associates $(x-1)/p$ to x is clearly well-defined on γ_p and presents interesting homomorphic properties. In particular,

$$\forall x, y \in \gamma_p \quad \log(xy \bmod p^2) = \log(x) + \log(y) \bmod p$$

whereby, as a straightforward generalisation,

$$\forall g \in \gamma_p, m \in \mathbf{Z}_p \quad \log(g^m \bmod p^2) = m \log(g) \bmod p$$

Key setup: Generate two k-bit primes p and q (typically $3k = 1023$) and set $n = p^2q$. Randomly select and publish a number $g < n$ such that

$$g_p = g^{p-1} \bmod p^2$$

is of order p in $\mathbf{Z}^*_{p^2}$ and keep g_p secret (note that $g_p \in \gamma_p$). Similarly, choose $g' < n$ at random and publish

$$h = g'^n \bmod n$$

The triple $\{n, g, h\}$ forms the public key. The secret key is $\{p, q\}$.

Encryption: Pick $r < n$ uniformly at random and encrypt the $(k-1)$-bit message m by:

$$c = g^m h^r \bmod n$$

Decryption: Proceed as follows:
(i) $c' = c^{p-1} \bmod p^2 = g^{m(p-1)} g'^{nr(p-1)} = g_p^m \bmod p^2$
(ii) $m = \log(c')\log(g_p)^{-1} \bmod p$.

We refer the reader to [1] for a thorough description of the scheme. Although provably equivalent to factoring [2] as far as chosen-plaintext attacks are concerned, the scheme suffers from the fact that ciphertexts are about three times longer than plaintexts. Note that step (i) of the decryption process requires $O(k^3)$ bit operations.

Proposed variant: As pointed out by Paillier [3], OU's trapdoor is inherently new in the sense that it profoundly differs from the RSA and Diffie-Hellman schemes. There is no doubt that this technique could be used in various ways to design new public-key cryptosystems in near future.

To reduce OU's complexity to $O(k^2)$ while preserving equivalence to factoring, we select a p such that $p-1$ has a large (160 bit) prime factor t; let $p - 1 = tu$ and modify the scheme's specifications as follows:

Randomly select a number $g < n$ such that

$$g_p = g^{p-1} \bmod p^2$$

is of order p in $\mathbf{Z}^*_{p^2}$, compute $G = g^u \bmod n$ and keep g_p secret. Similarly, choose $g' < n$ at random and publish

$$H = g'^{nu} \bmod n$$

The triple $\{n, G, H\}$ forms the public key. The secret key is $\{p, q\}$.

Encryption: Pick $r < n$ uniformly at random and encrypt the $(k-1)$-bit message m by

$$c = G^m H^r \bmod n$$

Decryption: Proceed as follows:
(i) $c' = c^t \bmod p^2 = g^{m(p-1)} g'^{nr(p-1)} = g_p^m \bmod p^2$
(ii) $m = \log(c')\log(g_p)^{-1} \bmod p$.

The cubic-complexity has thus been replaced by a quadratic-complexity (here t has a fixed size); equivalence to factoring is easily derived from the original security proof included in [1].

4 January 1999
Electronics Letters Online No: 19990229
DOI: 10.1049/el:19990229

J.-S. Coron (*École Normale Supérieure, 45 rue d'Ulm, Paris, F-75230, France*)

E-mail: coron@clipper.ens.fr

D. Naccache and P. Paillier (*Gemplus Card International, 34 rue Guynemer, Issy-les-Moulineaux, F-92447, France*)

J.-S. Coron: Also with Gemplus Card International, 34 rue Guynemer, Issy-les-Moulineaux, F-92447, France

P. Paillier: Also with École Nationale Supérieure des Télécommunications, 46 rue Barrault, Paris, F-75013, France

References

1 OKAMOTO, T., and UCHIYAMA, S.: 'A new public-key cryptosystem as secure as factoring'. Advances in Cryptology, Proc. Eurocrypt'98, Paper LNCS 1403, (Springer-Verlag, 1998,), pp. 308–318
2 OKAMOTO, E., and PERALTA, R.: 'Faster factoring of integers of a special form', *IEICE Trans. Fundamentals,* 1996, **E79-A**, (4), pp. 489–493
3 PAILLIER, P.: 'A trapdoor permutation equivalent to factoring'. Proc. PKC'99, Paper LNCS, (Springer-Verlag, 1999)

Criterion for selecting model order in identification

Soon H. Kwon and M. Sugeno

The authors propose a consistent and bias corrected extension of Akaike's information criterion (AIC), AIC_{bc}. Empirical performances of the AIC_{bc} are studied. The AIC_{bc} was found to result in relatively better model order choices of a linear regression model using a Monte Carlo experiment.

Introduction: The conventional theory of filtering and control assumes the availability of a mathematical model which adequately describes the system concerned. In system identification, the choice of an appropriate model is a fundamental difficulty. A general principle for addressing this problem, Occam's razor, states that an adequate but parsimonious model is preferable to others. Since Akaike's influential paper on AIC [1], several approaches to model selection have been developed and are still being refined [2 – 4]. When the data in the true model has infinite order, AIC provides an asymptotically efficient selection of a finite order model. However, when the data in the true model has finite order, minimising AIC does not produce consistent model order selection. This overfitting leads to more unsatisfactory model order selection when the sample size is small, or when the number of free parameters is relatively larger than the sample size. This Letter describes a consistent and bias corrected extension of AIC.

Model selection criterion: Suppose that independent random variables $X_1, ..., X_n$ form random samples $x_1, ..., x_n$ from a continuous distribution for which the probability density function is $f(\mathbf{x}|\theta)$, where the parameter vector $\theta = \theta_K = (\theta_1, \theta_2, ..., \theta_K)$ belongs to some K-dimensional parameter space Ω. Assume that a true parameter vector θ^* of θ with its probability density function $f(\mathbf{x}|\theta^*)$ is included in the Ω. A model defined by restricting the parameter space with $\theta_h = 0$ for all $h > k$ is given by MODEL(k): $f(\mathbf{x}|\theta_k)$, $\theta_k = \{(\theta_1, \theta_2, ..., \theta_k)|\theta_h = 0 \text{ for all } h > k\}$. The statistical model identification may be carried out by selecting a restricted model $f(\mathbf{x}|\theta_k)$, where the θ_k is the closest to the true parameter vector θ^*, based on the given n observations.

Proposition 1: The consistent and bias corrected extension of AIC is

$$AIC_{bc}(k) = -2\ell(\hat{\theta}_k) + k \log \frac{n}{2\pi} + \log |J(\hat{\theta}_k)| + 2\frac{nk}{n-k-2} \tag{1}$$

Proof of proposiation 1: Under the assumption that a given data set is a realisation of a random variable vector $\mathbf{X}$ in which random variables are independent and identically distributed, the mean log likelihood function for the set of data is given by

$$\ell_n(\theta) \equiv \frac{1}{n}\ell(\theta) = \frac{1}{n}\log L(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log f(x_i|\theta) \tag{2}$$

where $\ell(\theta) \equiv \log L(\theta) = \Sigma_{i=1}^n \log f(x_i|\theta)$ and $L(\theta) = f(x_1, x_2, ..., x_n|\theta) = \Pi_{i=1}^n f(x_i|\theta)$. From the efficiency of the maximum likelihood estimator, we observe that the mean log likelihood is a natural estimator of the expected log likelihood. Therefore, we have to minimise the expected mean log likelihood of the true model given by

$$\ell_n^* \equiv E\{\ell^*(\hat{\theta}_k)\}$$
$$\cong \ell(\hat{\theta}_k) - \ell(\theta_k^*) + \frac{n}{2}\|\theta^* - \theta_k^*\|_J^2$$
$$+ (\theta_k^* - \theta^*)nJ(\theta^* - \hat{\theta}_K)^T - E\left\{n\|\hat{\theta}_k - \theta_k^*\|_J^2\right\} \quad (3)$$

where $\ell^*(\theta^*) \equiv nE\{\ell_n(\theta^*)\} = E\{\ell(\theta^*)\} = E\{\Sigma_{i=1}^n \log f(x_i|\theta^*)\}$, and J is the positive definite Fisher information matrix. For sufficiently large n, we have $\hat{\theta}_k \to N(\theta_k^*, (nJ(\theta_k^*))^{-1})$. In this case, $\ell(\theta_k^*)$ of eqn. 3 can be approximated as follows [4]:

$$\ell(\theta_k^*) \cong \log h(\mathbf{x}) + \frac{k}{2}\log\frac{n}{2\pi} + \frac{1}{2}\log|J(\hat{\theta}_k)| + O(n^{-1/2}) \quad (4)$$

where $h(\mathbf{x})$ is independent of a parameter vector θ. Here, we assume that the true parameter θ^* is situated near θ_k^*. By ignoring $\log h(\mathbf{x})$ in eqn. 4 and multiplying both sides of eqn. 3 by –2, we have

$$-2\ell_n^*(k) \cong -2\ell(\hat{\theta}_k) + k\log\frac{n}{2\pi} + \log|J(\hat{\theta}_k)|$$
$$+ 2E\left\{n\|\hat{\theta}_k - \theta_k^*\|_J^2\right\} \quad (5)$$

For sufficiently large n, $\sqrt{n}(\hat{\theta}_k - \theta_k^*)$ is asymptotically multivariate normal. In this case, the last term in eqn. 5 is equal to $2\mathrm{tr}(J^{-1}R)$. We note that $\mathrm{tr}(J^{-1}R)$ is the well-known Lagrange-multiplier test statistic. Because θ_k^* is not directly observable, we use its maximum likelihood estimator $\hat{\theta}_k$. Because the computation of $\mathrm{tr}(J^{-1}R)$ is very cumbersome, it is reasonable to approximate it by a simpler form for effective use. The last term in eqn. 5, $\|\hat{\theta}_k - \theta_k^*\|_J^2$, under the expectation can be approximated as

$$\|\hat{\theta}_k - \theta_k^*\|_J^2 = \frac{\left(\frac{1}{P(\theta_k^*)}\left(\frac{\partial\ell(\theta)}{\partial\theta}\right)_{\theta_k^*}\right)^2}{(b(\theta))^2} \quad (6)$$

where

$$b(\theta) = \frac{1}{-P^2(\theta_k^*)}\left[\left\{\frac{1}{f(\mathbf{X}|\theta)}\left(\frac{\partial^2 f(\mathbf{X}|\theta)}{\partial\theta\partial\theta^T}\right)\right\}_{\theta_k^*} - \left(\frac{\partial\ell(\theta)}{\partial\theta}\right)_{\theta_k^*}^2\right]$$

and

$$P^2(\theta_k^*) = J(\theta_k^*)$$

The distribution of the numerator on the right of eqn. 6 becomes a χ^2 distribution with k degrees of freedom. The distribution of the denominator on the right of eqn. 6 also becomes a χ^2 distribution with $(n–k)$ degrees of freedom and they are independent. Thus, $[(n–k)/k]\|\theta_k^* - \hat{\theta}_k\|_J^2$ is approximately distributed as $F(k, n–k)$. Thus, taking the expectation of eqn. 6, AIC_{bc} of eqn. 1 is obtained. The proof of the consistency of AIC_{bc} is omitted for space reasons.

Numerical example: To investigate the empirical performance of the AIC_{bc}, we provide the results of a Monte Carlo study. We assume that the approximating family includes the true model, that is, the degree of the regression model k is less than or equal to a given K, which is a universally accepted assumption for such model selection problems. In our Monte Carlo study with $K = 7$, 1000 realisations for each sample size (n = 15, 100 and 500) were generated from the following linear regression model:

$$y = x_1 + 2x_2 + 3x_3 + \varepsilon \qquad \varepsilon \simeq N(0, \sigma^2 = 1) \quad (7)$$

Seven candidate variables stored in an $n \times 7$ matrix $\mathbf{X}$ of independent identically distributed normal random variables were considered. All the computations were carried out using MATLAB®. The candidate models included the columns of $\mathbf{X}$ in a sequentially nested fashion; i.e. the candidate model of dimension k consisted of columns 1, ..., k of $\mathbf{X}$. For each realisation, we studied the relative performances of AIC [1], BIC [2], AIC_c [3], CAICF [4], and AIC_{bc} given as follows:

$$AIC(k) = -2\ell(\hat{\theta}_k) + 2k \quad (8)$$

$$BIC(k) = -2\ell(\hat{\theta}_k) + k\log n \quad (9)$$

$$AIC_c(k) = -2\ell(\hat{\theta}_k) + \frac{n^2 + n(k-1)}{n-k-1} \quad (10)$$

$$CAICF(k) = -2\ell(\hat{\theta}_k) + k(\log n + 2) + \log|J(\hat{\theta}_k)| \quad (11)$$

For each sample size, the results of the Monte Carlo study are given in Table 1.

Table 1: Frequency of order selected by various criteria in 1000 realisations of regression model

Sample size	Criterion	Selected order						
		1	2	3*	4	5	6	7
n = 15	AIC	0	0	438	134	82	139	207
	AIC_c	0	0	935	51	11	2	1
	BIC	0	0	582	132	65	93	128
	CAICF	0	0	918	46	13	11	12
	AIC_{bc}	0	0	981	19	0	0	0
n = 100	AIC	0	0	749	110	52	51	38
	AIC_c	0	0	796	103	40	36	25
	BIC	0	0	967	27	4	2	0
	CAICF	0	0	989	10	1	0	0
	AIC_{bc}	0	0	980	16	3	1	0
n = 500	AIC	0	0	735	130	53	45	37
	AIC_c	0	0	744	126	52	43	35
	BIC	0	0	986	13	1	0	0
	CAICF	0	0	996	4	0	0	0
	AIC_{bc}	0	0	988	11	1	0	0

From the results for the small sample (n = 15) in Table 1, we see that AIC_{bc} provides the best selection of the correct degree of model for all criteria. The other criteria show a tendency to overfit the model. These support our belief that the fourth term in eqn. 1 penalises the overparametrisation more strongly for small samples than the other criteria other than AIC_{bc}. For large samples (n = 100 and 500), AIC_{bc}, BIC, and CAICF show consistent model order selection as we previously discussed, but AIC and AIC_c do not. From the results in Table 1, we can conclude that AIC_{bc} has relatively better performance than other criteria studied in this Letter across almost all sample sizes, and provides a consistency of order selection.

Conclusions: We have proposed a consistent and bias corrected model selection criterion (AIC_{bc}). Empirical performances of AIC_{bc} over small and large sample sizes showed relatively better order choices of a linear regression model than those of other criteria used in this experiment.

Acknowledgments: The authors thank M. Ueno for helpful discussions.

23 November 1998

Electronics Letters Online No: 19990201

DOI: 10.1049/el:19990201

Soon H. Kwon (*School of E & E, Yeungnam University, 214-1 Daedong, Kyongsan, Kyongbuk 712-749, Korea*)

M. Sugeno (*Department of Computational Intelligence, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226, Japan*)

References

1 AKAIKE, H.: 'A new look at the statistical model identification', *IEEE Trans.*, 1974, **AC-19**, pp. 716–723

2 SCHWARZ, G.: 'Estimating the dimension of a model', *Annal. Stat.*, 1978, **6**, pp. 461–464

3 HURVICH, C.M., and TSAI, C.L.: 'Regression and time series model selection in small samples', *Biometrika*, 1989, **76**, pp. 297–307

4 BOZDOGAN, H.: 'Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions', *Psychometrika*, 1987, **52**, pp. 345–370