Available online at www.elixirpublishers.com (Elixir International Journal)

Applied Mathematics

ISSN: 2229-712X

Elixir Appl. Math. 63C (2013) 18099-18105

Considering the Consistency for the Information Criterion AIC

Masumeh Ghahramani

Department of Mathematics, Gilan-E-Gharb Branch, Islamic Azad University, Gilan-E- Gharb, Iran.

ARTICLE INFO

Article history:

Received: 18 September 2013;

Received in revised form:

25 September 2013;

Accepted: 1 October 2013;

Keywords

Bias,

Consistency,

Information Criterion,

Kullbake-Leibler Risk,

Model Selection.

ABSTRACT

Akaike information criterion, AIC is widely used for model selection. The AIC as the estimator of asymptotically unbias for the second term Kullbake-Leibler risk considers the divergency between true model and offered models. The AIC, is an inconsistent estimator. In this article the proposed approach the problem the inconsistency of AIC, it is the use of consistent offered information criterion, called M_1IC (Masumeh information criterion). At the end of these two information criteria of model selection of classic and linear models have been considered by the simulation of Monte-Carlo.

Introduction

Statistical modeling is used for investigating a random phenomenon that isn't completely predictable. One of the criteria that have usage of the frequency in model selection is Kullbake-Leibler (KL) information criterion (see Kullback and Leibler 1951). This information criterion was introduced as one risk in model selection. Akaike (1973) introduced information criterion, AIC as asymptotically the unbiased of an estimator for the second term the KL risk and to form penalty likelihood function. Akaike stated modeling isn't only finding a model which describes the behavior of the observed data, but its main aim is predicted as a possible good, the future of the process under investigation. Hall (1987) by using the Kullbake-Leibler risk considered bias and variance in the approximate density function. Konishi and Kitagawa (1996) considered the analysis of bias with the method of bootstrap in the model selection. Bozdogan (2000) with the error distinction in the model selection considered two errors from bias and variance in the estimation of model selection. Choi and Kiffer (2006), Cawley and Talbot (2010) have considered the over fitting in model selection, and they showed the over fitting is resulting from the bias when modeling phenomena have been considered. Ghahramani (2013) has considered the inconsistent of information criterion KIC. The during these years has been made the corrections on penalty term, and criteria such as AIC (Akaike 1973), TIC (Takeuchi 1976), and KIC (Cavanaugh 1994) are introduced. In section 2, is stated the Kullbake-Liebler risk, and necessary of definitions. In section 3, a consistent information criterion is proposed instead of the AIC. In section 4, we present the results of our simulation studies.

Kullbake-Leibler (KL)

Let X $=(X_1, X_2, \dots, X_n)$ is a (i.i.d) random sample from true model and unknown, $h(.)$, and the family F_{θ_k}={$f(.;\theta_k) = f_{\theta_k}$; $\theta_k \epsilon \Theta \subseteq R^k$ } from offered models has been considered for approximate true model.

Definition 1). The family F_{θ_k} is well specified, if there is a $\theta_0 \epsilon \Theta$ such that $h(.) = f(.;\theta_0)$; otherwise it is mis specified

Definition 2). The KL risk definds for generate model and unknown $h(.)$, and offered model f_{θ_k} as

$$KL(h, f_{\theta_k}) = E_h\left[\log\left(\frac{h(.)}{f(.;\theta_k)}\right)\right] = E_h[\log h(.)] - E_h[\log f(.;\theta_k)] \quad (1)$$

Where the expectation is taken with respect to the unknown model $h(.)$. The first term in the right hand side of (1) is called irrelevant part, because it doesn't depend on θ_k, and the second term is called relevant part. Based on the properties of the KL risk, the smaller value showed the closeness of the offered model to the unknown and true model. Therefore the problem reduces to obtain a

Tele:

E-mail addresses: mghdatagilan@gmail.com

good estimate of the expected log-likelihood. Since the expectation is with respect to the model with unknown parameters, one estimator is

$E_h \{ \log f(. ; \hat{\theta}_n)\} = \frac{1}{n} \sum_{i=1}^{n} \log f(x_i ; \hat{\theta}_n)$.

So that $\hat{\theta}_n$ is the maximum likelihood estimator of θ_k and $f(. ; \hat{\theta}_n)$ is the maximum likelihood function. The bias of maximum log-likelihood is as

bias estimator = $E_h \{\log f(. ; \hat{\theta}_n) - nE_h \{\log f(Z ; \hat{\theta}_n)\}$

So that Z is a random variable (i.i.d) with X_i s. The general form of the information criterion that has been shown by IC, as

IC = -2 (log-likelihood of statistical model – bias estimator)

$$= -2\sum_{i=1}^{n} \log f(X_i ; \hat{\theta}_n) + 2\{\text{ bias estimator}\} = -2\, l_f(\hat{\theta}_n) + 2\{\text{ bias estimator}\}.$$

Akaike, when offered family is well specified, size of bias is estimated with dimensional parameter $\hat{\theta}_n$, means k, and Akaike information criterion, is stated as

$$\text{AIC} = -2\sum_{i=1}^{n} \log f(X_i ; \hat{\theta}_n) + 2k = -2\, l_f(\hat{\theta}_n) + 2k.$$

With attention to form the AIC by increasing the number of parameters in the offered model the penalty term, 2k will be increased and the term $-2\sum_{i=1}^{n} \log f(X_i ; \hat{\theta}_n)$ will be decrease. Penalty term is constant to chance of size sample in the information criterion AIC, and by increasing the size sample, AIC cannot distinguish the true model with the probability one. Therefore this problem is the same concept of inconsistency for an information criterion. Following the inconsistency of information criterion AIC, based on the definition similar to the definition of AIC, a consistent of information criterion which called M_1IC has presented. Akaike information criterion, by Akaike for model selection is introduced, but this useful criterion is inconsistent (see Akaike 1973). In this selection the bias term has used in the general form information criterion is considered from another perspective. We obtain the information criterion that furthermore has nice specials the information criterion AIC, it's also consistent. In the beginning the bais of the log-likelihood function as follows:

$$b = E_h \{\log f(. ; \hat{\theta}_n) - nE_h \log f(Z ; \hat{\theta}_n)\}$$

Where in the second term of the right hand side the inner expectation is calculated with respect to h(z) and the outer expectation is calculated with respect to h(x). By evaluating the bias it is composed as follows:

$$b = E_h \{ \log f(. ; \hat{\theta}_n) - \log f(. ; \theta_0)\} + E_h \{ \log f(. ; \theta_0) - n E_h \{\log f(Z; \theta_0)\}\}$$
$$+ n E_h \{ E_h\{\log f(Z ; \theta_0) - E_h \{\log f(Z; \hat{\theta}_n)\}\} = b_1 + b_2 + b_3 .$$

We calculate the three expectations separately b_1, b_2 and b_3.

a) For calculation of b_1 by writing $l_f(\theta_0) = \log f(. ; \theta_0)$ and by applying a Taylor series expansion around the maximum likelihood estimator $\hat{\theta}_n$, we have

$$l_f(\theta_0) = l_f(\hat{\theta}_n) + (\theta_0 - \hat{\theta}_n)^T \frac{\partial l_f(\theta)}{\partial \theta}\Big|_{\theta = \hat{\theta}_n} + \frac{1}{2}(\theta_0 - \hat{\theta}_n)^T \frac{\partial^2 l_f(\theta)}{\partial \theta\, \partial \theta^T}\Big|_{\theta = \hat{\theta}_n} (\theta_0 - \hat{\theta}_n) + o_p(1),$$

$o_p(1)$ is expression of quantity that in the probability tends to zero.

With attention to, the $\frac{\partial l_f(\theta)}{\partial \theta}\Big|_{\theta = \hat{\theta}_n} = 0$ and $\frac{1}{n}\frac{\partial^2 l_f(\theta)}{\partial \theta\, \partial \theta^T}\Big|_{\theta = \hat{\theta}_n}$ is converge to $J(\theta_0)$. (for more study see Akaike 1973). So

$$J(\theta_0) = -E_h\left[\frac{\partial^2 l_f(\theta)}{\partial \theta\, \partial \theta^T}\right]\Big|_{\theta = \theta_0} .$$

Thus, the relation above can be approximated, as

$$l_f(\hat{\theta}_n) - l_f(\theta_0) \approx \frac{n}{2}(\theta_0 - \hat{\theta}_n)^T J(\theta_0) (\theta_0 - \hat{\theta}_n) + o_p(1),$$

This based on the b_1 can be written as follow

$$b_1 = E_h\{l_f(\hat{\theta}_n) - l_f(\theta_0)\} \approx E_h\{ \frac{n}{2}(\theta_0 - \hat{\theta}_n)^T J(\theta_0) (\theta_0 - \hat{\theta}_n) \} \quad (2).$$

b) The b_2 doesn't contain an estimator and it can easily be written as

$$b_2 = E_h \{ \log f(. ; \theta_0) - n E_h \{\log f(Z;\theta_0)\}\} = 0 \qquad (3).$$

c) For calculation of value the b_3, first, the phrase $E_h \{\log f(Z;\theta_0)\}$ be definded equally of $\eta(\hat{\theta}_n)$. By using from Taylor expectation $\eta(\hat{\theta}_n)$ around θ_0, we have

$$\eta(\hat{\theta}_n) = \eta(\theta_0) + (\hat{\theta}_n - \theta_0)^T \frac{\partial \eta(\theta)}{\partial \theta}|_{\theta=\theta_0} + \frac{1}{2}(\hat{\theta}_n - \theta_0)^T \frac{\partial^2 \eta(\theta)}{\partial \theta \partial \theta^T}|_{\theta=\theta_0}(\hat{\theta}_n - \theta_0) + o_p(1),$$

with attention to the $\frac{\partial \eta(\theta)}{\partial \theta}|_{\theta=\theta_0} = 0$. Thus when n tends to infinity, the relation above can be approximated as $\eta(\hat{\theta}_n) \approx \eta(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^T J(\theta_0)(\hat{\theta}_n - \theta_0) + o_p(1)$.

Thus the b_3 can be written as

$$b_3 = n E_h\{ E_h\{ \log f(Z ; \theta_0)\} - E_h \{\log f(Z ; \hat{\theta}_n)\}\} \approx \frac{n}{2} E_h\{(\theta_0 - \hat{\theta}_n)^T J(\theta_0)(\theta_0 - \hat{\theta}_n)\} \qquad (4)$$

If the family of F_{θ_k} is well specified, with attention to quadratic forms in relations (2) and (4), that converge to centrally distributed chi-square with k degrees of freedom. Therefore b_1 and b_3 can be written as

$$b_1 = b_3 = \frac{n}{2}k \qquad (5).$$

So by combining of b_1 and b_3, in relation (5) and b_2, in relation (3), bias the b is as follow

$$b = b_1 + b_2 + b_3 = nk.$$

With replacing the value of b in the general form of the information criterion, the offered information criterion called, M_1IC is obtained as

$$M_1IC = -2\sum_{i=1}^{n} \log f(X_i ; \hat{\theta}_n) + 2nk = -2\, l_f(\hat{\theta}_n) + 2nk \qquad (6)$$

In the offered information criterion M_1IC, penalty term 2nk changes will change with sample size changes. So, if sample size will be very large, information criterion M_1IC, with the probability of one, find the true model data. In other words information criterion M_1IC, is the only consistent information criterion, that has been obtained based on Kullback-Leibler risk. To show consistency of information criterion M_1IC, let the maximum likelihood function estimator for the offered model ($f(.;\theta_k) = f(\theta_k)$) and optimal model ($f(.;\theta_{k_0}) = f(\theta_{k_0})$) with respectively $l_f(\hat{\theta}_{k(n)})$ and $l_f(\hat{\theta}_{k_0(n)})$. With regard to relation (6) information criterion M_1IC, for the model $f(\theta_k)$ and $f(\theta_{k_0})$, we have

$$M_1IC(f(\theta_k)) = -2l_f(\hat{\theta}_{k(n)}) + 2nk,$$

$$M_1IC(f(\theta_{k_0})) = -2l_f(\hat{\theta}_{k_0(n)}) + 2n k_0$$

If there is $k > k_0$, consistency for information criterion M_1IC is given by

$$P(M_1IC f(\theta_k) - M_1IC f(\theta_{k_0}) > 0)$$

$$= P(-2l_f(\hat{\theta}_{k(n)}) + 2nk - (-2l_f(\hat{\theta}_{k_0(n)}) + 2n k_0) > 0)$$

$$= P(2l_f(\hat{\theta}_{k(n)}) - 2l_f(\hat{\theta}_{k_0(n)}) < 2nk - 2n k_0)$$

$$= P(U_n < 2n(k - k_0)) = F(2n(k - k_0)) \xrightarrow{p} F(\infty) = 1 \qquad (7)$$

In relation (7), U_n is $2l_f(\hat{\theta}_{k(n)}) - 2l_f(\hat{\theta}_{k_0(n)})$ And the distribution function of chi-square has been shown by F. Therefore it tends in of the probability to one. Thus M_1IC is a consistent information criterion. (For further study about the consistency of an information criterion, see Hu and Shao 2008).

Simulation study

This simulation has been accomplished for usage and comparison of the offered information criterion, M_1IC, with the information criterion AIC, by using Monte -Carlo simulation, for linear regression and classic models. This simulation of linear regression model is supposed that well specified family $F_{\theta_k}=\{f(.;\theta_k)=f_{\theta_k};\ \theta_k\epsilon\,\Theta\subseteq R^k\}$, and mis specified family $G_{\beta_d}=\{g(.;\beta_d)=g_{\beta_d};\ \beta_d\epsilon\,\mathrm{B}\subseteq R^d\}$ are given for estimating the true model. Let $f: y_i=0.2+0.7x_{i1}+x_{i2}+0.6x_{i3}+\varepsilon_{i1}$ i=1,…,n is as the true model so that ε_{i1}, has been generated as random from distribution N(0,2). Models $f_1: y_i=\hat{\theta}_0+\hat{\theta}_1x_{i1}+\hat{\theta}_2x_{i2}+\hat{\theta}_3x_{i3}$ i=1,…,n

And $f_2: y_i=\hat{\theta}_0+\hat{\theta}_1x_{i1}+\hat{\theta}_2x_{i2}+\hat{\theta}_3x_{i3}+\hat{\theta}_4x_{i4}\quad i=1,\dots,n$ offered models, which have been generated from F_{θ_k}. Also we have $g: y_i=0.5+0.4z_{i1}+2z_{i2}+0.9z_{i3}+\varepsilon_{i2}$ i=1,…,n

So that ε_{i2}, has been generated as random from distribution N(0,1), and Models $g_1: y_i=\hat{\beta}_0+\hat{\beta}_1z_{i1}+\hat{\beta}_2z_{i2}+\hat{\beta}_3z_{i3}$ i=1,…,n and $g_2: y_i=\hat{\beta}_0+\hat{\beta}_1z_{i1}+\hat{\beta}_2z_{i2}+\hat{\beta}_3z_{i3}+\hat{\beta}_4z_{i4}\quad i=1,\dots,n.$

Offered models, are generated from G_{β_d}. This simulation is achieved by using from software R, and the number of repetitions are 10^4, and samples n = 50, 100, 150, 200, 350, 500, have been considered. The results of simulation are presented in the table(1).

Table(1): Comparison of AIC with M$_1$IC by using from Monte -Carlo simulation for linear regression models, f_1, f_2, g_1 and g_2.

Size	Model	AIC	M$_1$IC	ΔAIC	Δ M$_1$IC
n=50	f_1	-3379	-2987	-	-
	f_2	-3378	-2888	1	99
	g_1	240	632	3619	3619
	g_2	237	727	3616	3714
n=100	f_1	-6853	-6061	-	-
	f_2	-6851	-5861	2	200
	g_1	488	1279	7341	7340
	g_2	489	1479	7342	7540
n=150	f_1	-10116	-8924	-	-
	f_2	-10114	-8624	2	300
	g_1	720	1912	10836	10836
	g_2	722	2212	10838	11136
n=200	f_1	-13677	-12085	-	-
	f_2	-13675	-11685	2	400
	g_1	979	2571	14656	14656
	g_2	979	2969	14656	15054
n=350	f_1	-24218	-21426	-	-
	f_2	-24213	-20723	5	703
	g_1	1735	4527	25953	25953
	g_2	1736	5226	25954	26652
n=500	f_1	-33735	-29743	-	-
	f_2	-33732	-28742	3	1001
	g_1	2390	6382	36125	36125
	g_2	2392	7382	36127	37125

In the third and fourth columns of table (1), the value of AIC and M$_1$IC are presented in order to various values of n and for offered models, f_1, f_2, g_1 and g_2. Therefore the relation between values AIC for offering models is obvious as AIC $(f_1)<$ AIC$(f_2)<$ AIC$(g_1)<$ AIC(g_2).

Since family F_{θ_k} is well specified and family G_{β_d}, mis specified. Thus this the relation is logical. With attention to the fourth column of, table(1) recent relation also is confirmed for M$_1$IC. In other word

M$_1$IC$(f_1)<$ M$_1$IC$(f_2)<$ M$_1$IC$(g_1)<$ M$_1$IC(g_2).

With increasing n, the value of M$_1$IC has been increased for the offered models, but the direction is confirmed unequally. In the fifth and sixth columns in order the absolute magnitude difference of the value AIC and M$_1$IC between the model of f_1 and any which from other models have presented to confirm for any n. The absolute magnitude differences have been shown by the symbols of ΔAIC and ΔM$_1$IC. If there are symbols, as

$$\Delta\mathrm{AIC}_{|f_1-f_2|}=|\mathrm{AIC}(f_1)-\mathrm{AIC}(f_2)| \quad \text{and} \quad \Delta\mathrm{AIC}_{|f_1-g_j|}=|\mathrm{AIC}(f_1)-\mathrm{AIC}(g_j)|,\quad j=1,2$$

$$\Delta\mathrm{M_1IC}_{|f_1-f_2|}=|\mathrm{M_1IC}(f_1)-\mathrm{M_1IC}(f_2)| \quad \text{and} \quad \Delta\mathrm{M_1IC}_{|f_1-g_j|}=|\mathrm{M_1IC}(f_1)-\mathrm{M_1IC}(g_j)|,\quad j=1,2$$

for n=50, 100, 150, 200, 350, 500, and models f_1, f_2, g_1 and g_2 will be confirmed the relation as

$\Delta AIC|_{f_1-f_2|} < \Delta AIC|_{f_1-g_1|} < \Delta AIC|_{f_1-g_2|}$ and $\Delta M_1IC|_{f_1-f_2|} < \Delta M_1IC|_{f_1-g_1|} < \Delta M_1IC|_{f_1-g_2|}$.

With attention to these relations the direction of similarity the model selection for information criteria AIC and M_1IC for various n have been shown With this the quality that the criterion M_1IC is a consistent information criterion.

Table (2): comparison of AIC with M_1IC by using Monte -Carlo simulation, for the state that generate model data is Normal standard and offered models are from a Laplace family with different parameters.

Size	Model	AIC	M_1IC	ΔAIC	ΔM_1IC
n=50	f_1=lap(0,1.3)	-211	-15	-	-
	f_2=lap(0,1)	-190	6	21	21
	f_3=lap(2,1)	-87	109	124	124
	f_4=lap(-2,1)	-75	121	136	136
n=100	f_1=lap(0,1.3)	-413	-17	-	-
	f_2=lap(0,1)	-367	29	45	45
	f_3=lap(2,1)	-154	242	259	259
	f_4=lap(-2,1)	-149	247	263	263
n=150	f_1=lap(0,1.3)	-628	-32	-	-
	f_2=lap(0,1)	-554	42	74	74
	f_3=lap(2,1)	-355	241	273	273
	f_4=lap(-2,1)	-206	390	422	422
n=200	f_1=lap(0,1.3)	-818	-22	-	-
	f_2=lap(0,1)	-737	58	80	80
	f_3=lap(2,1)	-334	462	484	484
	f_4=lap(-2,1)	-330	466	488	488
n=350	f_1=lap(0,1.3)	-1401	-5	-	-
	f_2=lap(0,1)	-1267	129	134	134
	f_3=lap(2,1)	-582	814	819	819
	f_4=lap(-2,1)	-560	836	841	841
n=500	f_1=lap(0,1.3)	-2061	-65	-	-
	f_2=lap(0,1)	-1880	116	181	181
	f_3=lap(2,1)	-974	1022	1087	1087
	f_4=lap(-2,1)	-880	1116	1181	1181

In the third and fourth columns of table (2) values of AIC and M_1IC for n=50, 100, 150, 200, 350 and 500, have been respectively considered Laplace offered models f_1, f_2, f_3 and f_4. Therefore the relation between values AIC for offered models of laplace family is obvious as $AIC(f_1) < AIC(f_2) < AIC(f_3) < AIC(f_4)$.

With attention to the fourth column in the table (2), the recent relation is also confirmed for M_1IC. In other word

$M_1IC(f_1) < M_1IC(f_2) < M_1IC(f_3) < M_1IC(f_4)$.

In the fifth and sixth columns the absolute magnitude difference have been presented respectively for the v alue AIC and M_1IC between the model of f_1 and any which from other models to confirm with any n, symbols of ΔAIC and ΔM_1IC has been shown. With attention to these two columns for n's different have ΔAIC = ΔM_1IC. If we have these symbols, as

$\Delta AIC|_{f_i-f_j|} = |AIC(f_i) - AIC(f_j)| \quad i \neq j$ and $\Delta M_1IC|_{f_i-f_j|} = |M_1IC(f_i) - M_1IC(f_j)| \quad i \neq j$

for any n= 50,100, 150, 200, 350, 500, and models f_1, f_2, f_3 and f_4 Confirms the relation as

$\Delta AIC|_{f_1-f_2|} < \Delta AIC|_{f_1-f_3|} < \Delta AIC|_{f_1-f_4|}$ and $\Delta M_1IC|_{f_1-f_2|} < \Delta M_1IC|_{f_1-f_3|} < \Delta M_1IC|_{f_1-f_4|}$.

With attention to these relations the direction of similarity model selection for information criteria AIC and M_1IC for various n has been shown. With the quality that the criterion M_1IC is the consistent information criterion.

Table (3): comparison of AIC with M_1IC by using Monte -Carlo simulation for the state that generate model data is Normal standard and offered models are from a Cauchy family with different parameters.

Size	Model	AIC	M_1IC	ΔAIC	ΔM_1IC
n=50	g_1=cuch(0,0.8)	-205	-9	-	-
	g_2=cuch(0,0.7)	-176	20	29	29
	g_3=cuch(-2,0.5)	-111	85	94	94
	g_4=cuch(2,1)	-75	121	130	130
n=100	g_1=cuch(0,0.8)	-410	-14	-	-
	g_2=cuch(0,0.7)	-353	43	57	57
	g_3=cuch(-2,0.5)	-157	239	253	253
	g_4=cuch(2,1)	-153	243	257	257
n=150	g_1=cuch(0,0.8)	-627	-31	-	-
	g_2=cuch(0,0.7)	-544	52	83	87
	g_3=cuch(-2,0.5)	-280	316	347	347
	g_4=cuch(2,1)	-239	357	388	388
n=200	g_1=cuch(0,0.8)	-826	-30	-	-
	g_2=cuch(0,0.7)	-718	78	108	108
	g_3=cuch(-2,0.5)	-280	315	469	469
	g_4=cuch(2,1)	-332	464	494	494
n=350	g_1=cuch(0,0.8)	-1459	-63	-	-
	g_2=cuch(0,0.7)	-1258	-138	201	201
	g_3=cuch(-2,0.5)	-600	796	859	859
	g_4=cuch(2,1)	-578	818	881	881
n=500	g_1=cuch(0,0.8)	-2115	-199	-	-
	g_2=cuch(0,0.7)	-1840	156	275	275
	g_3=cuch(-2,0.5)	-955	1041	1160	1160
	g_4=cuch(2,1)	-877	1119	1238	1238

In the third and fourth columns of table (3) values of AIC and M_1IC for n=50, 100, 150, 200, 350 and 500, have been considered respectively for Cauchy offered models g_1, g_2, g_3 and g_4. Therefore the relation between values of AIC for offered models of Cauchy family is obvious as

$$M_1IC(g_1) < M_1IC(g_2) < M_1IC(g_3) < M_1IC(g_4).$$

In the fifth and sixth columns the absolute magnitude difference has been presented respectively for the value AIC and M_1IC between the model of g_1 and any which from other models to confirm with any n has been shown with symbols of ΔAIC and ΔM_1IC. With attention to these two columns for n's different have $\Delta AIC = \Delta M_1IC$. If there are symbols, as

$$\Delta AIC_{|g_i-g_j|} = |AIC(g_i) - AIC(g_j)| \quad i \neq j \quad \text{and} \quad M_1IC_{|g_i-g_j|} = |M_1IC(g_i) - M_1IC(g_j)| \quad i \neq j$$

for any n=50,100, 150, 200, 350, 500, and models g_1, g_2, g_3 and g_4 Confirms the relation as

$$\Delta AIC_{|g_1-g_2|} < \Delta AIC_{|g_1-g_3|} < \Delta AIC_{|g_1-g_4|} \quad \text{and} \quad \Delta M_1IC_{|g_1-g_2|} < \Delta M_1IC_{|g_1-g_3|} < \Delta M_1IC_{|g_1-g_4|}.$$

The total concepts in table (2) are confirmed for table (3).

Discussion and results

In this article with investigation of the inconsistent information criterion AIC, and by eliminate of inconsistency problem a method for achieving an information criterion, has been presented based on Kullback-Leibler risk and The consistent information criterion M_1IC has been obtained. Therefore this information criterion is the only consistent information criterion and asymptotically unbiased. Which is obtained based on Kullback-Leibler risk. In section (4), by using from simulation for linear regression and classic models, the quality of model selection has been shown throughout the two information criterion, AIC an d

M_1IC. According to consistent information criterion of M_1IC, it is possible for further discussion refine the other information criteria which are according to Kullback-Leibler risk (as AICc and KICc) and add the consistency feature to the criteria.

References

Akaike, H.,1973. Information theory and an extension of maximum likelihood principle, in second intention symposium on information theory 267-281.

Bozdogan , H., 2000. Akaike's information criterion and recent developments in information complexity, j. Math. Psych. 44, 62-91. Mr1770002

Cawley, G. Talbot, N., 2010. On over-fitting model selection and subsequent selection bias in performance evaluation. Journal of machine learning research 11, 2079- 2107.

Choi, H. Kiffer. N. M., 2006. Differential geometry and bias correction in nonnested hypothesis testing.

Ghahramani, M. Nazari, K. 2013. One consistent information criterion for the model selection, elixir statistics 54, 12490-12493. Hall, P. 1987. On Kullbake-Leibler loss and density estimation. 15, 4, 1491-1519.

Hu, B. Shao, J., 2008. Generalized linear model selection using r, journal of statistical planning and information 138, 3705-3712. Konishi, S. and Kitagawa, G. 1996. Generalized information criteria in model selection. Biometrika 83, 875-890.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Statist. 22, 1,76-86.

Takeuchi, K., 1976. Distribution of information statistics and criteria for adequacy of models, mathematical sciences. 153, 12-18 (in japan).