Statistics and Probability Letters 83 (2013) 2077–2080

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

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

Tuned iterated filtering

Erik Lindström

Division of Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Box 118, SE-22100 Lund, Sweden

ARTICLE INFO

Article history:
Received 2 November 2012
Received in revised form 14 March 2013
Accepted 16 May 2013
Available online 24 May 2013

MSC:
60J05
62M05
62M10

Keywords:
Hidden Markov models
Sequential Monte Carlo methods
Maximum likelihood estimation

ABSTRACT

Iterated filtering is an algorithm for estimating parameters in partially observed Markov process (POMP) models. The real-world performance of the algorithm depends on several tuning parameters. We propose a simple method for optimizing the parameter governing the joint dynamics of the hidden parameter process (called the Σ matrix).

The tuning is implemented using a fixed-lag sequential Monte Carlo expectation–maximization (EM) algorithm. We introduce two different versions of the tuning parameter, the approximately estimated Σ matrix, and a normalized version of the same matrix.

Our simulations show that the finite-sample performance for the normalized matrix outperform the standard iterated filter, while the naive version is doing more harm than good.

1. Introduction

Iterated filtering (IF), see Ionides et al. (2006, 2011), is a popular plug-and-play method for inference in partially observed Markov process models; see Bretó et al. (2009); He et al. (2010). Plug-and-play methods require only simulations from the model, and being able to relate those simulations to measurements. Applications include cholera transmissions, see King et al. (2008), measles, see He et al. (2010), the role of climate in malaria transmissions, see Laneri et al. (2010), and finance, see Bhadra (2010). The method is implemented in the open-source R package `pomp`.[1]

A number of design parameters must be initialized prior to applying the iterated filtering algorithm to data. Some of the design parameters were linked in Lindström et al. (2012). This paper studies optimization of the joint dynamics of the hidden parameters, as expressed by the Σ matrix; see Eq. (1). The asymptotic rate of convergence does not depend on Σ, but the practical performance, in terms of covariance of the estimates, often does; cf. Fabian (1978).

The Σ matrix is often chosen as a diagonal matrix; cf. Ionides et al. (2006). This choice is robust, but will not utilize any dependence structures in the model. We introduce a simple (essentially free) method for finding a near-optimal value for the full Σ matrix. It is also possible to recover the near-optimal diagonal matrix within the proposed framework. The matrix is obtained using a fixed-lag sequential Monte Carlo (SMC) expectation–maximization (EM) method; see Olsson et al. (2008). Two closely related approximations are derived, with vastly different qualitative properties.

The paper is organized as follows. Section 2 reviews the iterated filtering algorithm, and presents the estimators of the Σ matrix, Section 3 evaluates these estimators when applying the iterated filtering algorithm to a simple model, and Section 4 concludes.

E-mail address: erikl@maths.lth.se.

[1] Available on CRAN at http://www.r-project.org.

0167-7152/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.05.019

2. Methods for partially observed Markov process models

Partially observed Markov processes are a class of processes satisfying the following conditions. Let $\{X_k\}$ be a hidden $\mathcal{X}$-valued Markov process with initial density $f_{X_0}(x_0;\theta)$ and transition density $f_{X_k|X_{k-1}}(x_k|x_{k-1};\theta)$. The $\mathcal{Y}$-valued observation process $\{Y\}$ depends on the hidden process at time k through the observation density $f_{Y_k|X_k}(y_k|x_k,\theta)$, and it is assumed that $Y_k|X_k$ is independent of all other observations. Finally, the model is governed by the parameter vector θ.

Iterated filtering is distantly related to the "combined state and parameter" estimation method that is popular in engineering, see Ljung (1979), statistics, see Kitagawa (1998), finance, see Lindström et al. (2008), and hydrology, see Evensen (2009). The idea is to augment the latent $\{X\}$ process with the parameters with vanishing dynamics. Let $\kappa(\cdot)$ be a density with compact support, and let $\{\zeta_k\}$ be independent draws from $\kappa(\cdot)$ such that

$$\mathbb{E}[\zeta_k]=0,\qquad \text{Var}[\zeta_k]=\Sigma,\quad \forall\, k. \tag{1}$$

The time-varying parameter dynamics is then given by

$$\Theta_0=\theta+\tau\zeta_0,\qquad \Theta_n=\Theta_{n-1}+\sigma\zeta_n, \tag{2}$$

where τ and σ are small positive numbers. The joint distribution of the augmented model is given by

$$g_{X_{0:N},\Theta_{0:N},Y_{1:N}}(x_{0:n},\theta_{0:N},y_{1:N};\theta,\sigma,\tau,\Sigma)=f_{X_{0:N},Y_{1:N}}(x_{0:n},y_{1:N};\theta_{0:N})g_{\Theta_{0:N}}(\theta_{0:N};\theta,\sigma,\tau,\Sigma), \tag{3}$$

where $f_{X_{0:N},Y_{1:N}}$ is the joint distribution of $\{X,\ Y\}$ conditional on the parameters and $g_{\Theta_{0:N}}$ is the distribution of the parameters.

Define the conditional mean and covariance, computed with respect to the joint distribution, of the parameter process as

$$\theta_n^F=\mathbb{E}_{\theta,\tau,\sigma,\Sigma}[\Theta_n|Y_{1:n}], \tag{4}$$

$$V_n^P=\text{Cov}_{\theta,\tau,\sigma,\Sigma}[\Theta_n|Y_{1:n-1}]. \tag{5}$$

It was shown in Ionides et al. (2011) that the score function can be approximated using these moments.

Theorem 2.1 (*Theorem 3 in Ionides et al., 2011*). *Let K_1 be a compact subset of $\mathbb{R}^p$, C_1 be a constant, τ be sufficiently small, and $\lim_{\tau\to 0}\sigma(\tau)/\tau=0$. It then holds that*

$$\sup_{\theta\in K_1}\left|\sum_{n=1}^{N}(V_n^P)^{-1}\left(\theta_n^F-\theta_{n-1}^F\right)-\log f_{Y_{1:N}}(y_{1:N};\theta)\right|\le C_1\left(\tau+\frac{\sigma^2}{\tau^2}\right). \tag{6}$$

These moments are not known for most models, but they can be accurately approximated using SMC (particle filter) methods. Let $\tilde{\theta}_n^F$ and $\tilde{V}_n^P$ be the empirical versions of Eqs. (4) and (5) computed using J particles. The score can still be accurately approximated using $\tilde{\theta}_n^F$ and $\tilde{V}_n^P$ as long as a sufficient number of particles is used; cf. Theorem 4 in Ionides et al. (2011).

The iterated filtering algorithm is a stochastic approximation algorithm, in which the biased and noisy approximation of the score function is used to iteratively update the estimate of the parameters, eventually arriving at the maximum likelihood estimate (MLE) (we use iteration index m; i.e., $\tilde{\theta}_{n,m}^F$ is the SMC estimate of θ_n^F during iteration m).

Theorem 2.2 (*Theorem 5 in Ionides et al., 2011*). *Let $\{a_m\}$, $\{\tau_m\}$, $\{\sigma_m\}$ and $\{J_m\}$ be positive sequences such that $\tau_m\to 0$, $\sigma_m\tau_m^{-1}\to 0$, $\tau_m J_m\to\infty$, $\sum_{m=1}^{\infty}a_m=\infty$ and $\sum_{m=1}^{\infty}a_m^2J_m^{-1}\tau_m^{-2}<\infty$, and define $\hat{\theta}_m$ according to*

$$\hat{\theta}_{m+1}=\hat{\theta}_m+a_m\sum_{n=1}^{N}\left(\tilde{V}_{n,m}^P\right)^{-1}\left(\tilde{\theta}_{n,m}^F-\tilde{\theta}_{n-1,m}^F\right). \tag{7}$$

The estimate will then converge to the MLE with probability 1: $\theta_m\xrightarrow{a.s.}\theta^\star$.

Ionides et al. (2011) noted that choosing $a_m=m^{-1}$, $\tau_m^2=m^{-1}$, $\sigma_m^2=m^{-(1+\delta)}$, and $J_m=m^{(1/2+\delta)}$, where $\delta>0$, satisfies the conditions in Theorem 2.2.

2.1. Tuning the Σ matrix

Maximum likelihood methods can be used to estimate the Σ parameter, as the joint distribution is known; cf. Eq. (3).

An integral part of the iterated filtering algorithm is applying SMC filters for estimating properties of the latent processes, including the joint posterior density. However, Olsson et al. (2008) found that the naive approximation of the posterior distribution degenerates quickly, and argued that a fixed-lag approximation

$$p(x_k|y_{1:N})\approx p(x_k|y_{1:\min(k+L,N)}) \tag{8}$$

is less variable. They also showed how this approximation can be used to estimate parameters within the EM algorithm. The resulting algorithm is computationally attractive, but the estimates are often slightly biased; cf. Kantas et al. (2009).