COMSTA: 6621 Model 3G pp. 1–17 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Sequential Bayesian inference for static parameters in dynamic state space models

Arnab Bhattacharya *, Simon P. Wilson

Department of Statistics, School of Computer Science and Statistics, Trinity College, Dublin 02, Ireland

ARTICLE INFO

Article history:
Received 14 August 2017
Received in revised form 20 May 2018
Accepted 25 May 2018
Available online xxxx

Keywords:
Sequential estimation
Static parameter
Dynamic state space models
Bayesian inference
Grid-based methods

ABSTRACT

A method for sequential Bayesian inference of the static parameters of a dynamic state space model is proposed. The method is able to use any valid approximation to the filtering and prediction densities of the state process. It computes the posterior distribution of the static parameters on a discrete grid that tracks the support dynamically. For inference of the state process, the Kalman filter and its extensions as well as cubature filtering have been used. It is illustrated with several examples including the stochastic volatility model and the challenging Kitagawa model and is compared to both online and offline methods. It is shown to provide a good trade off between speed and performance.

1. Introduction

Dynamic state space models (Durbin and Koopman, 2001), consisting of an unknown state Markov process represented by $\{X_t\}_{t\in\mathbb{N}}$ and noisy observations drawn from a process represented by $\{Y_t\}_{t\in\mathbb{N}}$, that are conditionally independent, are used in a wide variety of applications — e.g. wireless networks (Haykin et al., 2004), object tracking (Ristic et al., 2004) etc. Each of these are $\mathcal{X}$ and $\mathcal{Y}$-valued stochastic processes respectively. The model is specified by an initial density $p(x_0|\theta)$, a transition kernel $p(x_t|x_{t-1},\theta)$ and an observation density $p(y_t|x_t,\theta)$. These densities are defined in terms of a collection of K static (non-time varying) parameters $\theta = (\theta_1,\ldots,\theta_K)$. The joint model to time t is:

$$p(\mathbf{y}_{1:t},\mathbf{x}_{0:t},\theta) = \left(\prod_{j=1}^{t} p(y_j|x_j,\theta)p(x_j|x_{j-1},\theta)\right) p(x_0\mid\theta)p(\theta), \tag{1}$$

where $\mathbf{y}_{1:t} = (y_1,\ldots,y_t)$ and $\mathbf{x}_{0:t} = (x_0,\ldots,x_t)$.

In this paper, we propose a new approach for approximating $p(\theta|\mathbf{y}_{1:t})$, the posterior density function of static parameters θ conditional on the observations up to time t

$$p(\theta|\mathbf{y}_{1:t}) = \int p(\theta,\mathbf{x}_{0:t}|\mathbf{y}_{1:t})\mathrm{d}\mathbf{x}_{0:t}. \tag{2}$$

The principal idea is to approximate the posterior on a carefully constructed grid to which points are added or deleted where necessary. Our method is used alongside a large range of state process estimation algorithms making it highly flexible, even for non-linear or non-Gaussian models. We provide evidence corresponding to our claims through a series of examples.

* Corresponding author.
E-mail addresses: bhattaca@tcd.ie (A. Bhattacharya), Simon.Wilson@tcd.ie (S.P. Wilson).

https://doi.org/10.1016/j.csda.2018.05.018
0167-9473/

The literature has tended to focus on computation of the predictive densities $p(x_t|\mathbf{y}_{1:t-1}, \theta)$ and $p(y_t|\mathbf{y}_{1:t-1}, \theta)$, and the filtering density $p(x_t|\mathbf{y}_{1:t}, \theta)$ at time t. Updating of these densities is by the well-known forward equations (Arulampalam et al., 2002); for example for the prediction of the state process we need to compute

$$p(x_t|\mathbf{y}_{1:t-1}, \theta) = \int p(x_t|x_{t-1}, \theta)p(x_{t-1}|\mathbf{y}_{1:t-1}, \theta)\, \mathrm{d}x_{t-1}, \tag{3}$$

where $p(x_{t-1}|\mathbf{y}_{1:t-1}, \theta)$ is the filtering density from time $t-1$. For the linear Gaussian case with known parameters, these computations reduce to the closed form of the Kalman filter (Kalman, 1960). Generally, Gaussian forms of the model allows filtering and prediction to be done quickly, exactly and sequentially, without the need to store the data sequence (West and Harrison, 1997). Exact inference has also been achieved for models where the state process is discrete (Cappé et al., 2005).

Extending inference to non-linear and/or non-Gaussian models has proved to be challenging since analytical solutions to the above integrals do not exist. Functional approximation approaches, derived from the Kalman filter, such as the extended Kalman filter (Haykin, 2001), unscented Kalman filter (Julier and Uhlmann, 1997) etc. have been proposed. Sequential Monte Carlo (SMC) approaches such as the bootstrap filter (Gordon et al., 1993) and auxiliary particle filter (Pitt and Shephard, 1999) have also been widely used. Attempts have also been made to combine sampling based methods with functional approximations (Merwe et al., 2001).

As regards the static parameter estimation problem, few closed form solutions are available. Different approaches are employed in the sequential inference literature, often involving joint inference of static parameters and the state process (Kantas et al., 2015). We will discuss some of the existing methodologies in the next section.

Section 2 is a review of some of the online static parameter estimation methods, some of which are used later for comparison. Section 3 outlines the principle of the method. Section 4 describes the most important issues to be resolved in order to implement the method: approximations to one-step ahead filtering and prediction densities, updating the posterior density grid over time, choice of some heuristics in the algorithm and finally, computing posteriors corresponding to new grid points. A discussion on the theoretical aspects of error accumulation is provided in Section 5. Section 6 illustrates the method and assesses its performance against alternative approaches. Section 7 contains some concluding remarks.

2. Online parameter inference review

In this paper, interest is not on the joint density but rather on online inference of the posterior density $p(\theta|\mathbf{y}_{1:t})$ along with the state filtering and prediction densities. Several related approaches have been proposed in the literature; a recent review is Kantas et al. (2015):

Joint KF/EKF/UKF: A common practice in the engineering literature is to add dynamics to static parameters, such as assuming $\theta_t \sim N(\theta_{t-1}, V_t)$ with variance V_t decreasing with t, and make inference on $p(x_t,\ \theta_t|\mathbf{y}_{1:t-1})$ at each time point using a single Kalman filter or extensions thereof. One advantage of this is that the state and parameters are typically correlated *a posteriori*, even in linear systems (Haykin, 2001); however this is known to suffer from numerical instability issues. A more fundamental problem with this approach is that it makes little sense to treat static parameters as time-varying.

Dual KF/EKF/UKF: Dual filters assume that the state and the parameter have separate state space representations, and thus two filters can be run concurrently (Wan and Nelson, 1997). The prediction $p(x_t|\mathbf{y}_{1:t-1}, \hat{\theta}_{\text{old}})$ is derived using the parameter mean and $p(\theta_t|\mathbf{y}_{1:t-1}, \hat{x}_{\text{old}})$ is updated using the filtered state mean. Because each posterior only uses the first moment of the other and ignores the variance, these methods are known to produce low variance posteriors.

Online gradient method: Sequential optimization of $\log(p(\mathbf{y}_{1:t}\,|\,\theta))$ is also possible. If $\hat{\theta}_{t-1}$ is the estimate after the first $t-1$ observations, then it is updated to $\hat{\theta}_t$ after receiving new data y_t; see Poyiadjis et al. (2011) and Nemeth et al. (2016) for details. The problem of evaluating the gradient for the whole data $\mathbf{y}_{1:t}$ has been bypassed in the case of hidden Markov models (LeGland and Mever, 1997). A typical problem associated with any gradient method is that it is extremely sensitive to initialization and may converge to a local maximum.

Online EM: Online versions of the EM algorithm (Dempster et al., 1977), suitable for a dynamic state space model with unknown model parameters, have been proposed in Cappé (2011). A major advantage is that it always attempts to maximize the likelihood, allowing methods such as variational inference to be used to estimate the parameters. However this method can also converge to a local maximum.

Liu and West filter: This is the most generic particle filter that performs dual estimation of the state and parameters. Artificial dynamics are introduced for the parameters and subsequently a kernel density estimate of $p(\theta|\mathbf{y}_{1:t})$ is proposed from which θ can be sampled (Liu and West, 2000). Shrinkage is introduced to control for over-dispersion in the kernel density function. A major drawback though is that it requires a significant amount of tuning for quantities such as the kernel bandwidth.

Storvik's filter: Storvik (2002) generates particles from the parameter's posterior distribution without assuming any associated random drift. It is further assumed that the posterior distribution of θ depends on a low-dimensional set of sufficient statistics that can be efficiently updated for each t. The choice of this set is the biggest stumbling block for this algorithm.

Particle learning: Carvalho et al. (2010) have provided a modified version of Storvik's filter which has proved to be more efficient. Sufficient statistics are derived for the class of conditional dynamic linear models (CDLM), thus providing additional structure to Storvik's algorithm. Further, the position of the resampling step in Storvik's algorithm is now interchanged with the propagation step, that allows particle deficiency of the posterior of θ to be reduced.

Please cite this article in press as: Bhattacharya A., Wilson S.P., Sequential Bayesian inference for static parameters in dynamic state space models. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.05.018.