Nouveaux schémas de pondération et de rééchantillonnage en population Monte Carlo

Víctor ELVIRA[1], Luca MARTINO[2], David LUENGO[3], Mónica BUGALLO[4]

[1]IMT Lille Douai & CRIStAL (UMR CNRS 9189).

[2]Image Processing Laboratory, University of Valencia (Spain).

[3]Dept. of Signal Theory and Communications, Universidad Politécnica de Madrid (Spain).

[4]Dept. of Electrical and Computer Eng., Stony Brook University (USA).

victor.elvira@imt-lille-douai.fr

Résumé – Dans cet article, nous nous concentrons sur de nouveaux schémas de population Monte Carlo présentant une flexibilité et des performance supérieures par rapport à l'approche standard. Ces nouveaux schémas combinent différentes stratégies de pondération et de rééchantillonnage afin d'obtenir des performances plus efficaces au prix d'un incrément de complexité de calcul raisonnable. Les simulations numeriques comparent les différentes alternatives lorsqu'elles sont appliquées au problème d'estimation de fréquence pour des sinusoïdes superposées.

Abstract – In this paper we focus on novel population Monte Carlo schemes that present added flexibility and superior performance compared to the standard approach. The new schemes combine different weighting and resampling strategies to achieve more efficient performance at the cost of a reasonable computational complexity increment. Computer simulations compare the different alternatives when applied to the problem of frequency estimation in superimposed sinusoids.

1 Introduction

Many practical statistical signal processing problems demand the use of Monte Carlo (MC) techniques for drawing random samples from probability distributions with non-standard forms in order to approximate intractable integrals [1, 2]. Among the MC methodologies, importance sampling (IS) and Markov chain Monte Carlo (MCMC) algorithms are two of the most popular approaches [3, 4, 5]. The MC methods generate samples from a proposal probability density function (pdf) and use them in order to approximate the target pdf (e.g. in Bayesian inference the target is the posterior distribution). The performance of the MC methods typically depends on the discrepancy between the proposal and target pdfs. The key problem is that good proposal functions are very hard to choose in advance, and adaptive procedures must be performed [6, 7, 8].

Adaptive importance sampling (AIS) algorithms have received a particular attention in recent years [9, 10, 11, **?**, 12]. The reason is that, with these algorithms, the proposal pdf is adapted online in order to reduce the mismatch (according to some discrepancy measure) with the target density. Moreover, the consistency of AIS schemes is easily guaranteed under weak IS conditions [1, 2, 6, 8], unlike adaptive MCMC (A-MCMC) algorithms where a careful theoretical study must be considered [13, 7, 14, 15]. Additionally, with AIS methods, it is possible to easily estimate the normalizing constant of the posterior target distribution (also called Bayesian evidence or marginal likelihood) [1, 2, 6]. This quantity is particularly useful, for instance, in model selection problems [16, 17].

The population Monte Carlo (PMC) methods are a specific family of AIS algorithms [18, 19, 20] used in a variety of problems such as missing data, tracking, or biological applications, among others [21, 16]. One important feature of the PMC algorithms is the use of the resampling step for adapting the proposal pdfs. The resampling is a fast, often dimensionality-free, procedure. However, the resampling schemes present some important drawbacks, such as the sample impoverishment.

In this work, we describe a generic and novel framework for PMC methodologies that generalizes the standard scheme [18]. Moreover, the new framework is presented under a flexible structure that can result in novel PMC algorithms with alternative weighting and resampling strategies [22]. Here we provide novel PMC algorithms, while some others will be discussed in future works.

The paper is structured as follows. Next section introduces the problem statement and Section 3 an overview of the standard PMC method. Section 4 discusses the proposed method with alternative novel weighting and resampling schemes. Finally, Section 5 presents some numerical results that show the validity of the new approach and the last section concludes the paper with closing remarks and a discussion about possible improvements of the method.

2 Problem formulation

In many applications, the goal consists of computing some moment of the r.v. $\mathbf{X}$, that is distributed according to a possibly complicated pdf, $\tilde{\pi}(\mathbf{x})$. The moment of interest is

$$I = \int_{\mathcal{X}} f(\mathbf{x})\tilde{\pi}(\mathbf{x})d\mathbf{x} = \frac{1}{Z}\int_{\mathcal{X}} f(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \qquad (1)$$

where f can be any integrable function w.r.t. $\pi(\mathbf{x})$. In many scenarios, Eq. (1) cannot be computed or the target pdf can be evaluated only up to a constant. In that case, the target pdf can be expressed as $\tilde{\pi}(\mathbf{x}) = \frac{\pi(\mathbf{x})}{Z}$, where $\pi(\mathbf{x})$ is the unnormalized target function that can be evaluated and Z is the unknown partition function. Due to the aforementioned difficulties, Eq. (1) cannot be generally computed and Monte Carlo methods are used to approximate the moment of interest, I.

3 Standard population Monte Carlo

The PMC algorithm was proposed in [18] and constitutes a possible implementation of the AIS methodology. For sake of conciseness, we introduce a generic PMC algorithm in Table 1 that allows us to present first the standard PMC of [18], and later the novel modifications proposed in this paper.

In the standard PMC, the parameters of N proposal densities are initialized. In particular, location parameters $\{\boldsymbol{\mu}_i^{(1)}\}_{i=1}^N$ will be adapted, while the parameters $\{\mathbf{C}_i\}_{i=1}^N$ are static (e.g., with Gaussian proposal densities, the location and the static parameters are the mean vectors and the covariance matrices, respectively). The standard PMC works as follows. At each iteration, in Eq. (2), exactly $K = 1$ sample is drawn from each proposal, and its weight is computed in Eq. (3). In the standard PMC, this weight is $w_i^{(t)} = \frac{\pi(\mathbf{x}_i^{(t)})}{q_i^{(t)}(\mathbf{x}_i^{(t)}|\boldsymbol{\mu}_i^{(t)},\mathbf{C}_i)}, i = 1,\dots,N$. The N location parameters for the next iteration are i.i.d. drawn from the random measure $\hat{\pi}_t^N(\mathbf{x}) = \sum_{i=1}^N \bar{w}_i^{(t)}\delta(\mathbf{x}-\mathbf{x}_i^{(t)})$.

The standard PMC algorithm presents some drawbacks. For instance, the aforementioned standard IS weights have been shown to be suboptimal in terms of variance of the IS estimator [23]. Moreover, the standard PMC suffers from sample degeneracy problems, which can seriously affect the diversity [20, 22].

4 Alternative population Monte Carlo

Table 1 describes the novel framework, and here we summarize the differences w.r.t. the standard PMC. First, note that more than one sample per proposal can be drawn in Eq. (2). The advantage of this strategy is that proposals in good areas have several opportunities to draw samples with larger weights, which increases the survival chances of those proposals. Second, different IS weights can be used in Eq. (3). Finally, note in steps 2(c)-(e) of Table 1, different resampling schemes can be performed for adapting the parameters. In particular, the NK samples at each iteration are clustered into M_t disjoint sets, and the resampling step is performed at each subset. Below, we present alternative IS weighting schemes and two specific ways of performing the partition. Note the parallelism with the static partial multiple IS approach proposed in [24].

TABLE 1 – Generic Population Monte Carlo.

1. **[Initialization]** : Select the adaptive $\mathcal{P}^{(1)} = \{\boldsymbol{\mu}_1^{(1)}, ..., \boldsymbol{\mu}_N^{(1)}\}$ and the static parameters, $\{\mathbf{C}_i\}_{i=1}^N$ of the N proposals.
2. **[For $t = 1$ to T]** :
   (a) Draw K samples from each proposal pdf,
   $$\mathbf{x}_{i,k}^{(t)} \sim q_i^{(t)}(\mathbf{x}|\boldsymbol{\mu}_i^{(t)}, \mathbf{C}_i), \qquad (2)$$
   with $i = 1,\dots,N$ and $k = 1,\dots,K$.
   (b) Compute the importance weights,
   $$w_{i,k}^{(t)} = \frac{\pi(\mathbf{x}_{i,k}^{(t)})}{\Phi_{i,k}^{(t)}(\mathbf{x}_{i,k}^{(t)})}, \qquad (3)$$
   where $\Phi_{i,k}^{(t)}$ is a suitable function (see Section 4).
   (c) Choose a partition of the bidimensional set $\{1,\dots,N\} \times \{1,\dots,K\}$, formed by M_t disjoint subsets of pairs of indices $\mathcal{S}_m = \{i_{m,r}, k_{m,r}\}_{r=1}^{R_m}$ of samples, i.e.,
   $$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \ldots \cup \mathcal{S}_{M_t} = \{1,\dots,N\} \times \{1,\dots,K\},$$
   and clearly $\sum_{m=1}^{M_t} R_m = NK$ (note that $|\mathcal{S}_m| = R_m$).
   (d) Compute the normalized weights in each subset, i.e.,
   $$\bar{w}_{i,k}^{(t)} = \frac{w_{i,k}^{(t)}}{\sum\limits_{\{j,h\}\in\mathcal{S}_m} w_{j,h}^{(t)}}, \quad \{i,k\} \in \mathcal{S}_m, \qquad (4)$$
   and $m = 1,\dots,M_t$.
   (e) Considering the subset of particles $\mathcal{X}_m = \{\mathbf{x}_{i,k}^{(t)}\}$ with $\{i,k\} \in \mathcal{S}_m$, generate N new parameters $\boldsymbol{\mu}_i^{(t+1)}$ drawing i.i.d. samples within each set $\mathcal{X}_m$ according to the normalized weights $\bar{w}_{i,k}^{(t)}$, for each $\{i,k\} \in \mathcal{S}_m$ and $m = 1,\dots,M_t$.
3. **[Output, $t = T$]** : Return the pairs $\{\mathbf{x}_{i,k}^{(t)}, w_{i,k}^{(t)}\}$, for $i = 1,\dots,N$, $k = 1,\dots,K$ and $t = 1,\dots,T$.

4.1 Alternative weighting schemes

Note that the IS weights of the standard PMC imply that $\Phi_{i,k} = q_i$ in Eq. (3). However, when multiple proposals are available, many choices of $\Phi_{i,k}$ can be used (see [23, Section 4] for a detailed explanation). In particular, the so-called deterministic mixture (DM) weights present the best performance in terms of the variance of the estimators [23, Theorem 1, Theorem 2]. Therefore, we propose to use the weights

$$w_{i,k}^{(t)} = \frac{\pi(\mathbf{x}_{i,k}^{(t)})}{\frac{1}{N}\sum_{j=1}^N q_j^{(t)}(\mathbf{x}_{i,k}^{(t)}|\boldsymbol{\mu}_j^{(t)}, \mathbf{C}_j)}, \qquad (5)$$

i.e., $\Phi_{i,k}(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^N q_j^{(t)}(\mathbf{x}|\boldsymbol{\mu}_j^{(t)}, \mathbf{C}_j)$ is the mixture of the N proposals. With $K = 1$ and the trivial partition of all samples in the same set, the method is called deterministic mixture PMC (DM-PMC).

4.2 Alternative resampling schemes

While the description of the generic framework of Table 1 is very flexible, here we provide specific examples where all subsets have the same number of elements. Figure 1 illustrates two different resampling schemes. For each scheme, we indicate with a box the elements that participate in the corresponding resampling strategy.

4.2.1 Global resampling (GR-PMC)

The location parameters of next iteration are i.i.d. sampled from the the set of all KN samples given by

$$\mathcal{X}^{(t)} = \{\mathbf{x}_{1,1}^{(t)}, ..., \mathbf{x}_{1,K}^{(t)}, ..., \mathbf{x}_{N,1}^{(t)}, ..., \mathbf{x}_{N,K}^{(t)}\}$$

with probabilities corresponding to the associated normalized weights

$$\bar{w}_{i,k}^{(t)} = \frac{w_{i,k}^{(t)}}{\sum_{j=1}^{N}\sum_{\ell=1}^{K} w_{j,\ell}^{(t)}} \tag{6}$$

for $i = 1, ..., N$ and $k = 1, ..., K$.

4.2.2 Local resampling (LR-PMC)

Under this scheme, the $\boldsymbol{\mu}_i^{(t+1)}$ is resampled from the set of K samples generated by the proposal located at $\boldsymbol{\mu}_i^{(t)}$, i.e., from the set

$$\mathcal{X}_i^{(t)} = \{\mathbf{x}_{i,1}^{(t)}, ..., \mathbf{x}_{i,K}^{(t)}\} \tag{7}$$

using the multinomial probability mass function with probabilities

$$\bar{w}_{i,k}^{(t)} = \frac{w_{i,k}^{(t)}}{\sum_{\ell=1}^{N} w_{i,\ell}^{(t)}}, \qquad k = 1, \dots, K. \tag{8}$$

where $w_{i,k}^{(t)}$ denote the unnormalized weights. Therefore, in this case there is no loss of diversity since exactly one sample per proposal survives at each iteration.

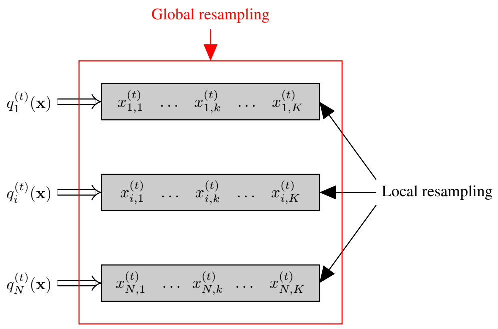

FIGURE 1 – Sketch of the global and local resampling schemes with N proposals at the t-th iteration, $q_i^{(t)}$ for $i = 1, \dots, N$ and $t = 1, \dots, T$, and K samples per proposal.

Method		$P = 5$ obs.	$P = 10$ obs.	$P = 50$ obs.
standard PMC		0.00275	0.0057	0.0218
GR-PMC	$K = 5$	0.00022	0.0219	0.0235
	$K = 10$	0.00008	0.0234	0.0246
	$K = 50$	0.00006	0.0173	0.0517
LR-PMC	$K = 5$	**0.00004**	0.0008	0.0265
	$K = 10$	0.00006	**0.0003**	0.0238
	$K = 50$	0.00009	0.0009	**0.0183**

TABLE 2 – Relative MSE in the estimation of the mean of $\bar{\pi}$.

5 Numerical Results

In order to show the advantages of the proposed methodologies, we consider the problem of estimating the frequencies of a weighted sum of sinusoids. The observation is given by

$$y_c(\tau) = A_0 + \sum_{i=1}^{S} A_i \cos(2\pi f_i \tau + \phi_i) + r(\tau), \quad \tau \in \mathbb{R},$$

where S is the number of sinusoids, A_0 is a constant term, $\{A_i\}_{i=1}^S$ is the set of amplitudes of the sinusoids, $\{f_i\}_{i=1}^S$ represents the set of frequencies, $\{\phi_i\}_{i=1}^S$ are the phases, and $r(\tau)$ are i.i.d. Gaussian samples. Here we focus on the inference of the set of frequencies $\{f_i\}_{i=1}^S$, which can be a problem of interest in many applications in different fields such as signal processing , control, or digital communications (see [25] and the references therein). Let us assume that we have P equally spaced data obtained discretizing $y_c(\tau)$ with period $T_s < \frac{\pi}{\max_{1 \le i \le S} 2\pi f_i}$ fulfilling the sampling theorem [26],

$$y[p] = A_0 + \sum_{i=1}^{S} A_i \cos(\Omega_i k + \phi_i) + r[p], \quad p = 1, \dots, P,$$

where $y[p] = y_c(pT_s)$ for $p = 0, 1, \dots, P-1$, $\Omega_i = 2\pi f_i T_s$ for $i = 1, \dots, S$, and $r[p] \sim \mathcal{N}(0, \sigma_w^2)$. Therefore, considering the hypercube $\mathcal{D} = \left[0, \frac{1}{2}\right]^S$ as domain of the target (note that is periodic outside $\mathcal{D}$), and a uniform prior on $\mathcal{D}$, the posterior given the data is $\bar{\pi}(\mathbf{x}) \propto \exp(-V(\mathbf{x}))$, where

$$V(\mathbf{x}) = \frac{1}{2\sigma_w^2} \sum_{k=1}^{K} \left(y[k] - A_0 - \sum_{i=1}^{S} A_i \cos(x_i k + \phi_i) \right)^2 \mathbb{I}_{\mathcal{D}}(\mathbf{x}).$$

We tackle a bi-dimensional problem with $S = 2$ sinusoids where we consider that the values of the frequencies $\{f_i\}_{i=1}^2 = [0.27 \;\; 0.43]$ are unknown. We set $A_0 = A_1 = A_2 = 1$ and $\phi_1 = \phi_2 = 0$. The problem consists on characterizing the posterior pdf of the frequencies $\{f_i\}_{i=1}^2$ given the data. To that end, we run the different algorithms presented above. In all cases, we use isotropic Gaussian distributions with $\boldsymbol{C}_i = \sigma^2 \mathbf{I}_2$ for $i = 1, ..., N$ where $\sigma = 0.05$. The means of the proposals are initialized randomly and uniformly in $\mathcal{D}$. For the proposed algorithms we set $K \in \{5, 10, 50\}$, and we always set $N = 10$. In all cases, we select T in such a way the number of target evaluations is fixed to $E = NKT = 2 \cdot 10^4$. Table 1 shows the relative mean square error (MSE) in the estimation of the mean of the target. Note that the proposed algorithms outpeform the standard PMC. Finally, note that the relative MSE increases for large P since the target pdf is sharper, and hence the adaptive problem is more challenging.

6 Conclusions

In this paper, we have presented a flexible framework of PMC algorithms that includes several methods in the literature and opens the door for new methods. We have discussed alternative weighting and resampling schemes, giving specific examples of PMC algorithms. Finally, we have tested the PMC algorithms in a problem of frequency estimation in superimposed sinusoids embedded in Gaussian noise. Future work includes the study of novel parallelizable PMC methods within the proposed framework and numerical results in high-dimensional examples with real data.

Références

[1] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 2004.

[2] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, 2004.

[3] A. Doucet and X. Wang, "Monte Carlo methods for signal processing : a review in the statistical signal processing context," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 152–170, 2005.

[4] J. K. O. Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Proc.*, Springer, 1996.

[5] W. J. Fitzgerald, "Markov chain Monte Carlo methods with applications to signal processing," *Signal Processing*, vol. 81, no. 1, pp. 3–18, January 2001.

[6] M. F. Bugallo, L. Martino, and J. Corander, "Adaptive importance sampling in signal processing," *Digital Signal Processing*, vol. 47, pp. 36–49, 2015.

[7] C. Andrieu and J. Thoms, "A tutorial on adaptive MCMC," *Statistics and Computing*, vol. 18, no. 4, pp. 343–373, 2008.

[8] L. Martino, V. Elvira, D Luengo, and F. Louzada, "Adaptive Population Importance Samplers : A General Perspective," *IEEE Sensor Array and Multichannel Signal Processing Workshop, (SAM)*, 2016.

[9] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler," *IEEE International Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 8088–8092, 2014.

[10] J. M. Cornuet, J. M. Marin, A. Mira, and C. P. Robert, "Adaptive multiple importance sampling," *Scand. Journal of Stat.*, vol. 39, no. 4, pp. 798–812, December 2012.

[11] V. Elvira, L. Martino, L. Luengo, and J. Corander, "A gradient adaptive population importance sampler," in *IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 4075–4079.

[12] I. Schuster, "Gradient importance sampling," *arXiv :1507.05781*, pp. 313–316, 2015.

[13] Y. F. Atchade and J.S. Rosenthal, "On adaptive Markov chain Monte Carlo algorithms," *Bernoulli*, vol. 11, no. 5, pp. 815–828, 2005.

[14] H. Haario, E. Saksman, and J. Tamminen, "An adaptive Metropolis algorithm," *Bernoulli*, vol. 7, no. 2, pp. 223–242, April 2001.

[15] L. Martino, J. Read, and D. Luengo, "Independent doubly adaptive rejection Metropolis sampling within Gibbs sampling," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3123–3138, 2015.

[16] M. Hong, M. F Bugallo, and P. M Djuric, "Joint model selection and parameter estimation by population monte carlo simulation," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 3, pp. 526–539, 2010.

[17] L. Martino, V. Elvira, and F. Louzada, "Weighting a Resampled Particle in sequential Monte Carlo," *IEEE Sensor Array and Multichannel Signal Processing Workshop, (SAM)*, 2016.

[18] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert, "Population Monte Carlo," *Journal of Comp. and Graphical Statistics*, vol. 13, no. 4, pp. 907–929, 2004.

[19] O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. P. Robert, "Adaptive importance sampling in general mixture classes," *Stat. and Comp.*, vol. 18, pp. 447–459, 2008.

[20] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, vol. 68, no. 3, pp. 411–436, 2006.

[21] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert, "Population Monte Carlo for ion channel restoration," *Journal of Comp. and Graphical Statistics*, 2002.

[22] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Improving Population Monte Carlo : Alternative weighting and resampling schemes," *Signal Processing*, vol. 131, no. 12, pp. 77–91, 2017.

[23] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Generalized multiple importance sampling," *preprint : arXiv :1511.03095*, 2015.

[24] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Efficient multiple importance sampling estimators," *Signal Proc. Letters, IEEE*, vol. 22, no. 10, pp. 1757–1761, 2015.

[25] P. M. Djurić, B. Shen, and M. F. Bugallo, "Population Monte Carlo methodology a la Gibbs sampling," in *European Signal Processing Conf. (EUSIPCO)*, 2011.

[26] J. G. Proakis, *Digital Communications*, McGraw-Hill, Singapore, 1995.