Criterion for Selection of Model and Controller Design based on I/O Data

K. Tsumura and H. Kimura

Department of Mathematical Engineering and Information Physics, Faculty of Engineering
The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656
tsumura@crux.t.u-tokyo.ac.jp, kimura@crux.t.u-tokyo.ac.jp

Abstract

In this paper, we propose information criteria not only for model estimation and selection but also for selection of design method of controller. The criteria are derived by the same approach of AIC/TIC, however we consider the modeling of the distribution of the control performance substituting for that of system index. The criteria are composed of a log-likelihood function and a bias term similar to AIC/TIC, and especially the bias term depends on the controller. Moreover, when the controller is designed under some conditions, the criteria become stochastic variables and their expectations are further added by the other bias terms, which are in proportion to the dimension of the control parameter. This shows that model estimation, selection and control designing are influenced simultaneously by the complexity of model and that of controller.

1 Introduction

In designing controller, a model or a model set of a plant is necessary, and it is constructed or estimated from a priori information or given data. In particular, there are a long history and a lot of results on the estimation problem, and we can apply the results to the modeling of plant to be controlled. In model estimation, in general, some universal criterion, e.g. Kullback-Leibler information, is defined to evaluate a estimated model. Such universalism which treats every model equivalently, is of course in the right, however some open question arises, that is, a model for a plant to be controlled should not be evaluated apart from the attained performance of the closed loop systems which is designed from that model, because the most important purpose of control is the control performance itself. This indivisibility of modeling/estimation and designing has been discussed in several literatures, and also, we can see several results on the unification of system identification and controller designing. However unfortunately, an important point of view, that is, model selection and selection of controller design method has not been discussed enough.

Now there is a well-known result in statistical estimation; select a simple model among the comparable models for a estimated system. This result is well explained by AIC or any other criteria [1, 7, 6], and they reveal the relation between “data” and “model”. On the other hand, we also prefer simple controllers as similar to the case of model. There are several miscellaneous reasons for this fact, but essential one has not been discussed enough. This problem is one of main problems in this paper.

Finally, in the field of robust control and identification, the probabilistic evaluation for the uncertainties of controlled systems or the performances of the closed loop systems has been discussed in the last several years [2, 14, 8]. One of the motivations is to improve the conservatism of the uncertainties of parameters estimated by system identification and another is the application of the Monte Carlo method for robust control problem which can hardly be solved analytically. The researches on system identification and robust control on this probabilistic approach have been done independently, however, it can be easily guessed that they have a deep connection each other. For example, the complexity of model for plant decides that of the controller, and the complexity of the controller much influences on the reliability of the closed loop.

On these discussions, we proposed an information criterion for model estimation/selection with respect to controller designing in the previous papers [12, 13]. The criterion is a kind of derivation of AIC/TIC basically, but it reflects the selection of controller. In this paper, this result is extended to the more usual case of system identification of controlled systems by using a set of the I/O data. Here we consider two cases: 1) a controller is designed by using a model which is estimated from the data, 2) there exist system perturbations and a controller is designed to compensate them. In these cases, it is shown that the complexity of the controller influences on the selection of model for plant.

2 Measurable space and variable transformation

At first assume a model for plant is parametrized by some index. For example,

$$F_\theta(s) = \frac{s+a}{bs+2},\ \theta = (a,b),$$

where $F_\theta(s)$ is a model for plant and θ is its index, and also assume the index is unknown a priori. A possible way here is to give bounds of them, but now we consider to estimate the distribution from I/O data in probabilistic approach in this paper.

To estimate system parameters from I/O data is a standard problem, and there exist many results. For example, let $f_1(a)$ and $f_2(a)$ be possible models for the distribution of parameter $a \in A$, then, standard model estimation/selection problem is to decide which one is better for the real distribution of a. And, of course, that decision is not influenced by the closed loop system which is composed of $F_a(s)$ and some controller which is designed from $F_a(s)$, $f_1(a)$ or $f_2(a)$. Now suppose that $f_1(a)$ and $f_2(a)$ show good similarity with the true distribution in some subset $A_1 \subset A$ and $A_2 \subset A$ respectively, and the attained control performance $\gamma(a)$ of some controller and the model $F_a(s)$ is bad in A_1 or A_2. When the purpose of the control is to avoid bad performance, then, the model will be selected with respect to the above mentioned conditions, and this selection may be different from the normal model selection. Here we should note that the final purpose of control is the improvement of the control performance, the model should be selected with respect to it [9, 10, 11].

With this discussion, in this paper, we consider modeling of distribution of control performance substituting for that of system index. Let $x \in X$ be an index of system, $f(x)$ be a probability density of x, and $\gamma_\phi(x) \in \Gamma$ be a performance index of a closed loop system composed of the system x and a controller ϕ. Now the controller ϕ is a mapping, $\phi : X \to \Gamma$, and $f(x)$ is transformed to the density of the performance index as $\phi : f(x) \to g(\gamma)$. To consider the modeling of $g(\gamma)$, several basic notations about variable transformation of probability density is necessary, and let show them in the followings.

Let $(\mathcal{A}, X, \mu)$ and $(\mathcal{B}, Y, \nu)$ be two measurable spaces, where $\mathcal{A}$ and $\mathcal{B}$ are σ-algebras composed of all the subsets of X and Y, μ and ν are measures on X and Y, and f or g are probability densities on $(\mathcal{A}, X, \mu)$ or $(\mathcal{B}, Y, \nu)$. We also define a transformation $S : X \to Y$, then, the followings are derived [5].

Proposition 2.1 *For any nonsingular transformation $S : X \to Y$ on measurable spaces $(\mathcal{A}, X, \mu)$, $(\mathcal{B}, Y, \nu)$, there exists a unique operator $P : \mathcal{L}^1 \to \mathcal{L}^1$ which satisfies*

$$\int_{B \in \mathcal{B}} Pf(y)\nu(dy) = \int_{S^{-1}(B)} f(x)\mu(dx). \qquad (1)$$

Here P is called a Frobenius-Perron operator.

Proposition 2.2 *For a nonsingular measurable transformation $S : X \to Y$ on measurable spaces $(\mathcal{A}, X, \mu)$, $(\mathcal{B}, Y, \nu)$,*

$$\begin{aligned} \int_{S^{-1}(B)} f(x)\mu(dx) &= \int_B f(S^{-1}(y))\mu S^{-1}(dy) \\ &= \int_B f(S^{-1}(y))J^{-1}(y)\nu(dy), \end{aligned} \qquad (2)$$

where μS^{-1}, J^{-1} are defined by

$$\mu S^{-1}(B) = \mu(S^{-1}(B)) = \int_B J^{-1}(y)\nu(dy), \ B \in \mathcal{B}.$$

From this proposition, we get a realization of the Frobenius-Perron operator as follows.

Proposition 2.3 *For an invertible nonsingular measurable transformation $S : X \to Y$ on measurable spaces $(\mathcal{A}, X, \mu)$ and $(\mathcal{B}, Y, \nu)$, let P be the corresponding Frobenius-Perron operator. Then,*

$$Pf(y) = f(S^{-1}(y))J^{-1}(y).$$

Moreover, when $S(y)^{-1}$ is differentiable, $J^{-1}(y)$ and $J(x)$ are given by

$$J^{-1}(y) = \left| \frac{dS^{-1}(y)}{dy} \right|, \ J(x) = \left| \frac{dS(x)}{dx} \right|.$$

In this paper, μ, ν are supposed as the Lebesgue measure. For more details, see [5].

3 I/O data and Bayesian model

In the papers [12, 13], we considered the same problem of this paper, however, under the case that the data $\{x_i\}$ of system index is given. However in usual case, only I/O data of the system is observed. Now we consider this case and estimate the parameters of systems by using the I/O data. The following is a typical example of systems,

$$x(t) = \sum_{i=1}^{n} \alpha_i x(t-i) + \sum_{i=1}^{m} \beta_i u(t-i) + \epsilon(t)$$

where $u(t)$ and $\epsilon(t)$ are signals from some stochastic processes. Now define vectors $\boldsymbol{y}_d$, $\boldsymbol{y}(i)$, $\boldsymbol{\theta}$ by

$$\begin{aligned} \boldsymbol{y}_d \ &:= \ [\, x(1) \quad x(2) \quad \cdots \quad x(n) \\ &\qquad u(1) \quad u(2) \quad \cdots \quad u(m)\,]^{\mathrm{T}} \\ \boldsymbol{y}(i) \ &:= \ [\, x(i-1) \quad x(i-2) \quad \cdots \quad x(i-n) \\ &\qquad u(i-1) \quad u(i-2) \quad \cdots \quad u(i-m)\,]^{\mathrm{T}} \\ \boldsymbol{\theta} \ &:= \ [\, \alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n \quad \beta_1 \quad \beta_2 \quad \cdots \quad \beta_m\,]^{\mathrm{T}}. \end{aligned}$$

When $\boldsymbol{y}_d$ and ϵ obey a normal distributions, the posterior distribution of the observed data $\boldsymbol{y}$ ($\boldsymbol{y}(1) \sim \boldsymbol{y}(N)$) is given by

$$\begin{aligned} & f(\boldsymbol{y}|\sigma^2, \boldsymbol{\theta}) = q(\boldsymbol{y}_d|\sigma^2, \boldsymbol{\theta}) \prod_{i=n+1}^{N} g(\boldsymbol{y}_1(i+1)|\sigma^2, \boldsymbol{\theta}) \\ & q(\boldsymbol{y}_d|\sigma^2, \boldsymbol{\theta}) := ((2\pi)^n \det \Sigma_d)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{y}_d^{\mathrm{T}}\Sigma_d^{-1}\boldsymbol{y}_d\right) \\ & \Sigma_d := \mathbf{E}[\boldsymbol{y}_d\boldsymbol{y}_d^{\mathrm{T}}] \\ & g(\boldsymbol{y}_1(i+1)|\sigma^2, \boldsymbol{\theta}) := (2\pi\sigma^2)^{-\frac{1}{2}} \\ & \qquad \times \exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{y}_1(i+1) - \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{y}(i)\right)^2\right). \end{aligned}$$

Here we apply the Bayesian approach, and suppose $\boldsymbol{\theta}$ is a stochastic variable with the probability density $\pi_\psi(\boldsymbol{\theta})$ where ψ is a parameter of the density, then the simultaneous probability density of $\boldsymbol{y}$ and $\boldsymbol{\theta}$ is given by

$$l(\boldsymbol{y}, \boldsymbol{\theta}) := f(\boldsymbol{y}|\boldsymbol{\theta})\,\pi_\psi(\boldsymbol{\theta}) =: f_{\boldsymbol{\theta}}(\boldsymbol{y})\pi_\psi(\boldsymbol{\theta}). \tag{3}$$

Note that this system and assumptions are examples and the other cases may be possible. The important point is that when the Bayesian approach is applied, the conditional distribution of the observed I/O data and the prior distribution of system index can be defined. In the following of this paper, let $f_\theta(y)$ and $\pi_\psi(\theta)$ be the conditional distribution density and the prior distribution density respectively for simplicity, where y is a set of I/O data, θ is an index of systems, and ψ is a parameter of the distribution model of θ.

4 Estimation and criteria with controller design

In this section, we will introduce information criteria for two cases:

Case 1. y^i $(= \{y^i_j\},\ j = 1 \sim n)$: a series of observed I/O data, y: a set of y^i $(i = 1 \sim m)$, θ: system index, $f_\theta(y)$: probability density of y of system θ, $\pi_\psi(\theta)$: probability density of θ, ψ: parameter of the probability density $\pi_\psi(\theta)$. In this case, the probability density of the system index θ is supposed unchanged and ψ is common for all the data $\{y^i\}$.

Case 2. y^i $(= \{y^i_j\},\ j = 1 \sim n)$: a series of observed I/O data, y: a set of y^i $(i = 1 \sim m)$, θ^i: system index for the data y^i, $f_{\theta^i}(y^i)$: probability density of y^i of system θ^i, $\pi_{\psi^i}(\theta^i)$: probability density of θ^i, ψ^i: parameter of the probability density $\pi_{\psi^i}(\theta^i)$, $k(\psi)$: probability density of ψ. In this case, we suppose the probability density of the system index θ is supposed variable at every observation of y^i, and ψ is also a stochastic variable.

In the following of this paper, we will address these cases.

Case 1:

Based on the Kullback-Leibler information, a model which maximizes the following quantity

$$\int g(\theta) \log \pi_\psi(\theta) d\theta, \tag{4}$$

is considered a good approximation of $g(\theta)$, where $g(\theta)$ is the true probability density of θ. However $g(\theta)$ is unknown and therefore we should substitute (4) by the following log-likelihood,

$$\int \frac{1}{m} \sum_{i=1}^{m} f_\theta(y^i) \log \pi_\psi(\theta) d\theta, \tag{5}$$

where an example of $f_\theta(y^i)$ is

$$f_\theta(y^i) = \frac{1}{n} \sum_{j=1}^{n} f_\theta(y^i_j).$$

In usual case, the modeling of $f_\theta(y)$ on the space of system index θ is considered, but with the discussion in [12, 13], it is reasonable to consider the modeling of distribution on the space of the performance index γ of closed loop systems, where θ is transformed to γ by a controller ϕ as

$$S_\phi(\theta) = \gamma. \tag{6}$$

Note that S_ϕ is supposed as an invertible nonsingular measurable transformation. This condition seems to be somewhat restrictive and in some cases it does not hold. However in usual cases of control problem, by introducing appropriate deformation of the space of γ, this condition holds. Here omit the details.

By this transformation, the maximization of (5) is also transformed to

$$\max_\psi \int \frac{1}{m} \sum_{i=1}^{m} \hat{f}_\gamma(y^i) \log \hat{\pi}_\psi(\gamma) d\gamma, \tag{7}$$

and this is a proposing estimation method in Case 1. By using the relations,

$$\begin{aligned} \hat{\pi}_\psi(\gamma) &= P_\phi \pi_\psi(\gamma) \\ \hat{f}_\gamma(y^i) &= f_{S_\phi^{-1}(\gamma)}(y^i) = f_\theta(y^i), \end{aligned} \tag{8}$$

where P_ϕ is a Frobenius-Perron operator, we get the followings.

$$\begin{aligned} \int \hat{f}_\gamma(y^i) \log \hat{\pi}_\psi(\gamma) d\gamma &= \int f_{S_\phi^{-1}(\gamma)}(y^i)\, \log P_\phi \pi_\psi(\gamma) d\gamma \\ &= \int f_\theta(y^i) \log F_{\phi,\psi}(\theta) d\theta \end{aligned} \tag{9}$$

$$K_\phi(\theta) := J_\phi^{-1}\left(S_\phi(\theta)\right),\ F_{\phi,\psi}(\theta) := K_\phi(\theta)\pi_\psi(\theta)$$

Now define

$$\psi_o := \arg\max_\psi \int g(\theta) \log F_{\phi,\psi}(\theta) d\theta \tag{10}$$

$$\hat{\psi} := \arg\max_\psi \int \frac{1}{m} \sum_{i=1}^{m} f_\theta(y^i) \log F_{\phi,\psi}(\theta) d\theta. \tag{11}$$

By solving (7) with the I/O data, we can get $\hat{\psi}$, and note that the optimal estimation model reflects the difference of controller ϕ.

On the other hand, (7) contains a model π_ψ and each model may give different $\hat{\psi}$, therefore now we should consider which model is best. Some criterion is necessary to decide it and

$$\int g(\theta) \log F_{\phi,\hat{\psi}}(\theta) d\theta \tag{12}$$

is a reasonable one, because the maximization of (12) is the original purpose of this parameter estimation. However (12) is unknown because it contains $g(\theta)$. AIC of Akaike's information criterion [1] or TIC of Takeuchi's information criterion [7, 4] is known as an estimation of (12) except for the variable transformation considered in this paper. Here we show the following theorem which is derived by the same process to introduce AIC/TIC.

Theorem 4.1 *Define the followings,*

$$\mathrm{IC}_1(\phi,\,\hat{\psi},\,\psi_o) := \int \frac{1}{m}\sum_{i=1}^{m} f_\theta(y^i)\log F_{\phi,\hat{\psi}}(\theta)d\theta - \frac{1}{m}\mathrm{tr}\,J(\phi,\,\psi_o)^{-1}I(\phi,\,\psi_o) \quad (13)$$

$$J(\phi,\,\psi) := -\int \frac{\partial^2}{\partial\psi\partial\psi'}\mathcal{F}_{\phi,\psi}(y)\mathcal{G}(y)dy$$

$$I(\phi,\,\psi) := \int \frac{\partial}{\partial\psi}\mathcal{F}_{\phi,\psi}(y)\frac{\partial}{\partial\psi'}\mathcal{F}_{\phi,\psi}(y)\mathcal{G}(y)dy$$

$$\mathcal{F}_{\phi,\psi}(y) := \int f_\theta(y)\log F_{\phi,\psi}(\theta)d\theta$$

$$\mathcal{G}(y) := \int f_\theta(y)g(\theta)d\theta,$$

then, the next holds.

$$\mathbf{E}\left[\mathrm{IC}_1(\phi,\,\hat{\psi},\,\psi_o) - \int g(\theta)\log F_{\phi,\hat{\psi}}(\theta)d\theta\right] = o\left(m^{-1}\right) \quad (14)$$

The equation (14) shows that $\mathrm{IC}_1(\phi,\,\hat{\psi},\,\psi_o)$ is a reasonable estimation of (12). However note that ψ_o in $\mathrm{IC}_1(\phi,\,\hat{\psi},\,\psi_o)$ is unknown and $\hat{\psi}$ is considered as a possible substitution. Moreover $J(\phi,\,\psi_o)$ and $I(\phi,\,\psi_o)$ are also unknown because it contains $g(\theta)$ and their empirical averages is possible to be used. Note that $\mathrm{IC}_1(\phi,\,\hat{\psi},\,\hat{\psi})$ is basically identical to TIC except for the terms which contain ϕ, and it is well-known that when $\int f_\theta(y)\pi_\psi(\theta)d\theta = g_o(y)$ where $g_o(y)$ is the true density of y, then

$$\frac{1}{m}\mathrm{tr}\,J(\phi,\,\psi_o)^{-1}I(\phi,\,\psi_o) = \frac{\dim\psi}{m}$$

and TIC is equal to AIC of Akaike's information criterion [1]. However the most important point is that IC_1 depends not only on $\pi_\psi(\theta)$ but also the controller $S_\phi(\theta)$. This means that the model selected by IC_1 will changes depending on the class of controller.

Case 2:

In this case, the different distribution model $\pi_{\psi^i}(\theta^i)$ of θ is assumed for every data y^i, and consider to maximize

$$\int g_i(\theta^i)\log\pi_{\psi^i}(\theta^i)d\theta^i. \quad (15)$$

The true probability density $g_i(\theta^i)$ is unknown, therefore, (15) is substituted by an empirical distribution

$$\int \frac{1}{n}\sum_{j=1}^{n} f_{\theta^i}(y_j^i)\pi_{\psi^i}(\theta^i)\log\pi_{\psi^i}(\theta^i)d\theta^i \quad (16)$$

with the data y_j^i. By the variable transformation to the space of the performance of the resultant closed loop system, the maximization of (16) is also transformed to

$$\max_{\psi_i}\int \frac{1}{n}\sum_{j=1}^{n} \hat{f}_{\gamma^i}(y_j^i)\hat{\pi}_{\psi^i}(\gamma^i)\log\hat{\pi}_{\psi^i}(\gamma^i)d\gamma^i$$

$$= \int \frac{1}{n}\sum_{j=1}^{n} f_{\theta^i}(y_j^i)P_\phi\pi_{\psi^i}(\gamma^i)\log P_\phi\pi_{\psi^i}(\gamma^i)d\gamma^i$$

$$= \int \frac{1}{n}\sum_{j=1}^{n} f_{\theta^i}(y_j^i)K_\phi(\theta^i)\pi_{\psi^i}(\theta^i) \times \log K_\phi(\theta^i)\pi_{\psi^i}(\theta^i)d\theta^i, \quad (17)$$

and this is a proposing estimation method in Case 2. Now define

$$F_{\phi,\psi^i}(\theta^i) := K_\phi(\theta^i)\pi_{\psi^i}(\theta^i) \quad (18)$$

$$\psi_o^i := \arg\max_{\psi^i}\int g_i(\theta^i)\log F_{\phi,\psi^i}(\theta^i)d\theta^i \quad (19)$$

$$\hat{\psi}^i := \arg\max_{\psi^i}\int \frac{1}{n}\sum_{j=1}^{n} f_{\theta^i}(y_j^i)F_{\phi,\psi^i}(\theta^i) \times \log F_{\phi,\psi^i}(\theta^i)d\theta^i, \quad (20)$$

then, we also get the next information criterion similar to IC_1.

Theorem 4.2 *Define the followings,*

$$\mathrm{IC}_2(\phi,\,\hat{\psi}^i,\,\psi_o^i) := \int \frac{1}{n}\sum_{j=1}^{n} f_{\theta^i}(y_j^i)F_{\phi,\hat{\psi}^i}(\theta^i) \times \log F_{\phi,\hat{\psi}^i}(\theta^i)d\theta^i - \frac{1}{n}\mathrm{tr}\,J(\phi,\,\psi_o^i)^{-1}I(\phi,\,\psi_o^i) \quad (21)$$

$$J(\phi,\,\psi^i) := -\int \frac{\partial^2}{\partial\psi^i\partial\psi^{i\prime}}\mathcal{F}_{\phi,\psi^i}(y)\mathcal{G}_i(y)dy$$

$$I(\phi,\,\psi^i) := \int \frac{\partial}{\partial\psi^i}\mathcal{F}_{\phi,\psi^i}(y)\frac{\partial}{\partial\psi^{i\prime}}\mathcal{F}_{\phi,\psi^i}(y)\mathcal{G}_i(y)dy$$

$$\mathcal{F}_{\phi,\psi^i}(y) := \int f_{\theta^i}(y)F_{\phi,\psi^i}(\theta^i)\log F_{\phi,\psi^i}(\theta^i)d\theta^i$$

$$\mathcal{G}_i(y) := \int f_{\theta^i}(y)g_i(\theta^i)d\theta^i,$$

then, the next holds.

$$\mathbf{E}\left[\mathrm{IC}_2(\phi,\,\hat{\psi}^i,\,\psi_o^i) - \int g_i(\theta^i)\log F_{\phi,\hat{\psi}^i}(\theta^i)d\theta^i\right] = o\left(n^{-1}\right) \quad (22)$$

Here omit the comment on IC_2 because it is similar to that of IC_1.

5 Selection of the class of controller

In the previous section, information criteria for model estimation and selection are induced. In this section, at first we consider concrete controller design methods, and then we show that the criteria become stochastic variables. Finally we also calculate the estimations of such criteria.

Case 1:

Let $P_\phi\pi_\psi(\gamma)$ be a probability density of a performance distribution of the closed loop system which depends on the model, then, one of reasonable evaluations of controller ϕ is the average of the attained performance as

$$\int \gamma\, P_\phi \pi_\psi(\gamma) d\gamma. \tag{23}$$

Therefore, the optimal controller ϕ in this case is defined by

$$\phi_o(\psi) := \arg\min_{\phi} \int \gamma\, P_\phi \pi_\psi(\gamma) d\gamma. \tag{24}$$

Note that (23) is also written as

$$\int \gamma\, P_\phi \pi_\psi(\gamma) d\gamma = \int S_\phi(\theta)\pi_\psi(\theta) d\theta. \tag{25}$$

The minimization problem of (24) or (25) is very hard to be solved in general. A reasonable way is applying the Monte Carlo method, that is, at first generate sample points $\hat{\theta}^i$ ($i = 1 \sim r$) which obey probability density $\pi_\psi(\theta)$ and transform them to the performance space, and finally the controller is selected as the optimizer of the following summation.

$$\hat{\phi} := \arg\min_{\phi} \frac{1}{r}\sum_{i=1}^{r} \gamma^i = \arg\min_{\phi} \frac{1}{r}\sum_{i=1}^{r} S_\phi(\theta^i) \tag{26}$$

Here note that when π_ψ is nonsingular, the sample points $\hat{\theta}_i$ which obey π_ψ are generated by a uniformly distributed sample data $\{z^i\}$ ($=: z^{(r)}$) on a space Z and a variable transformation T_ψ which is defined by π_ψ as [3, 12]

$$\hat{\theta}^i = T_\psi(z^i). \tag{27}$$

Therefore (26) is written as

$$\begin{aligned} \hat{\phi}(\psi,\, z^{(r)}) &= \arg\min_{\phi} \frac{1}{r}\sum_{i=1}^{r} U_{\phi,\psi}(z^i) \\ U_{\phi,\psi}(z) &:= S_\phi(T_\psi(z)). \end{aligned} \tag{28}$$

Corresponding to $\hat{\phi}$, $\phi_o(\psi)$ is also given by

$$\phi_o(\psi) = \arg\min_{\phi} \int h(z) U_{\phi,\psi}(z) dz \tag{29}$$

where $h(z)$ is the density function of a uniform distribution. Here also the following convergence of $\hat{\phi}$ to ϕ_o holds.

$$\begin{aligned} &\sqrt{r}(\hat{\phi} - \phi_o) \\ &\quad \to_{r\to\infty} N(0,\, R(\phi_o,\, \psi)^{-1} Q(\phi_o,\, \psi) R(\phi_o,\, \psi)^{-1}), \\ &R(\phi,\, \psi) := -\int h(z) \frac{\partial^2}{\partial\phi\partial\phi'} U_{\phi,\psi}(z) dz \\ &Q(\phi,\, \psi) := \int h(z) \frac{\partial}{\partial\phi} U_{\phi,\psi}(z) \frac{\partial}{\partial\phi'} U_{\phi\psi}(z) dz. \end{aligned}$$

Now note that $\mathrm{IC}_1(\hat{\phi},\, \hat{\psi},\, \psi_o)$ becomes a stochastic variable. By taking the expectation of $\mathrm{IC}_1(\hat{\phi},\, \hat{\psi},\, \psi_o)$ on z and the following is given.

$$\begin{aligned} &\int h(z) \mathrm{IC}_1(\hat{\phi},\, \hat{\psi},\, \psi_o) dz^{(r)} dz \\ &= \int \frac{1}{m}\sum_{i=1}^{m} f_\theta(y^i) \log F_{\hat{\phi}(\hat{\psi}, z^{(r)}), \hat{\psi}}(\theta) d\theta \\ &- \frac{1}{m} \mathrm{tr}\, J(\hat{\phi},\, \psi_o)^{-1} I(\hat{\phi},\, \psi_o) + W \\ &- \frac{1}{mr} \mathrm{tr}\, R(\phi_o,\, \hat{\psi})^{-1} Q(\phi_o,\, \hat{\psi}) R(\phi_o,\, \hat{\psi})^{-1} V(\phi_o,\, \hat{\psi}) \\ &\quad + o\left(m^{-1}\right) + o\left(r^{-1}\right) \end{aligned} \tag{30}$$

$$V(\phi,\, \psi) := -\frac{\partial^2}{\partial\phi\partial\phi'} \int \sum_{i=1}^{m} f_\theta(y^i) \log F_{\phi,\psi}(\theta) d\theta$$

$$\begin{aligned} W := \frac{1}{m}\sum_{i=1}^{m} \frac{\partial}{\partial\phi} \int \sum_{i=1}^{m} f_\theta(y^i) \\ \times \log F_{\hat{\phi}(\hat{\psi},\, z^{(r)}), \hat{\psi}}(\theta) d\theta (\phi_o - \hat{\phi}) \end{aligned}$$

In the above equation, the true value of W is unknown but its expectation is 0, and approximate it to 0, and also with the approximation of $R(\phi_o,\, \hat{\psi}) \to R(\hat{\phi},\, \hat{\psi})$ and so on, we get an information criterion for the model estimation/selection and also the selection of controller design method as follows.

$$\begin{aligned} \mathrm{ICC}_1(\hat{\phi},\, \hat{\psi}) &= \int \frac{1}{m}\sum_{i=1}^{m} f_\theta(y^i) \log F_{\hat{\phi}(\hat{\psi}, z^{(r)}), \hat{\psi}}(\theta) d\theta \\ &- \frac{1}{m} \mathrm{tr}\, J(\hat{\phi},\, \hat{\psi})^{-1} I(\hat{\phi},\, \hat{\psi}) \\ &- \frac{1}{mr} \mathrm{tr}\, R(\hat{\phi},\, \hat{\psi})^{-1} Q(\hat{\phi},\, \hat{\psi}) R(\hat{\phi},\, \hat{\psi})^{-1} V(\hat{\phi},\, \hat{\psi}) \end{aligned} \tag{31}$$

The formula (31) shows that ICC_1 is added by a new bias term, which is approximately proportional to the number of parameters of the controller. This suggests that when the structure of controller is complex then a better controller for the given data will be induced. However, when θ is given by a stochastic process, the complex controller tends to be specialized only for the given data, and it lacks the generalization. The new bias term can be interpreted to recover such generalization of controller.

Case 2:

In this case, ψ is supposed a stochastic variable with a probability density $k(\psi)$. In this case, it is reasonable to consider that the optimal controller ϕ_o, which corresponds to (24) in the previous case, is defined by

$$\phi_o := \arg\min_{\phi} \int k(\psi) \left(\int \gamma P_\phi \pi_\psi(\gamma) d\gamma \right) d\psi. \tag{32}$$

And also define $\hat{\phi}$ which is optimal for $\psi^{(m)}$ ($:= \{\psi^1,\, \psi^2,\, \ldots,\, \psi^m\}$) as

$$\begin{aligned} \hat{\phi}(\psi^{(m)}) &:= \arg\min_{\phi} \frac{1}{m}\sum_{i=1}^{m} \int \gamma P_\phi \pi_{\psi^i}(\gamma) d\gamma \\ &= \arg\min_{\phi} \frac{1}{m}\sum_{i=1}^{m} U_\phi(\psi^i) \end{aligned} \tag{33}$$

$$U_\phi(\psi^i) := \int S_\phi(\theta)\pi_{\psi^i}(\theta) d\theta. \tag{34}$$

Note that ϕ_o is also written as

$$\phi_o = \arg\min_{\phi} \int k(\psi) U_\phi(\psi) d\psi. \tag{35}$$

Now $\mathrm{IC}_2(\hat{\phi}(\psi^{(m)}),\, \hat{\psi}^i,\, \psi_o^i)$ also becomes a stochastic variable and its expectation on ψ is given as

$$\begin{aligned}
&\int k(\psi^{(m)}) \frac{1}{m} \sum_{i=1}^{m} \left(\int g_i(\theta^i) \log F_{\phi,\hat{\psi}^i}(\theta^i) d\theta^i \right) d\psi^{(m)} \\
&= \frac{1}{m} \sum_{i=1}^{m} \int \frac{1}{n} \sum_{j=1}^{n} f_{\theta^i}(y_j^i) F_{\phi,\hat{\psi}^i}(\theta^i) \log F_{\phi,\hat{\psi}^i}(\theta^i) d\theta^i \\
&\quad - \frac{1}{m} \sum_{i=1}^{m} \frac{1}{n} \mathrm{tr}\, J(\hat{\phi},\, \psi_o^i)^{-1} I(\hat{\phi},\, \psi_o^i) + W \\
&\quad - \frac{1}{m^2 n} \mathrm{tr}\, \left(R(\phi_o)^{-1} Q(\phi_o) R(\phi_o)^{-1} V(\phi_o) \right) \\
&\quad + o\left(n^{-1}\right) + o\left(m^{-1}\right)
\end{aligned} \tag{36}$$

$$R(\phi) := -\int k(\psi) \frac{\partial^2}{\partial\phi\partial\phi'} U_\phi(\psi) d\psi$$

$$Q(\phi) := \int k(\psi) \frac{\partial}{\partial\phi} U_\phi(\psi) \frac{\partial}{\partial\phi'} U_\phi(\psi) d\psi$$

$$\begin{aligned}
W := \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial\phi} \int \frac{1}{n} \sum_{j=1}^{n} f_{\theta^i}(y_j^i) F_{\phi,\hat{\psi}^i}(\theta^i) \\
\times \log F_{\phi,\hat{\psi}^i}(\theta^i) d\theta^i (\phi_o - \hat{\phi}).
\end{aligned}$$

Similar to ICC_1 in Case 1, we finally get the next criterion.

$$\begin{aligned}
&\mathrm{ICC}_2(\hat{\phi},\, \hat{\psi}) \\
&= \frac{1}{m} \sum_{i=1}^{m} \int \frac{1}{n} \sum_{j=1}^{n} f_{\theta^i}(y_j^i) F_{\phi,\hat{\psi}^i}(\theta^i) \log F_{\phi,\hat{\psi}^i}(\theta^i) d\theta^i \\
&\quad - \frac{1}{m} \sum_{i=1}^{m} \frac{1}{n} \mathrm{tr}\, J(\hat{\phi}(\hat{\psi}^{(m)}),\, \hat{\psi}^i)^{-1} I(\hat{\phi}(\hat{\psi}^{(m)}),\, \hat{\psi}^i) \\
&\quad - \frac{1}{m^2 n} \mathrm{tr}\, \Big(R(\hat{\phi}(\hat{\psi}^{(m)}))^{-1} Q(\hat{\phi}(\hat{\psi}^{(m)})) \\
&\qquad \times R(\hat{\phi}(\hat{\psi}^{(m)}))^{-1} V(\hat{\phi}(\hat{\psi}^{(m)})) \Big)
\end{aligned} \tag{37}$$

Here omit the comment on ICC_2 because it is similar to that of ICC_1.

6 Conclusion

In this paper, we proposed information criteria for model estimation/selection problem based on AIC/TIC with respect to the attained performance of closed loop systems. They are composed of the log-likelihood not only with a bias term which is proportional to the complexity of model but also a new term which is proportional to the complexity of controller.

The concrete control performance is not referred in this paper, however solving (26) or (33) with H_∞ norm as γ for example, needs enormous calculation power. The application of this paper for individual problems is left for future work.

The similar discussion based on MDL [6], which is related to the complexity of model and its estimation/selection problem, can be also applicable.

References

[1] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Control, AC-19-6, 716–723, 1974.

[2] M. A. Dahleh, T. V. Theodosopoulos, and J. N. Tsitsiklis, The sample complexity of worst–case identification of FIR linear systems, Systems & Control Letters, 20–3, 157–166, 1993.

[3] M. Fushimi, Stochastic method and simulation, Iwanami Seminar, Applied Mathematics, Iwanami, 1994 (in Japanese).

[4] S. Konishi and G. Kitagawa, Generalised information criteria in model selection, Biometrika, 83–4, 875–890, 1996.

[5] A. Lasota and M. C. Mackey, Chaos, fractals, and noise; Stochastic aspects of dynamics, Second edition, volume 97 of Applied mathematical sciences, Springer-Verlag, New York, 1995.

[6] J. Rissanen, Universal coding, information, prediction and estimation, IEEE Trans. Information Theory, IT-30, 629–636, 1984.

[7] K. Takeuchi, Distribution of information statistics and criteria for adequacy of models, Mathematical Sciences, 153, 12–18, 1976.

[8] R. Tempo, E. W. Bai, and F. Dabbene, Probabilistic robustness analysis: explicit bounds for the minimum number of samples, Proceedings of the 35th CDC, 3424–3428, 1996.

[9] K. Tsumura and S. Shin, Decision of nominal model and closed loop system design, Proceedings of the 18th SICE Symposium on Dynamical System Theory, 285–288, 1995.

[10] K. Tsumura and S. Shin, Robust control with distribution condition of system parameters, Proceedings of the 25th SICE Symposium on Control Theory, 247–252, 1996.

[11] K. Tsumura and S. Shin, Simultaneous modeling and data-distribution-dependent robust control system design, Preprints of the 11th IFAC Symposium on System Identification, 129–134, 1997.

[12] K. Tsumura, Model estimation and selection for control, Proceedings of the 28th SICE Symposium on Control Theory, 133–136, 1999 (in Japanese).

[13] K. Tsumura, Information criterion for model selection with controller design, Proceedings of the 12nd IFAC Symposium on System Identification, 2000.

[14] M. Vidyasagar, A theory of learning and generalization, Springer, London, 1997.