A NEW STRATEGY FOR EFFECTIVE LEARNING IN POPULATION MONTE CARLO SAMPLING

Mónica F. Bugallo[◇]*, Víctor Elvira*[⋆]*, and Luca Martino*[†]

[◇] Dept. of Electrical and Computer Engineering, Stony Brook University, New York (USA)
[⋆] Télécom Lille, CRIStAL (UMR 9189) (France)
[†] Dept. of Electrical Engineering, Universidad de Valencia (Spain)

ABSTRACT

In this work, we focus on advancing the theory and practice of a class of Monte Carlo methods, population Monte Carlo (PMC) sampling, for dealing with inference problems with static parameters. We devise a new method for efficient adaptive learning from past samples and weights to construct improved proposal functions. It is based on assuming that, at each iteration, there is an intermediate target and that this target is gradually getting closer to the true one. Computer simulations show and confirm the improvement of the proposed strategy compared to the traditional PMC method on a simple considered scenario.

***Index Terms*—** Importance sampling, Monte Carlo methods, population Monte Carlo.

1. INTRODUCTION

In Bayesian signal processing, all the information about the unknowns of interest is contained in their posterior distributions. The unknowns can be parameters of a model, or a model and its parameters. In many important problems, these distributions are impossible to obtain in analytic form. An alternative is to generate their approximations by Monte Carlo-based methods [1, 2]. In this work, we focus on Population Monte Carlo (PMC) sampling methods [3, 4] for dealing with problems with static parameters. PMC methodology belongs to the adaptive importance sampling (AIS) family [5, 6, 7, 8, 9, 10] and has been an active area of research for more than a decade since the publication of [3]. Since then, several variants have been proposed: the D-kernel algorithm [11, 12], the mixture population Monte Carlo algorithm [13], the nonlinear PMC [14] or the DM-PMC, GR-PMC and LR-PMC [4]. Other sophisticated AIS schemes have been recently proposed in the literature, e.g. the AMIS [15, 16, 17], APIS [18, 19] and LAIS [20, 21, 22] methods.

PMC is based on importance sampling, which amounts to generating samples from a selected distribution called instrumental, importance or proposal probability density function (pdf). More specifically, a set of proposal pdf s is considered. These distributions are different from the target distribution because we assume that direct sampling from the target distribution is unfeasible. Good proposal functions are the ones that are close to the target distribution. Once the samples are generated from the set of proposal pdfs, they are assigned weights. The samples with their associated weights represent an approximation of the target distribution. This approximation is used for adapting the location parameters of the proposal pdfs. The process repeats and the proposal functions keep adapting as we proceed with the iterations, which is why the PMC methodology belongs to the AIS family. In this process, learning takes place from samples and weights obtained in previous iterations.

In this paper, we devise a new method for efficient adaptive learning from past samples and weights to construct improved proposal functions. For the adaptation step, tempered target and weights are considered for improving the state space exploration. For the construction of the estimators, the standard importance weights are applied (differently from [23, 24]), so that the consistency is not jeopardized. The new method will be particularly useful for current signal processing challenges concerning complex systems. For instance, With large numbers of unknowns and/or data, the resulting target distributions are extremely peaky and very difficult to approximate. The novel PMC methodology uses a sequence of modified targets, which facilitates the adaptability of the algorithm and then improves its performance.

2. PROBLEM STATEMENT AND BASICS OF PMC

Suppose that we want to approximate a target distribution by a set of samples and weights. Here $\mathbf{x} \in \mathbb{R}^{d_x}$ is the unknown state of the system. The distribution of interest is most often a posterior, $\bar{\pi}(\mathbf{x}) = p(\mathbf{x}|\mathbf{y}_{1:N_\mathbf{y}})$, where $\mathbf{y}_n \in \mathbb{R}^{d_y}$ represents observations with information about $\mathbf{x}$, and $\mathbf{y}_{1:N_y} \equiv \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_{N_y}\}$. In general, we know $\bar{\pi}(\mathbf{x})$ only up to a normalizing constant, i.e., we can evaluate $\pi(\mathbf{x}) \propto \bar{\pi}(\mathbf{x})$. More precisely, we have

$$\pi(\mathbf{x}) = \ell(\mathbf{y}_{1:N_\mathbf{y}}|\mathbf{x})h(\mathbf{x}), \tag{1}$$

This work has been supported by the National Science Foundation under Award CCF-1617986 and by the European Research Council (ERC) through the ERC Consolidator Grant SEDAL ERC-2014-CoG 647423.

where ℓ is the likelihood function and h represents the prior pdf.

2.1. The basics of AIS

First, we review the concept of importance sampling [25, 1, 2]. Let the samples used for approximation be drawn from $\pi(\mathbf{x})$ itself and let them be denoted by $\mathbf{x}^{(m)}$, $m = 1, 2, \cdots, M$. The approximating distribution is

$$\widehat{\pi}^M(\mathbf{x}) \quad = \quad \sum_{m=1}^{M} \frac{1}{M} \delta(\mathbf{x}^{(m)}), \tag{2}$$

where $\delta(\mathbf{x}^{(m)})$ is the unit delta measure centered at $\mathbf{x}^{(m)}$. When it is difficult or impossible to draw samples from $\pi(\mathbf{x})$, the alternative is to use a proposal function $q(\mathbf{x})$, with a shape as close as possible to $\pi(\mathbf{x})$ and support larger than that of $\pi(\mathbf{x})$. If now $\mathbf{x}_0^{(m)} \sim q_0(\mathbf{x})$, then the approximation of $\pi(\mathbf{x})$ is

$$\widehat{\pi}_0^M(\mathbf{x}) \quad \propto \quad \sum_{m=1}^{M} \widehat{w}_0^{(m)} \delta(\mathbf{x}_0^{(m)}), \tag{3}$$

$$= \quad \sum_{m=1}^{M} \bar{w}_0^{(m)} \delta(\mathbf{x}_0^{(m)}), \tag{4}$$

where $\widehat{w}_0^{(m)} = \frac{\pi(\mathbf{x}_0^{(m)})}{q_0(\mathbf{x}_0^{(m)})}$ and $\bar{w}_0^{(m)} = \frac{w_0^{(m)}}{\sum_{j=1}^{M} w_0^{(j)}}$ are the unnormalized and normalized importance weights, respectively. The generation of samples from a proposal function and the corresponding assignment of weights is known as importance sampling (IS).

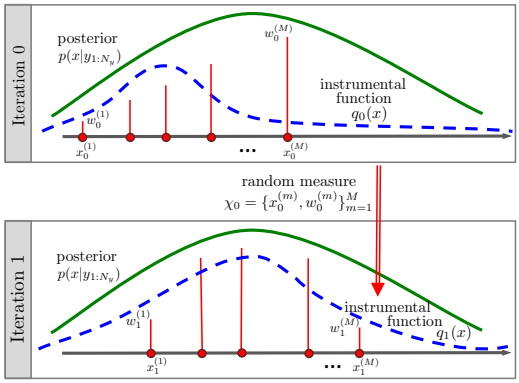


Fig. 1. AIS in two iterations.

The idea behind AIS is to use the random measure $\chi_0 = \{\mathbf{x}_0^{(m)}, w_0^{(m)}\}_{m=1}^M$ for creating a proposal function better than $q_0(\mathbf{x})$ (see Fig. 1 for a pictorial description of AIS). The new function, $q_1(\mathbf{x})$, is used to obtain a new random measure, $\chi_1 = \{\mathbf{x}_1^{(m)}, \widehat{w}_1^{(m)}\}$, which approximates the target by

$$\widehat{\pi}_1^M(\mathbf{x}) \quad \propto \quad \sum_{m=1}^{M} \widehat{w}_1^{(m)} \delta(\mathbf{x}_1^{(m)}), \tag{5}$$

where the subscript of $\widehat{\pi}_1^M(\mathbf{x})$ indicates the approximation of $\pi(\mathbf{x})$ at the first iteration. One can also combine the initial random measure χ_0 with χ_1 to improve the approximation (5), e.g., by

$$\widehat{\pi}_1^M(\mathbf{x}) \quad \propto \quad \sum_{m=1}^{M} \left(\widehat{w}_0^{(m)} \delta(\mathbf{x}_0^{(m)}) + \widehat{w}_1^{(m)} \delta(\mathbf{x}_1^{(m)}) \right). \tag{6}$$

The $2M$ weights must be normalized altogether in order to sum up to one. The process can obviously continue iteratively (in the sequel, we use the subscript i to denote iteration number). In the second iteration, one can use the random measure χ_1 and create yet an even better proposal function. Or, one can use both measures $\chi_{0:1}$ (here, $\chi_{0:1} \equiv \{\chi_0, \chi_1\}$) to obtain the new proposal function. The iterative process proceeds until a stopping condition is met.

Algorithm 1 PMC Sampling

1: **[Initialization]**: Select the adaptive parameters $\mathcal{P}_1 = \{\boldsymbol{\mu}_1^{(1)}, ..., \boldsymbol{\mu}_1^{(m)}\}$ and the static parameters, $\{\mathbf{C}^{(m)}\}_{m=1}^M$ of the M proposals.

2: **[For** $i = 1$ **to** I**]**:

(a) Draw one sample from each proposal pdf,

$$\mathbf{x}_i^{(m)} \sim q_i^{(m)}(\mathbf{x}|\boldsymbol{\mu}_i^{(m)}, \mathbf{C}^{(m)}), \qquad m = 1, \ldots, M.$$

(b) Compute the importance weights,

$$\widehat{w}_i^{(m)} = \frac{\pi(\mathbf{x}_i^{(m)})}{q_i^{(m)}(\mathbf{x}_i^{(m)}|\boldsymbol{\mu}_i^{(m)}, \mathbf{C}^{(m)})},$$

and normalize them, $\bar{w}_i^{(m)} = \frac{w_i^{(m)}}{\sum_{j=1}^{M} w_i^{(j)}}$, with $m = 1, \ldots, M$.

(c) Draw M independent parameters $\boldsymbol{\mu}_{i+1}^{(m)}$ from the discrete random measure, $\widehat{\pi}_i^M(\mathbf{x}) = \sum_{m=1}^{M} \bar{w}_i^{(m)} \delta(\mathbf{x} - \mathbf{x}_i^{(m)})$, obtaining the next population of parameters, $\mathcal{P}_{i+1} = \{\boldsymbol{\mu}_{i+1}^{(1)}, ..., \boldsymbol{\mu}_{i+1}^{(M)}\}$.

3: **[Output,** $i = I$**]**: Return the pairs $\{\mathbf{x}_i^{(m)}, \widehat{w}_i^{(m)}\}$ for $m = 1, \ldots, M$ and $i = 1, \ldots, I$.

2.2. Implementation of AIS: PMC sampling

The PMC algorithm constitutes a possible implementation of the AIS methodology. In its most basic implementation, it considers M proposals (as many as samples) that are iteratively adapted ($i = 1, \cdots, I$). In particular, location parameters $\{\boldsymbol{\mu}_i^{(m)}\}_{m=1}^M$ will be adapted, while the parameters $\{\mathbf{C}^{(m)}\}_{m=1}^M$ are static (e.g., with Gaussian proposal densities, the location and the static parameters are the mean vectors and the covariance matrices, respectively). At each iteration, exactly one sample is drawn from each proposal, and its weight

is computed. The M location parameters for the next iteration are drawn from the obtained discrete random measure. Algorithm 1 summarizes the basic PMC method.

3. PROPOSED METHOD

We investigate a novel strategy for effective adaptive learning based on assuming that the proposal function belongs to a family of mixtures with predefined kernels. The parameters of these kernels are learned from the generated samples and their weights as the iterations proceed. In the problems that we address, the likelihoods will be much more peaked than the priors, which implies that the posteriors will also be much more peaked.

Algorithm 2 PMC Sampling with Gradual Learning

[Initialization]: Select the adaptive parameters $\mathcal{P}_1 = \{\boldsymbol{\mu}_1^{(1)}, ..., \boldsymbol{\mu}_1^{(m)}\}$, the static parameters, $\{\mathbf{C}^{(m)}\}_{m=1}^M$ of the M proposals, and the sequence $[\lambda_1 < ... < \lambda_L]$.

[For $\ell = 1$ **to** L**]**:

[For $i = 1$ **to** I**]**:

(a) Draw one sample from each proposal pdf,

$$\mathbf{x}_i^{(m)} \sim q_i^{(m)}(\mathbf{x}|\boldsymbol{\mu}_i^{(m)}, \mathbf{C}^{(m)}), \qquad m = 1, \dots, M.$$

(b) Compute the *modified-tempered* weights,

$$w_i^{(m)} = \frac{\left(\ell(\mathbf{y}_{1:N_y}|\mathbf{x}_i^{(m)})\right)^{\lambda_\ell} h(\mathbf{x}_i^{(m)})}{q_i^{(m)}(\mathbf{x}_i^{(m)}|\boldsymbol{\mu}_i^{(m)}, \mathbf{C}^{(m)})},$$

and normalize them, $\bar{w}_i^{(m)} = \frac{w_i^{(m)}}{\sum_{j=1}^M w_j^{(m)}}$, with $m = 1, \dots, M$. Compute also the non-modified weights

$$\widehat{w}_i^{(m)} = \frac{\pi(\mathbf{x}_i^{(m)})}{q_i^{(m)}(\mathbf{x}_i^{(m)}|\boldsymbol{\mu}_i^{(m)}, \mathbf{C}^{(m)})},$$

with $m = 1, \dots, M$.

(c) Draw M independent parameters $\boldsymbol{\mu}_{i+1}^{(m)}$ from the discrete random measure, $\hat{\pi}_i^M(\mathbf{x}) = \sum_{m=1}^M \bar{w}_i^{(m)} \delta(\mathbf{x} - \mathbf{x}_i^{(m)})$, obtaining the new set of parameters $\mathcal{P}_{i+1} = \{\boldsymbol{\mu}_{i+1}^{(1)}, ..., \boldsymbol{\mu}_{i+1}^{(M)}\}$.

[Output, $i = I$**]**: Return the pairs $\{\mathbf{x}_i^{(m)}, \widehat{w}_i^{(m)}\}$ for $m = 1, \dots, M$ and $i = 1, \dots, I$.

This is especially the case when the number of data is large and/or the dimension of $\mathbf{x}$ is high, which will most likely entail that there will be no samples in parts of the space with large probability masses. To avoid that, we use the concept of gradual learning. Let us consider the first iteration $i = 1$. The underlying idea is that, *for the adaptation* (i.e., for the resampling step), we weight the generated samples by assuming that a *tempered* target proportional to the product $\left(\ell(\mathbf{y}_{1:N_y}|\mathbf{x})\right)^{\lambda_1} h(\mathbf{x})$ (this would be an "intermediate posterior" that we pursue), where λ_1 is some small positive number less than one. The weights are then given by Eq. (7). At the next iteration, the weight for the adaptation are computed by using as a target distribution $\left(\ell(\mathbf{y}_{1:N_y}|\mathbf{x})\right)^{\lambda_2} h(\mathbf{x})$, where $0 < \lambda_1 < \lambda_2 < 1$. By choosing a larger λ_2, we allow for more influence of the likelihood in the computation of the samples' weights.

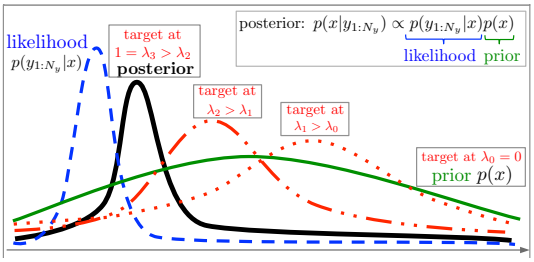


Fig. 2. New strategy for gradual learning.

Algorithm 2 summarizes the proposed methodology. The key point of the novel technique is the joint use of two importance weights:

- The standard importance weights,

$$\widehat{w}_i^{(m)} = \frac{\pi(\mathbf{x}_i^{(m)})}{q_i^{(m)}(\mathbf{x}_i^{(m)}|\boldsymbol{\mu}_i^{(m)}, \mathbf{C}^{(m)})},$$

  are used for inference purposes, i.e. for building the IS estimators.

- The modified-tempered weights

$$w_i^{(m)} = \frac{\left(\ell(\mathbf{y}_{1:N_y}|\mathbf{x}_i^{(m)})\right)^{\lambda_\ell} h(\mathbf{x}_i^{(m)})}{q_i^{(m)}(\mathbf{x}_i^{(m)}|\boldsymbol{\mu}_i^{(m)}, \mathbf{C}^{(m)})},$$

  are used for adaptation purposes. They are normalized as $\bar{w}_i^{(m)}$, $m = 1, \dots, M$, and used at the resampling step of each iteration.

A sequence of positive and increasing values is used to temper the target, i.e.,

$$0 < \lambda_1 < \lambda_2 < \ldots < \lambda_i \ldots < 1,$$

as shown in Figures 2. When $\lambda = 1$, the target distribution is the posterior and the corresponding samples and weights represent its approximation. We conjecture that in this way we can obtain a good proposal function for any dimension of $\mathbf{x}$. Obviously, the higher the dimension of $\mathbf{x}$ is, the more steps in learning the proposal function we need. We recall that the returned weights, used for building the IS estimators, $\widehat{w}_i^{(m)}$

are computed using the true unmodified target, so that the resulting IS estimators are always consistent. Note also that their calculation is almost free (only a new re-normalization is needed), since the modified weights of Eq. (7) have $\widehat{w}_i^{(m)}$ as intermediate result (i.e., the likelihood function is always evaluated for the tempered weights).

4. NUMERICAL RESULTS

In this section, we tackle the problem of estimating the frequencies of a weighted sum of sinusoids. The data are given by

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

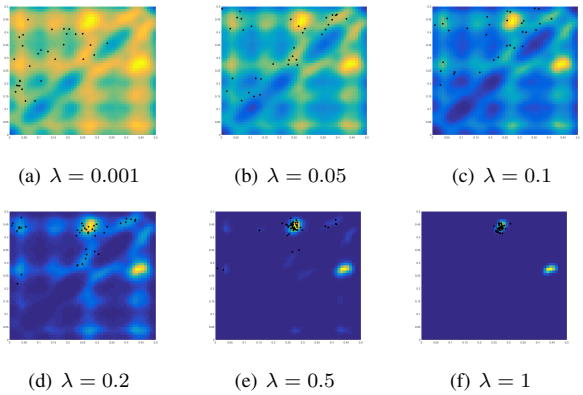
where S is the number of sinusoids, A_0 is a constant term, $\{A_i\}_{i=1}^S$ is the set of amplitudes of the sinusoids, $\{f_i\}_{i=1}^S$ represents the set of frequencies, $\{\phi_i\}_{i=1}^S$ are the phases, and $r(\tau)$ are i.i.d. zero-mean Gaussian samples with variance σ_w^2. Let us assume that we have N_y equally spaced samples $y_c(\tau)$ with period $T_s < \frac{\pi}{\max_{1\leq i\leq S} 2\pi f_i}$, i.e. fulfilling the sampling theorem [26],

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

where $y[p] = y_c(pT_s)$ for $p = 0, 1, \ldots, d_y - 1$, $\Omega_i = 2\pi f_i T_s$ for $i = 1, \ldots, S$, and $r[p] \sim \mathcal{N}(0, \sigma_w^2)$. The goal consists on inferring the set of unknown frequencies $\{f_i\}_{i=1}^S$. We set a uniform prior on $\mathcal{D}$, where $\mathcal{D} = \left[0, \frac{1}{2}\right]^S$ is the hypercube considered a domain of the target (which is periodic outside $\mathcal{D}$). Then the posterior given the data is $\bar{\pi}(\mathbf{x}) \propto \exp\left(-V(\mathbf{x})\right)$, 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 the address the bi-dimensional problem with $S = 2$ sinusoids where the true frequencies, $\{f_i\}_{i=1}^2 = [0.27\ 0.43]$, are unknown. For generating $N_y = 20$ observations, we set $A_0 = A_1 = A_2 = 1$ and $\phi_1 = \phi_2 = 0$. Therefore, the problem consists in characterizing the posterior pdf of the frequencies $\{f_i\}_{i=1}^2$ given the data. We use the DM-PMC method of [4], and we implement the gradual learning strategy proposed. We use a sequence of $L = 8$ modified targets, where we have set $\lambda_i = \Lambda(i)$ for $i = 1, ..., 8$, where $\Lambda = [0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 1]$. As proposals, we use isotropic Gaussian distributions with $\boldsymbol{C}_i = \sigma^2 \mathbf{I}_2$ for $i = 1, ..., N$ where $\sigma = 0.05$, and with means initialized randomly and uniformly in $\mathcal{D}$. We test the algorithms with $K \in \{1, 2, 5, 10\}$, and $N = 10$. In all cases, for sake of a fair comparison, we select T in such a way the number of target evaluations is fixed to $E = NKTL = 16000$ (note that for the standard PMC, $L = 1$). Table 1 shows the relative mean square error (RMSE) in the estimation of the true parameters. Note that the proposed gradual learning PMC sampling method outperforms the standard PMC for all sets of parameters. Moreover, when K is bigger the number of iterations T is smaller (since we keep $E = NKTL$ constant). Therefore, the new method is of special interest when few adaptive iterations can be performed. Finally, Fig. 3 depicts the evolution of the proposals as λ changes for the new method. It is obvious that when λ is small, the target is less peaky and exploring the space of unknowns is easier, while λ tends to one, the target gets more peaky. At the end all proposals move to the mode (true value is [0.27 0.43]).



(a) $\lambda = 0.001$ (b) $\lambda = 0.05$ (c) $\lambda = 0.1$

(d) $\lambda = 0.2$ (e) $\lambda = 0.5$ (f) $\lambda = 1$

Fig. 3. Sequence of different targets as λ changes. The black points show the evolution of the mean of the proposals.

Method		$K = 1$	$K = 2$	$K = 5$	$K = 10$
Standard PMC	$N = 5$	0.0521	0.0691	0.1504	0.1389
	$N = 10$	0.0862	0.1507	0.1298	0.1448
Gradual learning PMC	$N = 5$	0.0517	0.0513	0.0603	0.0574
	$N = 10$	0.0516	0.0632	0.0517	0.0687

Table 1. RMSE in the estimation of parameters $\{f_i\}_{i=1}^2$.

5. CONCLUSIONS

We discuss the implications of considering a new strategy for proposal generation in population Monte Carlo (PMC) sampling and compare the new method with the standard PMC in terms of performance. The new PMC approach is based on a gradual learning process that approximates the target more precisely with iterations. Computer simulations and comparisons reveal a good and advantageous performance of the proposed method when compared to the standard PMC. It is expected that in more challenging scenarios (higher dimension of the space of unknowns) the gain of the proposed algorithm will be more.

6. REFERENCES

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

[4] 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.

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

[6] V. Elvira, L. Martino, D. Luengo, and J. Corander, "A Grandient Adaptive Population Importance sampler," in *23th European Signal Processing Conference (EUSIPCO)*, 2015, pp. 1–5.

[7] G. H. Givens and A. E. Raftery, "Local adaptive importance sampling for multivariate densities with strong nonlinear relationships," *Journal of the American Statistical Association*, vol. 91, pp. 132–141, 1996.

[8] P. Zhang, "Nonparametric importance sampling," *Journal of the American Statistical Association*, vol. 91, pp. 1245–1253, 1996.

[9] L. Martino, V. Elvira, D. Luengo, and F. Louzada, "Adaptive population importance samplers: A general perspective," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2016, pp. 1–5.

[10] Z. I. Botev, P. L'Ecuyer, and B. Tuffin, "Markov chain importance sampling with applications to rare event probability estimation," *Statistics and Computing*, vol. 23, pp. 271–285, 2013.

[11] G. R. Douc, J.-M. Marin, and C. Robert, "Convergence of adaptive mixtures of importance sampling schemes," *Annals of Statistics*, vol. 35, pp. 420–448, 2007.

[12] G. R. Douc, J.-M. Marin, and C. Robert, "Minimum variance importance sampling via population Monte Carlo," *ESAIM: Probability and Statistics*, vol. 11, pp. 427–447, 2007.

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

[14] E. Koblents and J. Míguez, "A population Monte Carlo scheme with transformed weights and its application to stochastic kinetic models," *Statistics and Computing*, pp. 1–19, 2013.

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

[16] L. Pozzi and A. Mira, "A R adaptive multiple importance sampling (ARAMIS)," *Technical Report*, 2012.

[17] J.-M. Marin, P. Pudlo, and M. Sedki, "Consistency of the adaptive multiple importance sampling," *arXiv:1211.2548*, pp. 1–25, 2012.

[18] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.

[19] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler: Learning from the uncertanity," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4422–4437, 2015.

[20] L. Martino, V. Elvira, D. Luengo, and J. Corander, "MCMC-driven adaptive multiple importance sampling," in *Proceedings of the 12th Brazilian Meeting on Bayesian Statistics (EBEB)*, 2014.

[21] L. Martino, V. Elvira, D. Luengo, and J. Corander, "Layered adaptive importance sampling," *Statistics and Computing, to appear*, 2016.

[22] L. Martino, V. Elvira, D. Luengo, and J. Corander, "Interacting Parallel Markov Adaptive Importance Sampling," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015, pp. 1–5.

[23] 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.

[24] R. M. Neal, "Annealed importance sampling," *Statistics and Computing*, vol. 11, no. 2, pp. 125–139, 2001.

[25] B. D. Ripley, *Stochastic Simulation*, John Wiley & Sons, 1987.

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