Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-7 (2016) 508–513

On-line Full Probability Distribution Identification of ARX Model Parameters Based on Bayesian Approach

Amir Hosein Valadkhani * **Aminollah Khormali** * **Mahdi Aliyari Shoorehdeli** * **Hamid Khaloozadeh** * **Alireza Fatehi** *

* *Industrial Control Center of Excellence, Faculty of Electrical Engineering, K.N. Toosi University of Technology, Tehran, Iran (e-mail: amirhosein.valadkhani@ee.kntu.ac.ir)*

Abstract: In this contribution, full probability distribution of parameters of ARX model is obtained for on-line problems by means of Bayesian approach and Markov chain Monte Carlo method (MCMC), which provides the ability to be applied on time-varying ARX models as well. Full probability distribution of parameters represent whole available knowledge of parameters. So, decision makers can follow any policies to make decision about point estimation, like dynamic point estimation. Moreover, the Bayesian approach has great potential in combining sources of knowledge much more easier. To decrease the computational efforts, full probability of model parameters are updated based on size-varying partitions. Furthermore, incorporating the posterior probability of previous partition into the jump probability of current partition, in MCMC method, improves the performance of the proposed algorithm from the computation and convergence rate point of view. Simulation results demonstrate the effectiveness and validity of the proposed algorithm.

Keywords: Bayesian statistics, System identification, Markov chain Monte Carlo (MCMC), Metropolis-Hastings (MH) Method, ARX model

1. INTRODUCTION

One of the fundamental differences between the beliefs of the Bayesian and frequentist statisticians is that Bayesian ends up with a function of the model parameters given the observed data and inference is made in the form of probability distribution of parameters rather than a simple point estimate which is common in frequentist approach (Brooks, 2003).

Intrinsic potential of Bayesian approach brings superb help to combine sources of knowledge, utilize all the available information, and also give decision makers more knowledge to make decision. In system identification (SI), more specifically, Bayesian removes the limitation of normal assumption of distribution of noise, so Bayesian shows great results when the available data is limited (Ninness and Henriksen, 2010). Another advantage that encourages one to employ Bayesian in SI is that it provides the confidence about parameters as a byproduct of solution (Baldacchino et al., 2013). Besides, overfitting is not meaningful in Bayesian methods (Green, 2015). The main obstacle on the way of using Bayesian is its heavy computational efforts.

In control society, Peterka (1981), explained for the first time how Bayesian concept can be used in SI problems. How to solve the equations obtaining from Bayesian in SI problems was major hindrance to take the advantages of Bayesian in SI. So, some restrictive assumptions were made to simplify the equations like calculating just specific features of the model parameters distribution; mean and variance for example (Huang and Wang, 2006). Another approach was using conjugate probabilities to keep *posterior distribution* unchanged (O'Hagan and Forster, 2004). Along with progressing well-developed processors which facilitates computations, the simplifying assumptions were not necessary anymore and Markov chain Monte Carlo methods (MCMC) were employed in the Bayesian approach for SI problems. Some innovative contributions on how MCMC and Bayesian approach can be used in SI issues are available in Green (2015); Ninness and Henriksen (2010); Baldacchino et al. (2013).

Previous researches about using MCMC methods in SI with Bayesian approach were focused on off-line problems due to heavy computations. Moreover, obtaining full probability distribution of parameters in on-line problems was not their main concern; just *maximum* of the distribution or *expectation* of it was chosen in on-line problems (Huang and Wang, 2006). However, determining only one point from the distribution has somehow conflict with Bayesian concept in SI, in contrast to the frequentist method.

In this research, computing the full probability of model parameters stimulating from *posterior probability* using MCMC in Bayesian method is taken into account. Hence, full probability distribution of estimated output can be obtained readily. Due to the heavy computations of MCMC methods, partitions with variable size are considered; the size of the partitions are varied based on dynamic of systems. Also, to represent the concept of *forgetting factor* in SI, a factor is introduced in proposed method which

2405-8963 © 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.07.393

assigns the importance of *posterior probability* of previous partition into *posterior probability* of current partition. For reducing the severity of computational problems, along with using previous *posterior probability* distribution as prior knowledge of next partition, we propose combining the *posterior probability* of previous partition in *jump distribution* of current partition too.

The rest of the paper is structured as follows. In Section 2, formulation and concept of Bayesian in estimation of model parameters, off-line form of linear static and dynamic models, are portrayed. Section 3 discusses about the MCMC methods as a solution of solving equations demonstrated in Section 2; in addition, the main steps of Metropolis-Hastings algorithm as one of the well-known MCMC algorithm, which is used in this paper, is presented. Section 4 describes proposed algorithm for online Bayesian identification of ARX models to get the full probability distribution of parameters. In Section 5, the simulation results of applying proposed method are presented. Finally, the paper is concluded in Section 6.

2. SYSTEM IDENTIFICATION FROM BAYESIAN POINT OF VIEW

In Bayesian system identification, the parameters of the model are assumed as both stochastic and unknown phenomena (Anscombe, 1961; Bishop, 2006). Intuitively, the more one parameter is known, the fewer it is dispersed. The final aim in Bayesian approach is determining the probability distribution of parameters of the model which is called *posterior probability*. This distribution is obtained based on the observed data and prior knowledge. Bayes theorem light the way of how to incorporate these two sources of the information (Bishop, 2006; Khatibisepehr et al., 2013):

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}, \tag{1}$$

$$P(D) = \int P(D|\theta)P(\theta)\,\mathrm{d}\theta, \tag{2}$$

where, $P(\theta|D)$ is *posterior probability* of parameters given the data, $P(D|\theta)$ is *likelihood function* or *probability model*, $P(\theta)$ is *prior probability* of parameters, and $P(D)$ is called *evidence*. Peterka (1981) shows that if observations, i.e. input-output pairs, are assumed to be identically independent distribution (i.i.d), equation (1) can be extended to reach equation (3).

$$P(\theta|D_1^t) = \frac{(\prod_{\tau=t_1+1}^{t} P(y_\tau|u_\tau, D_1^{\tau-1}, \theta))P(\theta|D_1^{t_1})}{\int \prod_{\tau=t_1+1}^{t} P(y_\tau|u_\tau, D_1^{\tau-1}, \theta)P(\theta|D_1^{t_1})\mathrm{d}\theta}, \tag{3}$$

where,

$$D_1^\tau = \{(u_1, y_1), (u_2, y_2), \cdots, (u_\tau, y_\tau)\}. \tag{4}$$

In equations (3) and (4), u_τ and y_τ are input and output when time is τ, respectively.

The probability distribution of the noise is an essential key to obtain the probability of the model parameters because noise in SI problems is enumerated as the major source of uncertainty on parameters. In linear static models, equation (5), the *probability model* is achieved as equation (6) (Nelles, 2001; Lindley and Smith, 1972).

$$Y = X^T\theta + \upsilon. \tag{5}$$

In (5), Y is a $(N \times 1)$ vector of outputs, X is a $(n \times N)$ regression matrix and θ is a $n \times 1$ vector of parameters or coefficients. Moreover, υ is $N \times 1$ noise vector.

To obtain the probability of model parameters, the major part is calculating the *probability model* because by mixing it with *prior probability*, *posterior probability* can be easily acquired. Equation (6) shows the *probability model* of static linear model parameters.

$$P(Y|\theta, \eta, X) = P_\upsilon\left(Y - X^T\theta|\eta, X\right). \tag{6}$$

In (6), P_υ is probability distribution of noise and η is parameters of this probability distribution. There is no restriction on types of probability distribution of noise. Assuming normal distribution for it, can decrease some computational complexities; however, it is not necessarily needed.

For linear dynamic models presented in the following

$$y_t = G(q,\theta)u_t + H(q,\theta)\varepsilon_t, \tag{7}$$

one-step-ahead optimal predictor is equal to equation (8). In equations (7) and (8), u_t is a vector of observed exogenous input, $G(q,\theta)$ and $H(q,\theta)$ are transfer functions, rational in the forward shift operator q, and θ is a vector of model parameters (Nelles, 2001).

$$\hat{y}_{t|t-1}(\theta) = H^{-1}(q,\theta)G(q,\theta)u_t + [1 - H^{-1}(q,\theta)]y_t. \tag{8}$$

The goal of Bayesian approach in estimation of linear dynamic models can be defined as deriving the probability distribution of parameters of one-step-ahead optimal predictors.

Same as linear static models, by assuming observations as i.i.d, the *probability model* can be shown as follows.

$$P(Y_{t_1}^t|\theta, \eta, U_{t_1}^t) = P(y_0|\theta, \eta) \prod_{t=1}^{t} P_\varepsilon(y_t - \hat{y}_{t|t-1}(\theta)|\eta, D_{t_1}^{t-1}, u_t). \tag{9}$$

Equations (6) and (9) generally do not have close-form solution. In Section 3, methods of solving these equations are discussed.

3. OVERVIEW OF MCMC METHODS AND THEIR CONVERGENCE

3.1 Implementation of MCMC method

MCMC methods involve numerically computing the required probability distribution. In these methods instead of generating samples from desired probability distribution, $\pi(\theta)$, which cannot be done directly, producing Markov chain with equilibrium distribution of $\pi(\theta)$ is taken into account. Because stimulating samples from the latter probability distribution is more straightforward (Smith and Roberts, 1993).

There are a variety types of MCMC methods which can be found in Brooks et al. (2011). In this research, Metropolis-Hastings (MH) algorithm is used because of its strength in stimulating multivariate distribution. In addition, the conditional probabilities in this work is hard to find (Chib and Greenberg, 1995). The main steps of the MH algorithm are clearly presented as follow (Ninness and Henriksen, 2010).

MH Algorithm:

(1) Initializing θ_0 such that $P(\theta_0|Y) > \epsilon > 0$
(2) Set $k = 1$