 67, 86, 64, 69, 67, 79, 67, 67, 69, 57, 76, 90, 71, 93
Pre – Test 2: 88, 79, 67, 83, 79, 76, 71, 67, 69, 67, 69, 74, 81, 81
Post – Test 1: 74, 50, 64, 76, 64, 81, 74, 50, 60, 57, 62, 76, 71, 76
Post – Test 2: 79, 81, 83, 74, 76, 69, 71, 75, 64, 64, 64, 71, 74, 64

Source: *Steel, G. D. and Torrie, J. H. (1982), Principles and Procedures of Statistics, 2nd Ed. p. 145.*

The data in Tables 1, 2, and 3, serve as examples for the 2, 3, and 4 samples problems respectively. We evaluated our results on 3000 bootstrap configurations with appropriate choices of the transforming parameter θ. The application of equations (3.6) and (3.7) to the data of the three examples yielded the plots of the graphs of the distributions of the exact bootstrap empirical likelihood and that of the normalized bootstrap kernel density likelihood. These graphs are shown in figures 1, 2, and 3 for the respective examples. The approximate 95% confidence intervals so obtained for the bootstrap differences of means (example 1) and contrasts (examples 2 and 3) are given in Table 4. Also in Table 4 are the 95% confidence interval for the original data set using the t – test, the Tukey's method as well as the Scheffe's method.

Table 4: Summary of results for the 95% confidence intervals for the difference of means/contrasts

Procedure	**95% Confidence Interval**		
	2 Samples (Example 1)	**3 Samples (Example 2)**	**4 Samples (Example 3)**
t – test	–5.92, 6.04	–7.26, 6.26	–0.14, 9.21
Tukey's Method	–5.92, 6.04	–9.95, 8.95	– 4.22, 13.29
Scheffe's Method	–5.97, 5.97	–8.54, 8.54	–6.74, 6.74
Efron's Bootstrap	–5.45, 5.27	–6.08, 6.08	– 4.62, 4.49
Percentile Method	–5.26, 5.03	–5.95, 6.30	– 4.57, 4.57
Efron's Bootstrap Bias Corrected Method	–5.5, 5.24	– 4.78, 7.28	– 4.46, 4.6

5.0 Conclusion:

The exact empirical bootstrap likelihood and the kernel bootstrap likelihood have been examined and applied to contrasts of parameters for $k \geq 2$ samples problems. Results from 3000 bootstrap configurations were used in getting the bootstrap 95% confidence interval for the difference of means/contrasts. In the three examples considered, the bootstrap percentile method provided shorter confidence intervals. The graphs of the likelihood in figures 1 – 3, provide the distributions of the parameters examined. This is really something from nothing. This is the beauty of the procedure we have evolved in the generation of bootstrap likelihoods. Care must be exercised in the choice of epsilon, since the value of epsilon determines the smoothness of the exact empirical bootstrap. The results obtained for both bootstrap likelihoods compare favourable with other theoretical statistical estimation procedures like the t–test, the Tukey's method and the Scheffe's method for contrasts of parameters. In the three examples considered both likelihoods appears to give very close results. The graphs and indeed the distributions of both bootstrap likelihoods are symmetric about the actual difference of means/contrasts. The normalized kernel bootstrap likelihood presents smoother curves when compared with the exact empirical likelihood. The smoothness of the normalized kernel bootstrap likelihood (see figures 1 – 3) should not be seen as an edge over the exact empirical bootstrap likelihood since the kernel method is conditioned on the assumption of the Gaussian function as its kernel. In fact, the kernel likelihood makes use of

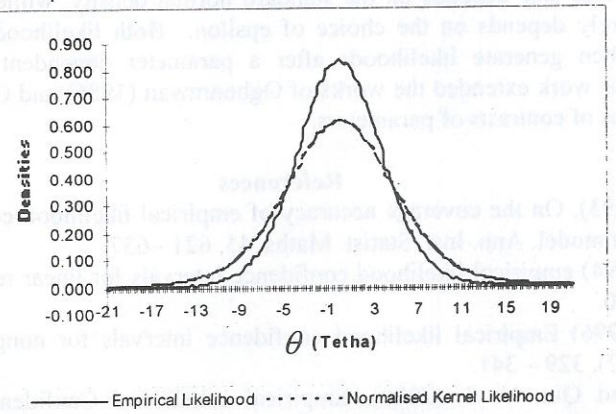

Figure 1: Empirical and Normalized Kernel Likelihood for 2-Sample Problems

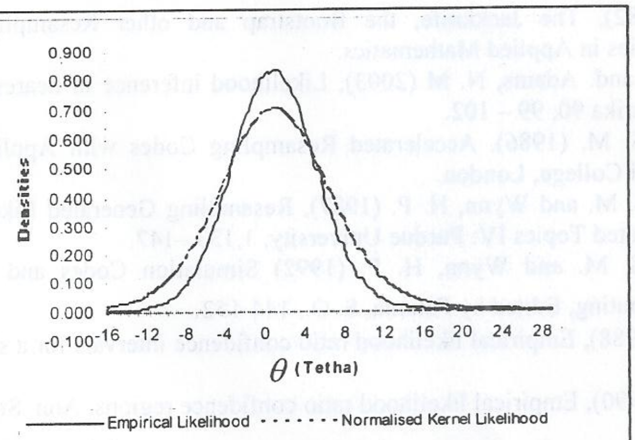

Figure 2: Empirical and Normalized Kernel Likelihood for 3-Sample Problems

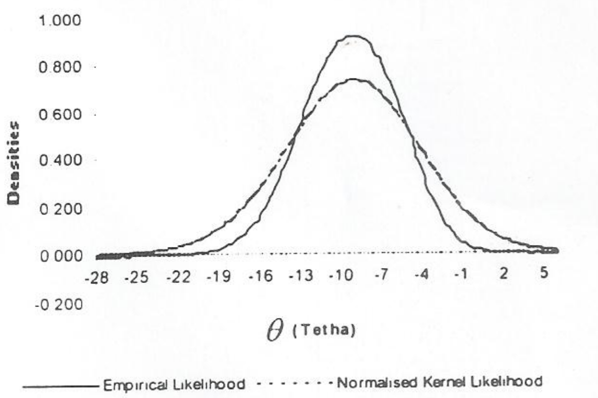

Figure 3: Empirical and Normalized Kernel Likelihood for 4-Sample Problems

a smoothing function whose size depends on the standard normal density, while the smoothness of the exact empirical likelihood merely depends on the choice of epsilon. Both likelihoods take into account only the presence of data and then generate likelihoods after a parameter dependent transformation followed by bootstrap. We have in this work extended the works of Ogbonmwan (1986) and Ogbonmwan and Wynn (1987, 1992) to the k≥2 problems of contrasts of parameters.

References

[1] Chen, S. X. (1993). On the coverage accuracy of empirical likelihood confidence interval regions for linear regression model, Ann. Inst. Statist. Maths. 45, 621 - 637

[2] Chen, S. X. (1994) empirical likelihood confidence intervals for linear regression coefficients J. Mult. Anal. 49, 24 – 40.

[3] Chen, S. X. (1996) Empirical likelihood confidence intervals for nonparametric density estimation Biometrika, 83(2), 329 – 341.

[4] Chen, S. X. and Qin, Y. S. (2000). Empirical Likelihood Confidence intervals for local linear smoothers, Biometrika, 87, 946 – 953.

[5] Davison, A. C. and Hinkley, D. V. (1988), Saddlepoint approximations in resampling methods, Biometrika, 75(3), 417 – 431.

[6] Diciccio, T. J., Hall, P. and Romano, J. P. (1991), Empirical likelihood is Bartlett correctable, Am Statist. 19,1053 – 1061.

[7] Efron, B. (1982), The Jackknife, the Bootstrap and other Resampling Plans, SIAM Regional Conference Series in Applied Mathematics.

[8] Holmes, C. C. and. Adams, N. M (2003), Likelihood inference in nearest – neighbour classification models; Biometrika 90, 99 – 102.

[9] Ogbnomwan, S. M. (1986), Accelerated Resampling Codes with Application To likelihood, Ph.D Thesis, Imperial College, London.

[10] Ogbonmwan, S. M. and Wynn, H. P. (1987), Resampling Generated Likelihood, Statistical Decision Theory and Related Topics IV: Purdue University, 1,133 –147.

[11] Ogbonmwan, S. M. and Wynn, H. P. (1992) Simulation Codes and Likelihood Generation: In Scientific Computing, Edited by Fatunla, S. O., 144-152.

[12] Owen, A. B. (1988), Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75, 237 – 249.

[13] Owen, A. B. (1990), Empirical likelihood ratio confidence regions, Ann, Statist. 18, 90 – 120