Signal Processing 131 (2017) 77–91

Contents lists available at ScienceDirect

Signal Processing

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

Improving population Monte Carlo: Alternative weighting and resampling schemes

Víctor Elvira [a,*], Luca Martino [b], David Luengo [c], Mónica F. Bugallo [d]

[a] Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Leganés, Spain
[b] Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland
[c] Department of Signal Theory and Communications, Universidad Politécnica de Madrid, Madrid, Spain
[d] Department of Electrical and Computer Engineering, Stony Brook University, NY, USA

ARTICLE INFO

Article history:
Received 27 November 2015
Received in revised form
2 July 2016
Accepted 8 July 2016
Available online 12 July 2016

Keywords:
Population Monte Carlo
Adaptive importance sampling
Proposal distribution
Resampling

ABSTRACT

Population Monte Carlo (PMC) sampling methods are powerful tools for approximating distributions of static unknowns given a set of observations. These methods are iterative in nature: at each step they generate samples from a proposal distribution and assign them weights according to the importance sampling principle. Critical issues in applying PMC methods are the choice of the generating functions for the samples and the avoidance of the sample degeneracy. In this paper, we propose three new schemes that considerably improve the performance of the original PMC formulation by allowing for better exploration of the space of unknowns and by selecting more adequately the surviving samples. A theoretical analysis is performed, proving the superiority of the novel schemes in terms of variance of the associated estimators and preservation of the sample diversity. Furthermore, we show that they outperform other state of the art algorithms (both in terms of mean square error and robustness w.r.t. initialization) through extensive numerical simulations.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Bayesian signal processing, which has become very popular over the last years in statistical signal processing, requires computing distributions of unknowns conditioned on observations (and moments of them). Unfortunately, these distributions are often impossible to obtain analytically in many real-world challenging problems. An alternative is then to resort to Monte Carlo (MC) methods, which approximate the target distributions with random measures composed of samples and associated weights [1].

A well-known class of MC methods are those based on the adaptive importance sampling (AIS) mechanism, such as Population Monte Carlo (PMC) algorithms [2,3], which have been used in missing data, tracking, biological applications, among others [4–8]. In these methods, a population of probability density functions (pdfs) is adapted for approximating a target distribution through an iterative importance sampling procedure. AIS is often preferred to other MC schemes, such as Markov Chain Monte Carlo (MCMC), since they present several advantages. On the one hand, all the generated samples are employed in the estimation (e.g., there is no "burn-in" period). On the other hand, the corresponding adaptive schemes are more flexible, since they present fewer theoretical issues than adaptive MCMC algorithms. Namely, the convergence of AIS methods can usually be guaranteed under mild assumptions regarding the tails of the distributions and the stability of the adaptive process, whereas adaptive MCMC schemes must be designed very carefully, since the adaptation procedure can easily jeopardize the ergodicity of the chain (e.g., see [9] or [1, Section 7.6.3]).

The most characteristic feature in PMC [3] is arguably the use of resampling procedures for adapting the proposal pdfs (see for instance [10] for a review of resampling methods in particle filtering). The resampling step is a fast, often dimensionality-free, and an easy way of adapting the proposal pdfs by using information about the target. However, resampling schemes present some important drawbacks, such as the sample impoverishment. At the resampling step, the proposal pdfs with poor performance (i.e., with low associated weights) are likely to be removed, thus yielding a reduction of diversity. Since the publication of the standard PMC [3], several variants have been considered, partly in an attempt to mitigate this issue. In the D-kernel algorithm [11,12], the PMC kernel is a mixture of different kernels and the weights of the mixture are iteratively adapted in an implicit expectation-maximization (EM) algorithm. This procedure is refined through a double Rao–Blackwelization in [13]. The mixture population Monte Carlo algorithm (M-PMC) proposed in [14] also adapts a

* Corresponding author.

http://dx.doi.org/10.1016/j.sigpro.2016.07.012
0165-1684/© 2016 Elsevier B.V. All rights reserved.

mixture of proposal pdfs (weights and parameters of the kernels). The M-PMC belongs to the family of AIS methods, since it iteratively draws the samples from the mixture that is updated at every iteration without an explicit resampling step. Since drawing from the mixture can be interpreted as an implicit multinomial resampling, this method retains some similarities with the standard PMC scheme. A nonlinear transformation of the importance weights in the PMC framework has also been proposed in [15]. Other sophisticated AIS schemes, such as the AMIS [16] and the APIS [17] algorithms, have been recently proposed in the literature.

In this paper, we study three novel PMC schemes that improve the performance of standard PMC approach by allowing a better exploration of the space of unknowns and by reducing the variance of the estimators. These alternatives can be applied within some other sophisticated AIS approaches as well, such as the SMC samplers [18]. For this reason, we mainly compare them with the standard PMC [3], since the novel schemes could be automatically combined with the more sophisticated AIS techniques.

First of all, we introduce an alternative form of the importance weights, using a mixture of the proposal pdfs in the denominator of the weight ratio. We provide an exhaustive theoretical analysis, proving the unbiasedness and consistency of the resulting estimator, and showing the reduction in the variance of the estimator w.r.t. the estimator obtained using the standard weights. We also prove that the use of this mixture decreases the averaged mismatch between the numerator (target) and the function in the denominator of the IS weight in terms of L_2 distance. Moreover, we test this alternative scheme in different numerical simulations, including an illustrative toy example in Section 5.1, showing its practical benefit.

In the second proposed scheme, we generate several samples from every proposal pdf (not only one, as in PMC) and then we resample them jointly (all the samples at once, keeping fixed the total number of proposal pdfs). In the third proposed scheme, we consider again the generation of several samples from every proposal pdf, but the resampling is performed separately on the set of samples coming from each proposal, therefore guaranteeing that there will be exactly one representative from each of the individual mixture components in the random measure.

We show, through extensive computer simulations in several different scenarios, that the three newly proposed variants provide a substantial improvement compared to the standard PMC. In addition, we test the proposed variants on a standard implementation of the SMC samplers [18], showing also an improvement of the performance. On the one hand, they yield unbiased estimators with a reduced variance, as also proved theoretically. On the other hand, they outperform the standard PMC in terms of preservation of sample diversity and robustness w.r.t. initialization and parameter choice.

2. Problem statement

Let us consider the variable of interest, $\mathbf{x} \in \mathbb{R}^{D_x}$, and let $\mathbf{y} \in \mathbb{R}^{D_y}$ be the observed data. In a Bayesian framework, the posterior probability density function (pdf), here referred to as *target*, contains all the information about the parameters of interest and is defined as

$$\bar{\pi}(\mathbf{x}|\mathbf{y}) = \frac{\ell(\mathbf{y}|\mathbf{x})p_0(\mathbf{x})}{Z(\mathbf{y})} \propto \pi(\mathbf{x}) = \ell(\mathbf{y}|\mathbf{x})p_0(\mathbf{x}), \tag{1}$$

where $\ell(\mathbf{y}|\mathbf{x})$ is the likelihood function, $p_0(\mathbf{x})$ is the prior pdf, and $Z(\mathbf{y})$ is the model evidence or partition function (useful in model selection).[1] The goal is to compute some moment of $\mathbf{x}$, i.e., an integral measure w.r.t. the target pdf,

$$I = \frac{1}{Z}\int f(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \tag{2}$$

where f can be any square integrable function of $\mathbf{x}$ w.r.t. $\pi(\mathbf{x})$, and $Z = \int \pi(\mathbf{x})d\mathbf{x}$.[2]

In many practical applications, both the integral (2) and Z cannot be obtained in closed form and must be approximated. Importance sampling methods allow for the approximation of both quantities by a set of properly weighted samples.

3. Population Monte Carlo (PMC)

3.1. Description of the original PMC algorithm

The PMC method [3] is a well-known iterative adaptive importance sampling technique. At each iteration it generates a set of N samples $\{\mathbf{x}_i^{(t)}\}_{i=1}^N$, where t denotes the iteration number and i denotes the sample index. In order to obtain the samples, the original PMC algorithm makes use of a collection of proposal densities $\{q_i^{(t)}(\mathbf{x})\}_{i=1}^N$, with each sample being drawn from a different proposal, $\mathbf{x}_i^{(t)} \sim q_i^{(t)}(\mathbf{x})$ for $i = 1, \ldots, N$. Then, they are assigned an importance weight, computed as $w_i^{(t)} = \frac{\pi(\mathbf{x}_i^{(t)})}{q_i^{(t)}(\mathbf{x}_i^{(t)})}$, i.e., the weight of a particular sample represents the ratio between the evaluation, at the sample value, of the target distribution and the evaluation at the sample value of the proposal used to generate it. The method proceeds iteratively (up to the maximum iteration step considered, T), building a global importance sampling estimator using different proposals at every iteration. The new proposals are obtained by updating the set of proposals in the previous iteration.

There are two key issues in the application of PMC methods: the adaptation of the proposals from iteration to iteration and the way resampling is applied. The latter is critical to avoid the degeneracy of the random measure, i.e., to avoid a few particles having extremely large weights and the rest negligible ones [1,19]. Through the resampling procedure one selects the most promising streams of samples from the first iteration up to the current one. Several resampling procedures have been proposed in the literature [20,21]. In the standard PMC [3], multinomial resampling is the method of choice, and consists of sampling N times from the discrete probability mass defined by the normalized weights. As a result of this procedure, the new set of parameters used to adapt the proposals for the generation of samples in the next iteration is selected. In summary, the standard PMC technique consists of the steps shown in Table 1.

3.2. Estimators and consistency

All the generated samples can be used to build a global approximation of the target. This can be done by first normalizing all the weights from all the iterations,

$$\bar{\rho}_i^{(t)} = \frac{w_i^{(t)}}{\sum_{\tau=1}^{t}\sum_{j=1}^{N} w_j^{(\tau)}}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, N, \tag{8}$$

and then providing the pairs $\{\mathbf{x}_i^{(t)}, \bar{\rho}_i^{(t)}\}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. This procedure to compute the weights is equivalent to applying a static importance sampling technique that considers NT different proposal pdfs and all the corresponding samples. If the

[1] From now on, we remove the dependence on $\mathbf{y}$ in order to simplify the notation.

[2] Let us recall that $f(\mathbf{x})$ is square integrable w.r.t. $\pi(\mathbf{x})$ if $f(\mathbf{x}) \in L_\pi^2$, i.e., if $\int_{\mathcal{X}} f(\mathbf{x})^2\pi(\mathbf{x})d\mathbf{x} < \infty$.

Download English Version:

https://daneshyari.com/en/article/6958038

Download Persian Version:

https://daneshyari.com/article/6958038

Daneshyari.com