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

Statistics

Elixir Statistics 54 (2013) 12490-12493

ISSN: 2229-712X

One consistent information criterion for the model selection

Masomeh Ghahramani[1] and Kamran Nazari[2]

[1]Department of Mathematical Statistics, Payam Noor University, Iran.
[2]Department of Business Management, Payam Noor University, Iran.

ARTICLE INFO

Article history:
Received: 2 November 2012;
Received in revised form:
27 December 2012;
Accepted: 10 January 2013;

Keywords
Model selection,
Kullback-leibler risk symmetric,
Consistent,
Information criterion KIC.

ABSTRACT
Information criterion, KIC, for the model selection based on kullback-leibler risk symmetric by cavanagh, be present for large sample. KIC as an asymptotically unbiased estimator kullback-leibler risk symmetric, is consider divergency between the true model and the candidate model. It is an inconsistent information criterion. All the criteria that exsist based on kullback-leibler risk (symmetric or asymmetric)are inconsistent. In this article based on information criterion KIC is defined consistent information criterion, MIC. This criterion, MIC also is made based on kullbake-leibler symmetric. At the end these two information criteria for linear regression models has been consider by simulation monte carlo.

1- Introduction

One of the most important problems confronting an investigator in statistical modeling is the choice of an appropriate model to characterize the underlying data. This determination can often be facilitated by the use of a model selection criterion, which judges the propriety of a fitted model by assessing whether it offers an optimal balance between goodness of fit and parsimony. the first model selection criterion to gain widespread acceptance was the Akaike (1973) information criterion, AIC. Many other criteria have been subsequently introduced and including well-known measures by Mallows (1973), Schwarz (1978), Akaike (1978), and Hurvich and Tsai (1989). AIC serves as an asymptotically unbiased estimator of a variant of Kullback's (1968) directed divergence between the true model and a fitted approximating model. The directed divergence, also known as the Kullback-Leibler (Kullback and Leibler, 1951) information, the I -divergence, or the relative entropy, assesses the dissimilarity between two statistical models. It is an asymmetric measure, meaning that an alternate directed divergence may be obtained by reversing the roles of the two models in the defnition of the measure.

The sum of the two directed divergences is Kullback's symmetric divergence, also known as the J -divergence. Since the symmetric divergence combines the information in two related though distinct measures, it functions as a gauge of model disparity which is arguably more sensitive than either of its individual components. With this motivation, we propose a model selection criterion, similalr to KIC, which serves as an asymptotically unbiased estimator, but to property of consistency. In Section 2, we present a short discussion about concept necessity. In Section 3, we discuss Kullback's symmetric divergence and is defined information criterion consistent, MIC. In Section 4, we present the results of our simulation study.

2-Concept of Necessity

Let $\underline{X}=(X_1,X_2,\dots,X_n)$ is a (i.i.d) random sample from true model and unknown, $h(.)$, and family $F_{\underline{\theta}_k}=\{f(.;\underline{\theta}_k)=f_{\underline{\theta}_k};\underline{\theta}_k\epsilon\Theta\subseteq R^k\}$ from candidate models has been consider for approximate truemodel.

Definition 1). family $F_{\underline{\theta}_k}$ is well-specified, if there is a $\underline{\theta}_0\epsilon\Theta$ such that $h(.)=f(.;\underline{\theta}_0)$; otherwise it is mis-specifid.

Definition 2). the general form of the information criterion that has been showed by IC, as

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

$=-2\sum_{i=1}^{n}log\ f(X_i\ ;\hat{\underline{\theta}}_{k(n)})+2\{$ bias estimator$\}$

$=-2\ l_f(\hat{\underline{\theta}}_{k(n)})$ + 2{ bias estimator}

So that $\hat{\underline{\theta}}_{k(n)}$, the maximum likelihood estimator $\underline{\theta}_k$ and $f(.\,;\hat{\underline{\theta}}_{k(n)})$, is the maximum likelihood function. (see, Akaike 1973).

3- Consistent information criterion, MIC

In the kullback-leibler risk asymmetric by reversing the roles of the two models be obtained, other the information criterion. the sum of the two directed divergences is kullback symmetric divergence. so that symmetric divergence was first introduced by jeffreys (see jeffreys 1946). Cavanagh (1994) information criterion KIC=$-2\sum_{i=1}^{n}log\ f(X_i\ ;\hat{\underline{\theta}}_{k(n)})+3k$ based on kullback-leibler risk asymmetric be present for the model selection in the large sample. Thus to increase number parameters in the candidate model the penalty term, 3k, be increase and term $-2log\,f(\underline{X}\,;\hat{\underline{\theta}}_{k(n)})$ be decrease. penalty term is constant to chance of size sample in the information criterion KIC, and by increase of size sample, cavanagh information criterion isn't distinguish the true model with probability one. Therefore this the problem is same concept of inconsistency for one information criterion. The follow for inconsistency information criterion, KIC, one consistent information criterion by name, MIC, is present based on kullback-leibler risk symmetric.

Tele:
E-mail addresses: kamrann0156@yahoo.com

Let $\underline{X}=(X_1,X_2,\dots,X_n)$ is a (i.i.d) random sample from has been generated according to an true and unknown parametric model $f(.;\underline{\theta}_0)$. We want to find a fitted parametric model which provides a suitable approximation to $f(.;\underline{\theta}_0)$. We present, consistent information criterion consistent CKIC, regard to definition Cavanagh for divergence between two arbitrary parametric densities $f(.;\underline{\theta})$ and $f(.;\underline{\theta}^*)$ with respect to $f(.;\underline{\theta})$. For two arbitrary parametric densities $f(.;\underline{\theta})$ and $f(.;\underline{\theta}^*)$, kullback's directed divergence between $f(.;\underline{\theta})$ and $f(.;\underline{\theta}^*)$ with respect to $f(.;\underline{\theta})$ is defined as

$$I(\underline{\theta};\underline{\theta}^*) = E_{\underline{\theta}}\left(\log\frac{f(.;\underline{\theta})}{f(.;\underline{\theta}^*)}\right). \quad (1)$$

Where $E_{\underline{\theta}}$ denotes the expectation under $f(.;\underline{\theta})$. thus $I(\underline{\theta}_0\,;\,\underline{\theta}_k)$ defines the directed divergence between $f(.;\underline{\theta}_0)$ and $f(.;\underline{\theta}_k)$ with respect to $f(.;\underline{\theta}_0)$. it is well known that $I(\underline{\theta}_0\,;\,\underline{\theta}_k)\geq 0$ with equality if and only if $\underline{\theta}_k=\underline{\theta}_0$ (see kullback 1968). Now for $f(.;\underline{\theta})$ and $f(.;\underline{\theta}_0)$, is defined as

$$d(\underline{\theta},\underline{\theta}_0)= E_{\underline{\theta}}(-2\log f(\underline{X};\underline{\theta}_0)). \quad (2)$$

From (1) and (2), note that we can write

$$2\,I(\underline{\theta}_0\,;\,\underline{\theta}_k)= d(\underline{\theta}_0,\underline{\theta}_k) - d(\underline{\theta}_0,\underline{\theta}_0). \quad (3)$$

Kullback's symmetric divergence is then defined as

$$J(\underline{\theta}_0\,;\,\underline{\theta}_k)= I(\underline{\theta}_0,\underline{\theta}_k) - I(\underline{\theta}_k,\underline{\theta}_0).$$

Note that $J(\underline{\theta}_0\,;\,\underline{\theta}_k)= J(\underline{\theta}_k\,;\,\underline{\theta}_0)$, whereas $I(\underline{\theta}_0\,;\,\underline{\theta}_k)\neq I(\underline{\theta}_k\,;\,\underline{\theta}_0)$ unless $\underline{\theta}_k=\underline{\theta}_0$; thus $J(\underline{\theta}_0\,;\,\underline{\theta}_k)$ is symmetric in its arguments whereas $I(\underline{\theta}_0\,;\,\underline{\theta}_k)$ is not.

using(1),(2), and (3), can be written

$$2J(\underline{\theta}_0\,;\,\underline{\theta}_k)=\{d(\underline{\theta}_0,\underline{\theta}_k) - d(\underline{\theta}_0,\underline{\theta}_0)\}+\{d(\underline{\theta}_k,\underline{\theta}_0) - d(\underline{\theta}_k,\underline{\theta}_k)\}. \quad (4)$$

Since $d(\underline{\theta}_0,\underline{\theta}_0)$ dose not depend on $\underline{\theta}_k$, thus taking into consideration $K(\underline{\theta}_0\,;\,\underline{\theta}_k)$, instead relation (4), as

$$K(\underline{\theta}_0\,;\,\underline{\theta}_k)= d(\underline{\theta}_0,\underline{\theta}_k) +\{d(\underline{\theta}_k,\underline{\theta}_0) - d(\underline{\theta}_k,\underline{\theta}_k)\}. \quad (5)$$

measures such as $K(\underline{\theta}_0\,;\,\underline{\theta}_k)$, $J(\underline{\theta}_0\,;\,\underline{\theta}_k)$, $d(\underline{\theta}_k\,;\,\underline{\theta}_0)$, and $I(\underline{\theta}_0\,;\,\underline{\theta}_k)$ are often called discrepancies(see linhart and zucchini 1986). Now evaluating (5) at maximum likelihood estimator $\underline{\theta}_k$, means $\hat{\underline{\theta}}_{k(n)}$, thus we have

$$K(\underline{\theta}_0\,;\,\hat{\underline{\theta}}_{k(n)})= d(\underline{\theta}_0,\hat{\underline{\theta}}_{k(n)}) +\{d(\hat{\underline{\theta}}_{k(n)},\underline{\theta}_0) - d(\hat{\underline{\theta}}_{k(n)},\hat{\underline{\theta}}_{k(n)})\}. \quad (6)$$

For clarity, we also have

$$d(\hat{\underline{\theta}}_{k(n)},\underline{\theta}_0)=E_{\underline{\theta}_k}\{-2\log f(\underline{X};\underline{\theta}_0)\}|_{\underline{\theta}_k=\hat{\underline{\theta}}_{k(n)}},$$

$$d(\underline{\theta}_0,\hat{\underline{\theta}}_{k(n)})=E_{\underline{\theta}_0}\{-2\log f(\underline{X};\underline{\theta}_k)\}|_{\underline{\theta}_k=\hat{\underline{\theta}}_{k(n)}},$$

$$d(\hat{\underline{\theta}}_{k(n)},\hat{\underline{\theta}}_{k(n)})=E_{\underline{\theta}_k}\{-2\log f(\underline{X};\underline{\theta}_k)\}|_{\underline{\theta}_k=\hat{\underline{\theta}}_{k(n)}}.$$

Although $K(\underline{\theta}_0\,;\,\hat{\underline{\theta}}_{k(n)})$ is inaccessible, one might speculate that it is possible to construct an asymptotically unbiased estimator with an expected value that asymptotically approaches the expected value of $K(\underline{\theta}_0\,;\,\hat{\underline{\theta}}_{k(n)})$, say

$$W(\underline{\theta}_0,k) = E_{\underline{\theta}_0}\{K(\underline{\theta}_0\,;\,\hat{\underline{\theta}}_{k(n)})\}$$

$$= E_{\underline{\theta}_0}\{d(\underline{\theta}_0,\hat{\underline{\theta}}_{k(n)}) +\{d(\hat{\underline{\theta}}_{k(n)},\underline{\theta}_0) - d(\hat{\underline{\theta}}_{k(n)},\hat{\underline{\theta}}_{k(n)})\}\},$$

to adding and subtract, statements $-2\log f(\underline{X};\hat{\underline{\theta}}_{k(n)})$ and $d(\underline{\theta}_0,\underline{\theta}_0)$ to $K(\underline{\theta}_0\,;\,\hat{\underline{\theta}}_{k(n)})$, in the relation (6), and have $W(\underline{\theta}_0,k)$, we can write $W(\underline{\theta}_0,k)$ as

$$W(\underline{\theta}_0,k) = E_{\underline{\theta}_0}\{-2\log f(\underline{X};\hat{\underline{\theta}}_{k(n)})\}$$

$$= [\,d(\underline{\theta}_0,\underline{\theta}_0)-E_{\underline{\theta}_0}\{-2\log f(\underline{X};\hat{\underline{\theta}}_{k(n)})\}\,] + [E_{\underline{\theta}_0}\{d(\underline{\theta}_0,\hat{\underline{\theta}}_{k(n)}) - d(\underline{\theta}_0,\underline{\theta}_0)\}]$$

$$+ E_{\underline{\theta}_0}\{d(\hat{\underline{\theta}}_{k(n)},\underline{\theta}_0)-d(\hat{\underline{\theta}}_{k(n)},\hat{\underline{\theta}}_{k(n)})\} = E_{\underline{\theta}_0}\{-2\log f(\underline{X};\hat{\underline{\theta}}_{k(n)})\} + W_1 + W_2 + W_3.$$

following, be shown, W_1 converge to nk and W_2 and W_3 converge to k.

Now if $\{-\frac{\partial^2 \log f(\underline{X};\underline{\theta}_k)}{\partial\underline{\theta}_k\,\partial\underline{\theta}_k^T}\}$, be shown with $\mathcal{L}(\underline{\theta}_k)$ and taylor's expansion $-2\log f(\underline{X};\underline{\theta}_0)$ around the maximum likelihood estimator, $\hat{\underline{\theta}}_{k(n)}$, is given by

$$-2\log f(\underline{X};\underline{\theta}_0) = -2\log f(\underline{X};\hat{\underline{\theta}}_{k(n)}) + (\hat{\underline{\theta}}_{k(n)}-\underline{\theta}_0)^T\,\mathcal{L}(\hat{\underline{\theta}}_{k(n)})\,(\hat{\underline{\theta}}_{k(n)}-\underline{\theta}_0)+o_p(1), \quad (7)$$

In relation (7), $\frac{1}{n}\mathcal{L}(\underline{\theta}_k)|_{\underline{\theta}_k=\hat{\underline{\theta}}_{k(n)}}$ converge to $J(\underline{\theta}_0)$. So that $J(\underline{\theta}_0)=-E_{\underline{\theta}_0}[\frac{\partial^2 \log f(\underline{X};\underline{\theta}_k)}{\partial\underline{\theta}_k\,\partial\underline{\theta}_k^T}]|_{\underline{\theta}_k=\underline{\theta}_0}$.

Therefore, relation (7) can be approximate, as

$$-2\log f(\underline{X};\underline{\theta}_0) -(-2\log f(\underline{X};\hat{\underline{\theta}}_{k(n)})) \approx n(\hat{\underline{\theta}}_{k(n)}-\underline{\theta}_0)^T\,J(\underline{\theta}_0)\,(\hat{\underline{\theta}}_{k(n)}-\underline{\theta}_0)+o_p(1), \quad (8)$$

taking the taylor's expansion $d(\underline{\theta}_0,\hat{\underline{\theta}}_{k(n)})$ around $\underline{\theta}_0$, and taylor's expansion $d(\hat{\underline{\theta}}_{k(n)},\underline{\theta}_0)$ around $\hat{\underline{\theta}}_{k(n)}$, is given by

$$d(\underline{\theta}_0,\hat{\underline{\theta}}_{k(n)}) = d(\underline{\theta}_0,\underline{\theta}_0)+(\hat{\underline{\theta}}_{k(n)}-\underline{\theta}_0)^T\,J(\underline{\theta}_0)\,(\hat{\underline{\theta}}_{k(n)}-\underline{\theta}_0)+o_p(1), \quad (9)$$

$$d(\hat{\underline{\theta}}_{k(n)},\underline{\theta}_0) = d(\hat{\underline{\theta}}_{k(n)},\hat{\underline{\theta}}_{k(n)})+(\hat{\underline{\theta}}_{k(n)}-\underline{\theta}_0)^T\,J(\hat{\underline{\theta}}_{k(n)})(\hat{\underline{\theta}}_{k(n)}-\underline{\theta}_0)+o_p(1). \quad (10)$$

Since the discrepancies $d(\underline{\theta}_0,\hat{\underline{\theta}}_{k(n)})$ and $d(\hat{\underline{\theta}}_{k(n)},\underline{\theta}_0)$ are respectively minimized at $\underline{\theta}_0$ and $\hat{\underline{\theta}}_{k(n)}$.

Taking the expectation of mathematical with respect to $\underline{\theta}_0$ from relations (8), (9), and (10), are given by

$W_1 = d(\underline{\theta}_0,\underline{\theta}_0)-$

Table (1):camparison, KIC with MIC by use from simulation monte carlo for linear regression models, f_1, f_2, g_1 and g_2.

Size sample	model	KIC	MIC	ΔKIC	ΔMIC
n=50	f_1 f_2 g_1 g_2	-3440 -3437 256 268	-3244 -3192 461 513	- 3 3705 3708	- 52 3705 3757
n=100	f_1 f_2 g_1 g_2	-6672 -6668 497 500	-6276 -6173 893 995	- 4 7169 7172	- 102 7169 7271
n=150	f_1 f_2 g_1 g_2	-10117 -10114 712 713	-9521 -9369 1308 1458	- 3 10829 10831	- 152 10829 10980
n=200	f_1 f_2 g_1 g_2	-13673 -13670 982 984	-12877 -12675 1779 1979	- 3 14655 14657	- 202 14656 14856
n=350	f_1 f_2 g_1 g_2	-23676 -23673 1692 1695	-22280 -21928 3088 3440	- 3 25368 25371	- 352 25369 25720
n=500	f_1 f_2 g_1 g_2	-33731 -33727 2394 2397	-31735 -31232 4390 4892	- 4 36125 36128	- 503 36125 36627

$$E_{\underline{\theta}_0}\{-2\log f(\underline{X};\hat{\underline{\theta}}_{k(n)})\} \approx nE_{\underline{\theta}_0}\{(\hat{\underline{\theta}}_{k(n)} - \underline{\theta}_0)^T \, J(\underline{\theta}_0)\,(\hat{\underline{\theta}}_{k(n)} - \underline{\theta}_0)\}, \quad (11)$$

$$W_2 = E_{\underline{\theta}_0}\{d(\underline{\theta}_0, \hat{\underline{\theta}}_{k(n)}) - d(\underline{\theta}_0, \underline{\theta}_0)\} = E_{\underline{\theta}_0}\{(\hat{\underline{\theta}}_{k(n)} - \underline{\theta}_0)^T \, J(\underline{\theta}_0)(\hat{\underline{\theta}}_{k(n)} - \underline{\theta}_0)\}, \quad (12)$$

$$W_3 = E_{\underline{\theta}_0}\{ d(\hat{\underline{\theta}}_{k(n)}, \underline{\theta}_0) - d(\hat{\underline{\theta}}_{k(n)}, \hat{\underline{\theta}}_{k(n)})\} = E_{\underline{\theta}_0}\{(\hat{\underline{\theta}}_{k(n)} - \underline{\theta}_0)^T \, J(\hat{\underline{\theta}}_{k(n)})(\hat{\underline{\theta}}_{k(n)} - \underline{\theta}_0)\}. \quad (13)$$

Now the quadratic forms in relations (11), (12), and (13), all converge to centrally distributed chi-square with k degrees of freedom. Thus, the expectations (under $\underline{\theta}_0$) of each of these quadratic forms is k. Therefore, is given by W_1=nk and W_2=W_3=k.

There is kullback-leibler risk symmetric, in the case information criterion MIC, as

$$MIC=-2\sum_{i=1}^{n} log\ f(X_i\ ;\hat{\underline{\theta}}_{k(n)})+k(n+2)$$

$$=-2\ l_f(\hat{\underline{\theta}}_{k(n)}) + k(n+2). \quad (14)$$

In offered information criterion MIC, penalty term k(n+2) is change, with change size sample. so if size sample is going very large, information criterion MIC, with probability one, to find the true model data. In other words information criterion MIC, is alone consistent information criterion, that has been obtained based on kullback-leibler risk symmetric.

For to show consistency, information criterion MIC, let the maximum likelihood function estimator for the candidate model $(f(.\,;\underline{\theta}_k) = f(\underline{\theta}_k))$ and optimal model $(f(.\,;\underline{\theta}_{k_0}) = f(\underline{\theta}_{k_0}))$ with respectively $l_f(\hat{\underline{\theta}}_{k(n)})$ and $l_f(\hat{\underline{\theta}}_{k_0(n)})$. Regard to in relation (14) information criterion MIC, for the model $f(\underline{\theta}_k)$ and $f(\underline{\theta}_{k_0})$ are given by

$$MIC\ (f(\underline{\theta}_k)) = -2l_f(\hat{\underline{\theta}}_{k(n)})+k(n+2),$$

$$MIC\ (f(\underline{\theta}_{k_0})) = -2l_f(\hat{\underline{\theta}}_{k_0(n)})+ k_0\ (n+2).$$

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

$$P(MIC\ f(\underline{\theta}_k) - MIC\ f(\underline{\theta}_{k_0}) > 0)$$

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

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

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

(15)

In relation (15), U_n is $2l_f(\hat{\underline{\theta}}_{k(n)}) - 2l_f(\hat{\underline{\theta}}_{k_0(n)})$ and distribution function of chi-square has been shown with F. So that it is tend in the probability to one. Thus MIC is one consistent information criterion.(for further study about the consistency one information criterion, see Hu and Shao2008).

4) Simulation study

In this simulation to be accomplished for usage and camparison the offered information criterion, MIC, with the information criterion KIC, by use from simulation monte carlo, for linear regression models. In this simulation be supposed that well-specifid family $F_{\underline{\theta}_k}=\{f(.;\underline{\theta}_k) = f_{\underline{\theta}_k};\ \underline{\theta}_k \epsilon \Theta \subseteq R^k\}$, and mis-specified family $G_{\underline{\beta}_d}=\{g(.;\underline{\beta}_d) = g_{\underline{\beta}_d};\ \underline{\beta}_d \epsilon \mathrm{B} \subseteq R^d\}$ 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 generate as random from distribution N(0,2). Models

$f_1: y_i=\hat{\theta}_0+\hat{\theta}_1 x_{i1}+\hat{\theta}_2 x_{i2}+\hat{\theta}_3 x_{i3}$ i=1,…,n

and

$f_2: y_i = \hat{\theta}_0 + \hat{\theta}_1 x_{i1} + \hat{\theta}_2 x_{i2} + \hat{\theta}_3 x_{i3} + \hat{\theta}_4 x_{i4} \quad i = 1, \dots, n$

candidate models, is generate from $F_{\underline{\theta}_k}$.

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 generate as random from distribution N(0,1), and Models $g_1: y_i=\hat{\beta}_0+\hat{\beta}_1 z_{i1}+\hat{\beta}_2 z_{i2}+\hat{\beta}_3 z_{i3}$ i=1,…,n

and

$g_2: y_i = \hat{\beta}_0 + \hat{\beta}_1 z_{i1} + \hat{\beta}_2 z_{i2} + \hat{\beta}_3 z_{i3} + \hat{\beta}_4 z_{i4} \quad i = 1, \dots, n$

candidate models, is generate from $G_{\underline{\beta}_d}$. this the simulation be achieved by use from software R, and number repeat is r $=10^4$, and samples n=50, 100, 150, 200, 350, 500, has been consider. the results of simulation are present in the table(1).

In the columns, third and fourth, in the table (1), are present value KIC and MIC for various values of n and for candidate models, f_1, f_2, g_1 and g_2. So that, obviously the relation between values KIC for candidate models, as

$\mathrm{KIC}(f_1) < \mathrm{KIC}(f_2) < \mathrm{KIC}(g_1) < \mathrm{KIC}(g_2)$.

Since family $F_{\underline{\theta}_k}$ is well- specified and family $G_{\underline{\beta}_d}$, mis-specified. Thus this the relation is logical.

Attention to the column of fourth, table(1) recent relation also is confirmed for MIC. In other word

$\mathrm{MIC}(f_1) < \mathrm{MIC}(f_2) < \mathrm{MIC}(g_1) < \mathrm{MIC}(g_2)$.

With be large n, has been large value MIC for the candidate models, but is confirmed direction unequally. The columns, fifth and sixth in order are absolute magnitude difference the value KIC and MIC between the model of f_1 and any which from other the models to conform with any n. symbols of $\Delta\mathrm{KIC}$ and $\Delta\mathrm{MIC}$ has been shown. If there is symbols, as

$\Delta\mathrm{KIC}_{|f_1-f_2|} = |\mathrm{KIC}(f_1) - \mathrm{KIC}(f_2)|$,

$\Delta\mathrm{KIC}_{|f_1-g_j|} = |\mathrm{KIC}(f_1) - \mathrm{KIC}(g_j)|$, j=1,2

$\Delta MIC_{|f_1-f_2|} = |\mathrm{MIC}(f_1) - MIC(f_2)|$,

$\Delta MIC_{|f_1-g_j|} = |\mathrm{MIC}(f_1) - MIC(g_j)|$, j=1,2

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

$\Delta\mathrm{KIC}_{|f_1-f_2|} < \Delta\mathrm{KIC}_{|f_1-g_1|} < \Delta\mathrm{KIC}_{|f_1-g_2|}$,

$\Delta MIC_{|f_1-f_2|} < \Delta MIC_{|f_1-g_1|} < \Delta MIC_{|f_1-g_2|}$.

Attention to these relations has been shown similarity direction the model selection for information criteria KIC and MIC for various n. With the this quality that criterion MIC, is the consistent information criterion.

Discuss and results

In this article to consider the inconsistent information criterion KIC, and to eliminate the problem of inconsistency, is present the method for to earn one information criterion, that is based on kullback-leibler risk symmetric. It is earning the consistent information criterion MIC. So that hitherto this the information criterion is alone consistent information criterion and asymptotically unbiased. It is earning based on kullback-leibler risk symmetric. In section (4) has been shown, by simulation for linear regression models, quality the model selection, through two the information criterion, KIC and MIC. Further work on other the model selection, may be considered, other the information criterioa, so that there is based on kullback-leibler risk (as AIC, AICc and KICc) and to add the consistency be considered for these criteria.

References

Akaike, H.,1973. Information theory and an extension of maximum likelihood principle, in second intrantion 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

Cavanagh, J. E., 1999. a large-sample model selection criterion based on kullback’s symmetric divergence, statistics and probability letters vol. 44, 333-344.

Hu, B. shao, J., 2008. generalized linear model selection using rr, journal of statistical planning and information 138, 3705-3712.

Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small samples. Biometrika 76, 297-307.

Jefreys, H., 1946. An invariant from for the prior probability in estimation problems. Proc. Roy. Soc. A 186, 453-461.

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

Linhart, H., Zucchini, W., 1986. Model Selection. Wiley, New York.

Mallows, C.L., 1973. Some comments on Cp. Technometrics 15, 661-675.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461-464.