Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 6, n. 2, 2018.

Trabalho apresentado no XXXVIII CNMAC, Campinas - SP, 2018.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

---

Improvements on Parameter Estimation based on Particle Approximations of the Gradient and Information Matrix in State Space Models

Uriel M. Silva[1]
Departamento de Estatística, UFMG
Rodolfo Santos Nunes Rodrigues[2]
Departamento de Estatística, UFMG
Felipe Carvalho Álvares da Silva[3]
Banco Inter
Luiz Henrique Duczmal[4]
Departamento de Estatística, UFMG
Denise Burgarelli Duczmal[5]
Departamento de Matemática, UFMG

Let $(X_t, Y_t)_{t\geq 1}$ be a homogeneous discrete-time bivariate stochastic process where $(X_t)_{t\geq 1}$ is a Markov chain and $(Y_t|X_t)_{t\geq 1}$ is a conditionally independent sequence such that each Y_t is determined almost surely by X_t. If we also assume that only $(Y_t)_{t\geq 1}$ is available for inference, i.e. that the Markov chain $(X_t)_{t\geq 1}$ is unobservable, then $(X_t, Y_t)_{t\geq 1}$ is usually called a *state space* or *hidden Markov* model (HMM). Usually, the law of $(X_t, Y_t)_{t\geq 1}$ is also taken to be indexed by a d-dimensional parameter θ taking values in Θ.

Upon measuring a sequence $(Y_t)_{t=1}^T$, there are basically three kinds of inference we can perform about the law of the *state* $(X_t)_{t=1}^T$, namely *prediction*, *filtering* and *smoothing*, which are respectively the computation of $X_t|Y_{1:t-1}$, $X_t|Y_{1:t}$ and $X_t|Y_{1:T}$ for all $t = 2, \ldots, T$. Since in general none of these distributions are available in closed-form, computational methods such as *particle filters* (also known as *sequential Monte Carlo*) are used to approximate it.

Particle filters are a class of importance sampling algorithms that exploit the sequential nature of HMMs to produce approximations of the probability density of $X_t|Y_{1:t}$, also known as *filtering density* or distribution by simulation [1]. In possession of a set of *particles* $(X_t^i)_{i=1}^N$ with respective *importance weights* $(w_t^i)_{i=1}^N$, we are in principle able to produce a Monte Carlo estimator of any measureable function of $X_t|Y_{1:t}$. Further

[1] urielmoreirasilva@gmail.com
[2] rodolforodrigues_1@hotmail.com
[3] felipe.alvares@gmail.com
[4] duczmal@gmail.com
[5] burgarelli@ufmg.br

exploitation of the Markovian nature of HMMs allow us to also estimate functions of predicted and smoothed states, i.e. $X_t|Y_{1:t-1}$ and $X_t|Y_{1:T}$, for all t.

However, when the law of $(X_t, Y_t)_{t\geq 1}$ is a parametric family, θ determines the distribution of the HMM almost surely. Hence, even if our only goal is to perform inference upon $(X_t)_{t\geq 1}$, we also have to deal with θ by either obtaining an estimate for it or "integrating it out"of the model as additional noise.

Typically, an estimate for θ can be obtained by maximum likelihood or bayesian estimation. However, both methodologies involve computing the likelihood of the observations $(Y_t)_{t=1}^{T}$ given θ, which in general is also not available in closed-form.

An interesting solution to this problem has been proposed in [3], in which the main idea is to directly approximate the score vector and observed information matrix and use these to estimate θ via a gradient-based method, such as the Newton-Raphson algorithm. Two algorithms have been proposed to accomplish this, the first having a linear computational cost in the number of particles N but in which the approximations suffer from a variance that increases at least quadratically over time and the second producing approximations with linearly increasing variances over time but suffering from a quadratic computational cost.

Recently, a major improvement of this methodology was proposed in [2], producing approximations with $\mathcal{O}(N)$ computational cost and variances that increase linear over time. This is basically made possible by combining a certain gaussian kernel density estimation approach to reduce particle degeneracy (and therefore estimator variance) with Rao-Blackwellisation to directly update the ensuing gaussian approximations (since a gaussian random variable is almost surely determined by its first two moments, the Rao-Blackwell theorem can be used in this context to obtain these quantities for the score and observed information matrix without additional sampling).

In this work, we illustrate substantial computational and parameter estimation efficiency gains that can be made with careful implementation of the above described methodology. In particular, we explore how different resampling devices, learning rates and filtering algorithms influence computing time and mean squared error of parameter estimates for a class of widely used nonlinear and non-gaussian state space models.

Referências

[1] Doucet A, De Freitas N, Gordon N. *Sequential Monte Carlo methods in practice.* Springer, New York, NY, 2001.

[2] Nemeth C, Fearnhead P, Mihaylova L. Particle approximations of the score and observed information matrix for parameter estimation in state–space models with linear computational cost. *Journal of Computational and Graphical Statistics*, Oct 1;25(4):1138-57, 2016.

[3] Poyiadjis G, Doucet A, Singh SS. Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika*, Feb 4;98(1):65-80, 2011.