Efficient Multiple Importance Sampling Estimators

Víctor Elvira, Luca Martino, David Luengo, and Mónica F. Bugallo

Abstract—**Multiple importance sampling (MIS) methods use a set of proposal distributions from which samples are drawn. Each sample is then assigned an importance weight that can be obtained according to different strategies. This work is motivated by the trade-off between variance reduction and computational complexity of the different approaches (classical vs. deterministic mixture) available for the weight calculation. A new method that achieves an efficient compromise between both factors is introduced in this letter. It is based on forming a partition of the set of proposal distributions and computing the weights accordingly. Computer simulations show the excellent performance of the associated *partial deterministic mixture* MIS estimator.**

Index Terms—**Adaptive importance sampling, deterministic mixture, Monte Carlo methods, multiple importance sampling.**

I. INTRODUCTION

IMPORTANCE SAMPLING (IS) methods approximate statistical moments of a variable of interest by sets of samples, drawn from a proposal distribution different from the targeted one, and weights, assigned to the samples in order to measure their adequacy in approximating the target [1], [2]. In its standard form, IS uses one proposal distribution from which all samples are drawn. However, a more powerful strategy to reduce the variance of the estimators consists of using a set of different proposal distributions. This is the basis of *multiple importance sampling* (MIS) techniques [3]. The samples drawn from the different proposals in MIS are then assigned weights proportional to the ratio between the target and the proposal densities evaluated at the sample values. Several strategies for calculating the weights have been considered, depending on the way in which the proposal evaluation in the denominator is interpreted. In the standard MIS, the evaluated proposal is exactly the one from which the sample was generated. This constitutes the simplest method in terms of computational complexity. A different approach, referred to as *deterministic mixture* (DM) MIS, interprets the set of generating proposals as a whole mixture distribution and calculates the weight of a given sample by considering the entire mixture as the global proposal [4]. This method attains a variance reduction at the expense of an increase in computational load [5], [6].

In this work, we propose a novel MIS method that provides an efficient tradeoff in terms of computational complexity and variance reduction of the associated IS estimators. The approach is based on creating a partition of the set of available proposals and considering that each partitioned set constitutes a mixture distribution. A sample drawn from a mixand in one of the partitions (mixtures) is then assigned a weight that only accounts for that particular mixture, instead of the entire composite mixture as in the full DM-MIS. A remarkable reduction in computational complexity is achieved by this approach, while the variance of the associated partial DM-MIS estimator remains comparable to that of the full DM-MIS estimator. The method can not only applied to IS methods with static distributions (i.e., characterized by fixed parameters), but also to IS methods that adapt the parameters of the proposal distribution in an iterative way (i.e., adaptive IS (AIS) methods [7]–[9]). Computer simulations show the excellent performance of the proposed scheme in terms of variance reduction for a given computational load.

II. PROBLEM STATEMENT AND BACKGROUND

In many applications, the interest lies in obtaining the posterior density function (pdf) of set of unknown parameters given the observed data. Mathematically, denoting the vector of unknowns as $\mathbf{x} \in \mathcal{D} \subseteq \mathbb{R}^n$ and the observed data as $\mathbf{y} \in \mathbb{R}^d$, the pdf is defined as

$$\widetilde{\pi}(\mathbf{x}|\mathbf{y}) = \frac{\ell(\mathbf{y}|\mathbf{x})g(\mathbf{x})}{Z(\mathbf{y})} \propto \pi(\mathbf{x}|\mathbf{y}) = \ell(\mathbf{y}|\mathbf{x})g(\mathbf{x}), \quad (1)$$

where $\ell(\mathbf{y}|\mathbf{x})$ is the likelihood function, $g(\mathbf{x})$ is the prior pdf, and $Z(\mathbf{y})$ is the normalization factor.[1] The computation of a particular moment of $\mathbf{x}$ is then given by

$$I = \frac{1}{Z} \int_{\mathcal{D}} f(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \quad (2)$$

where $f(\cdot)$ can be any integrable function of $\mathbf{x}$. In many practical scenarios, we cannot obtain an analytical solution of (2) and Monte Carlo methods are used to approximate it.

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

A. Importance Sampling (IS)

Let us consider N samples $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ drawn from a proposal pdf, $q(\mathbf{x})$, with heavier tails than the target, $\pi(\mathbf{x})$. The samples have associated importance weights given by

$$w_i = \frac{\pi(\mathbf{x}_i)}{q(\mathbf{x}_i)}, \quad i = 1, \ldots, N. \tag{3}$$

Using the samples and weights, the moment of interest can be approximated as

$$\hat{I}_{\mathrm{IS}} = \frac{1}{\frac{1}{N}\sum_{j=1}^{N} w_j} \sum_{i=1}^{N} w_i f(\mathbf{x}_i) = \frac{1}{N\hat{Z}} \sum_{i=1}^{N} w_i f(\mathbf{x}_i), \tag{4}$$

where $\hat{Z} = \frac{1}{N}\sum_{j=1}^{N} w_j$ is an unbiased estimator of $Z = \int_{\mathcal{D}} \pi(\mathbf{x}) d\mathbf{x}$ [1]. Note that Eq. (4) always provides a consistent estimator of I, but its variance is directly related to the discrepancy between $\pi(\mathbf{x})|f(\mathbf{x})|$ and $q(\mathbf{x})$ (for a specific choice of f) or to the mismatch between the target $\pi(\mathbf{x})$ and the proposal $q(\mathbf{x})$ (for a general and arbitrary f) [1], [10].

B. Multiple Importance Sampling (MIS)

Finding a good proposal pdf, $q(\mathbf{x})$, is critical and can also be very challenging [4]. An alternative and more robust strategy consists of using a population of different proposal pdfs. This scheme is often known in the literature as *multiple importance sampling* (MIS) [7]–[9]. Let us consider a set of N (normalized) proposal pdfs, $q_1(\mathbf{x}), \ldots, q_N(\mathbf{x})$, and let us assume that exactly one sample is drawn from each of them, i.e., $\mathbf{x}_i \sim q_i(\mathbf{x})$, $i = 1, \ldots, N$. The importance weights associated to these samples can then be obtained according to one of the following strategies:

- *Standard* MIS: $w_i = \frac{\pi(\mathbf{x}_i)}{q_i(\mathbf{x}_i)}, \quad i = 1, \ldots, N$.
- *Deterministic mixture MIS* (DM-MIS) [4]:

$$w_i = \frac{\pi(\mathbf{x}_i)}{\psi(\mathbf{x}_i)} = \frac{\pi(\mathbf{x}_i)}{\frac{1}{N}\sum_{j=1}^{N} q_j(\mathbf{x}_i)}, \quad i = 1, \ldots, N, \tag{5}$$

where $\psi(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N} q_j(\mathbf{x})$ is the mixture pdf, composed of all the proposal pdfs. This approach interprets the complete set of samples, $\{\mathbf{x}_i\}_{i=1}^{N}$, as being distributed according to the mixture $\psi(\mathbf{x})$, i.e., $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \sim \psi(\mathbf{x})$. See Appendix A for further details.

In both cases, the consistency of the estimators is ensured. The main advantage of the DM-MIS weights is that they yield more stable and efficient estimators [4], i.e., with less variance (as proved in Appendix B). However, the DM-MIS estimator is computationally more expensive, as it requires N evaluations of the proposal to obtain each weight instead of just one.[2] In some practical scenarios, this additional load may be excessive (especially for large values of N) and alternative efficient solutions must be developed.

C. Adaptive Importance Sampling (AIS)

In order to decrease the mismatch between the proposal and target pdfs, there are several Monte Carlo methods that itera-

tively adapt the parameters of the proposal pdf using the information of the past samples [7]–[9]. In this scenario, we have a set of proposal pdfs $\{q_j^{(t)}(\mathbf{x}), j = 1, \cdots, J\}_{t=1}^{T}$, where the superscript t indicates the iteration index and T is the total number of adaptation steps. Some of these well-known AIS methods, which are also based on MIS, are Population Monte Carlo (PMC) and its variants [7], [11]–[13], adaptive multiple importance sampling (AMIS) [8], and adaptive population importance sampling (APIS) [5], [9].

III. PARTIAL DETERMINISTIC MIXTURE APPROACH

In this section we develop a *partial* DM-MIS scheme, which groups the proposal pdfs, $\{q_i(\mathbf{x})\}_{i=1}^{N}$, into P mixtures composed of M mixands, with $PM = N$ (recall that $\mathbf{x}_i \sim q_i(\mathbf{x}), i = 1, \cdots, N$).[3] Namely, we define a partition of $\{1, \ldots, N\}$ into P disjoint subsets of M indices, $\mathcal{S}_p$ with $p = 1, \ldots, P$, s.t.

$$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \ldots \cup \mathcal{S}_P = \{1, \ldots, N\}, \tag{6}$$

where $\mathcal{S}_k \cap \mathcal{S}_\ell = \emptyset$ for all $k, \ell = 1, \ldots, P$ and $k \neq \ell$. Each subset, $\mathcal{S}_p = \{j_{p,1}, j_{p,2}, \ldots, j_{p,M}\}$, contains M indices, $j_{p,m} \in \{1, \ldots, N\}$ for $m = 1, \ldots, M$ and $p = 1, \ldots, P$. Following this strategy, the weights of the p-th mixture are computed as

$$w_i = \frac{\pi(\mathbf{x}_i)}{\psi_p(\mathbf{x}_i)} = \frac{\pi(\mathbf{x}_i)}{\frac{1}{M}\sum_{j \in \mathcal{S}_p} q_j(\mathbf{x}_i)}, \quad i \in \mathcal{S}_p. \tag{7}$$

The resulting *partial* DM-MIS (p-DM-MIS) estimator is then given by

$$\hat{I}_{\mathrm{p-DM-MIS}} = \frac{1}{\sum_{j=1}^{N} w_j} \sum_{i=1}^{N} w_i f(\mathbf{x}_i), \tag{8}$$

which coincides with the expression in (4), but using the weights given by Eq. (7). Note that the particular cases $P = 1$ and $P = N$ correspond to the *full* DM-MIS (f-DM-MIS) and the *standard* MIS (s-MIS) approaches, respectively. The number of evaluations of the proposal pdfs is then PM^2. Since $N \leq PM^2 = NM \leq N^2$, the computational cost is larger than that of s-MIS approach (M times larger), but lower than that of the f-DM-MIS approach (since $M \leq N$).

The good performance of the novel approach is ensured by Theorem 1 and Corollary 2 (see Appendix B) and can be summarized by the following expression:

$$\mathrm{Var}(\hat{I}_{\mathrm{f-DM-MIS}}) \leq \mathrm{Var}(\hat{I}_{\mathrm{p-DM-MIS}}) \leq \mathrm{Var}(\hat{I}_{\mathrm{s-MIS}}), \tag{9}$$

which holds regardless of the choice of P and the strategy followed to group the original proposals $\{q_i(\mathbf{x})\}_{i=1}^{N}$ into mixtures. Therefore, there is a tradeoff between performance and computational cost: using a smaller number of mixtures (P) leads to a reduced variance, but at the expense of an increase in the number of evaluations of the proposal densities.

A. Choice of the number of mixtures P

A simple strategy to choose the number of mixtures consists of starting with $P = N$ (which coincides with the standard MIS

---

[2]Note that the number of evaluations of the target $\pi(x)$ is the same regardless of whether the weights are calculated according to (3) or (5).

[3]For the sake of simplicity, we assume that all the mixtures contain the same number of proposal pdfs. However, any strategy that clusters the N proposals into P disjoint mixtures (regardless of their size) is valid.

scheme), computing the corresponding estimator in Eq. (8), and iteratively reducing the number of mixtures P (thus increasing $M = \frac{N}{P}$) while the estimation significantly changes w.r.t. the previous step. This iterative approach does not require a significant additional computational cost (since the proposal evaluations can be stored and re-used) and results in an efficient and automatic procedure to select P.

B. Design of the P mixtures

Developing an optimal strategy to cluster the proposals into P mixtures is a difficult task, since the number of possible configurations is extremely large. Indeed, unless this clustering strategy is computationally inexpensive, the additional computational effort might be better invested in decreasing P (thus increasing $M = \frac{N}{P}$ and reducing the variance of the estimators). Therefore, we propose applying a simple random clustering strategy, where $M = \frac{N}{P}$ different proposals are randomly assigned to each partition. This approach provides an excellent performance for large values of M (see Table II), so there seems to be little room for improvement (except maybe for small values of M). Note that this result is not surprising: randomness is the key element behind compressive sampling [14], and many randomized clustering algorithms have been developed for applications such as data mining [15], image processing [16] or blind channel identification [17].

C. Application to AIS schemes

For the sake of simplicity, we have focused on p-DM-MIS for a static framework, where the parameters of the proposals are fixed. However, all the previous considerations can be easily applied in AIS schemes, where the proposals are iteratively updated. In methods like PMC [7] or APIS [9], a population of J proposal densities is adapted during T iterations. At the t-th iteration, the j-th sample $\mathbf{x}_j^{(t)}$ is drawn from the j-th proposal, i.e., $\mathbf{x}_j^{(t)} \sim q_j^{(t)}(\mathbf{x})$ for $j = 1, \ldots, J$ and $t = 1, \ldots, T$. Thus, after T iterations we have $N = JT$ samples drawn from $N = JT$ different proposal pdfs.

Regardless of the adaptation procedure followed by each algorithm, different strategies can be used to design an efficient DM-MIS estimator when considering all the $N = JT$ proposal pdfs. Table I summarizes the weight calculation for three well-known AIS methods (PMC, AMIS and APIS), comparing them to the DM-MIS estimators. We also analyze the complexity in terms of the number of proposal evaluations and the performance in terms of the variance reduction. We can see that the novel p-DM-MIS approach provides a very good compromise between performance and computational cost. Finally, it is important to remark that Table I does not take into account the specific adaptive procedures of each algorithm, which can also have a large influence on the final performance.

IV. NUMERICAL EXAMPLE

We consider a bivariate multimodal target pdf, defined as a mixture of 5 Gaussians:

$$\pi(\mathbf{x}) = \frac{1}{5} \sum_{i=1}^{5} \mathcal{N}(\mathbf{x}; \nu_i, \Sigma_i), \quad \mathbf{x} \in \mathbb{R}^2, \qquad (10)$$

where $\nu_1 = [-10, -10]^\top$, $\nu_2 = [0, 16]^\top$, $\nu_3 = [13, 8]^\top$, $\nu_4 = [-9, 7]^\top$, $\nu_5 = [14, -14]^\top$, $\Sigma_1 = [2, 0.6; 0.6, 1]$, $\Sigma_2 = [2, -0.4; -0.4, 2]$, $\Sigma_3 = [2, 0.8; 0.8, 2]$, $\Sigma_4 = [3, 0; 0, 0.5]$ and

TABLE I
DIFFERENT STRATEGIES FOR WEIGHT CALCULATION IN AIS ALGORITHMS USING J PROPOSALS PER ITERATION (NOTE THAT $J = 1$ IN AMIS), AND T ITERATIONS

Method	Weight calculation	Complexity	Variance
PMC [7]	$\dfrac{\pi(\mathbf{x}_i^{(t)})}{q_i^{(t)}(\mathbf{x}_i^{(t)})}$	Lowest ($JT = N$)	Highest
APIS [9]	$\dfrac{\pi(\mathbf{x}_i^{(t)})}{\sum_{j=1}^{J} q_j^{(t)}(\mathbf{x}_i^{(t)})}$	Medium ($J^2T = JN$)	Medium
p-DM-MIS	$\dfrac{\pi(\mathbf{x}_i^{(t)})}{\sum_{j,\tau \in \mathcal{S}_P} q_j^{(\tau)}(\mathbf{x}_i^{(t)})}$	Suitable ($PM^2 = MN$)	Low
AMIS [8]	$\dfrac{\pi(\mathbf{x}_i^{(t)})}{\sum_{\tau=1}^{T} q^{(\tau)}(\mathbf{x}_i^{(t)})}$	Highest ($T^2 = N^2$)	Lowest
f-DM-MIS	$\dfrac{\pi(\mathbf{x}_i^{(t)})}{\sum_{\tau=1}^{T} \sum_{j=1}^{J} q_j^{(\tau)}(\mathbf{x}_i^{(t)})}$	Highest ($J^2T^2 = N^2$)	Lowest

TABLE II
MSE IN THE ESTIMATION OF THE MEAN AND NORMALIZING CONSTANT OF THE TARGET FOR DIFFERENT VALUES OF P AND M.

Parameters	MSE($E[\mathbf{X}]$)	MSE(Z)	Evaluations
$P = 4096$ ($M = 1$): s-MIS	6.8129	0.0743	4096
$P = 2048$ ($M = 2$)	3.3832	0.0273	8192
$P = 1024$ ($M = 4$)	1.4723	0.0120	16384
$P = 512$ ($M = 8$)	0.9678	0.0076	32768
$P = 256$ ($M = 16$)	0.8078	0.0064	65536
$P = 128$ ($M = 32$)	0.7730	0.0060	131072
$P = 64$ ($M = 64$)	0.7648	0.0058	262144
$P = 32$ ($M = 128$)	0.7506	0.0058	524288
$P = 16$ ($M = 256$)	0.7486	0.0058	1048576
$P = 8$ ($M = 512$)	0.7448	0.0058	2097152
$P = 4$ ($M = 1024$)	0.7416	0.0058	4194304
$P = 2$ ($M = 2048$)	0.7414	0.0058	8388608
$P = 1$ ($M = 4096$): f-DM-MIS	0.7406	0.0058	16777216

$\Sigma_5 = [2, -0.1; -0.1, 2]$. The goal is to approximate, using some Monte Carlo method, the expected value of $\mathbf{X} \sim \pi(\mathbf{x})$, i.e., $E[\mathbf{X}] = \int_{\mathbb{R}^2} \mathbf{x} \pi(\mathbf{x}) d\mathbf{x}$ and $Z = \int_{\mathbb{R}^2} \pi(\mathbf{x}) d\mathbf{x}$.

We apply the MIS algorithm in a setup with $N = 4096$ Gaussian proposal pdfs, $\{q_i(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_i, C_i)\}_{i=1}^{N}$, where $\boldsymbol{\mu}_i \sim \mathcal{U}([-20, 20] \times [-20, 20])$ is the randomly chosen location parameter and $\mathbf{C}_i = \sigma^2 \mathbf{I}_2$, with $\sigma = 5$, is the scale matrix. We proceed as follows. First, we draw a sample from each proposal. Then, we compute the corresponding weight according to (7). Finally, we build the estimator $\hat{I}_{\text{p-DM-MIS}}$ using (8) for different number of mixtures $P = 2^k$ for $k = 0, 1, \ldots, 12$ (i.e., $P = 1, 2, \ldots, 2048, 4096$). Since $PM = N$, the number of proposals per mixture is $M = 2^{12-k} = 4096, 2048, \ldots, 2, 1$. Note that the case $P = 1$ (i.e., $M = 4096$) corresponds to the f-DM-MIS approach, while $P = 4096$ (i.e., $M = 1$) corresponds to the s-MIS. A random assignment of the proposals to the mixtures is performed in all cases.

Table II shows the mean square error (MSE) in the estimation of the mean of the target $E[X]$ (averaged over both dimensions) and the normalizing constant Z. All results are averaged over 500 independent runs. The last column of Table II shows the total number of proposal evaluations for each value of M. Fig. 1 shows the MSE in the estimation of the mean of the target w.r.t. the total number of proposal evaluations. The results show that decreasing P reduces the MSE (as expected) at the expense of an increase in the computational cost (measured by the number of proposal evaluations). However, note that decreasing the number of mixtures below $P = 64$ does not improve the performance significantly. Indeed, the p-DM-MIS with $P = 64$
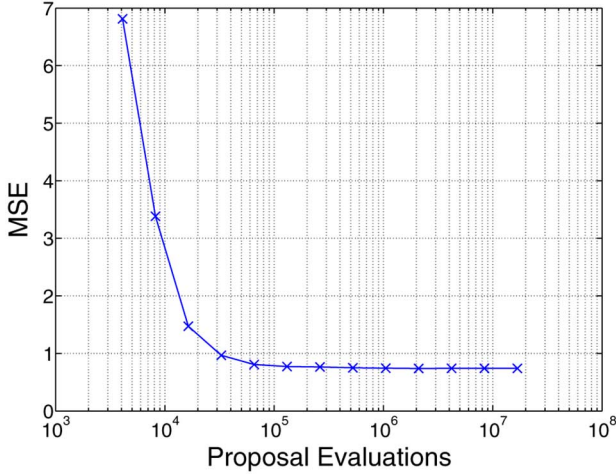
Fig. 1. MSE in the estimation of the mean of the target w.r.t. the total number of proposal evaluations.

obtains an MSE close to that of the f-DM-MIS estimator while performing 98.4% less proposal evaluations, thus attaining an excellent performance-cost tradeoff.

V. CONCLUSION

In this letter we propose a novel approach for the calculation of the weights in multiple importance sampling schemes that provides an efficient tradeoff between computational complexity and variance reduction of the associated estimator. The proposed scheme is based on constructing a random partition of the set of available proposals and then calculating the weight of each sample locally using only the corresponding subset of the partition. Computer simulations reveal a very good performance of the method, which is able to attain an excellent performance-computational cost tradeoff.

APPENDIX A
DRAWING SAMPLES FROM A MIXTURE OF PDFS

Let us consider a mixture of N normalized pdfs with equal weights, i.e.,

$$\psi(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} q_i(\mathbf{x}). \qquad (11)$$

The classical procedure for drawing N samples from $\psi(\mathbf{x})$ is (starting with $k = 1$):
1) Draw $j^* \in \{1, \dots, N\}$ with equal weights $\frac{1}{N}$.
2) Draw $\mathbf{x}_k \sim q_{j^*}(\mathbf{x})$.
3) Set $k = k + 1$ and repeat until $k = N$.

In this way, each sample $\mathbf{x}_k$ is distributed according to $\psi(\mathbf{x})$ and, as a consequence, $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \psi(\mathbf{x})$. An alternative procedure, more *deterministic* than the previous one, consists of the following steps (starting with $i = 1$):
1) Draw one sample from each $q_i(\mathbf{x})$, i.e., $\mathbf{x}_i \sim q_i(\mathbf{x})$.
2) Set $i = i + 1$ and repeat until $i = N$.

In this case, we have $\mathbf{x}_i \sim q_i(\mathbf{x})$ for $i = 1, \dots, N$, *but* the joint set is again distributed as $\psi(\mathbf{x})$, i.e., $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \sim \psi(\mathbf{x})$. This is the underlying theoretical argument of the deterministic mixture (DM) weights approach used throughout this work. Furthermore, given M indices $j_r \in \{1, \dots, N\}$ with $r = 1, \dots, M$:

$$\{\mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_M}\} \sim \frac{1}{M} q_{j_1}(\mathbf{x}) + \dots + \frac{1}{M} q_{j_M}(\mathbf{x}).$$

APPENDIX B
VARIANCE REDUCTION OF DETERMINISTIC MIXTURE WEIGHTS

A. Theorem 1

Consider a normalized target pdf, $\tilde{\pi}(\mathbf{x}) = \frac{1}{Z}\pi(\mathbf{x})$, and N samples drawn from a set of N normalized proposal pdfs (one from each pdf), $\mathbf{x}_i \sim q_i(\mathbf{x})$ for $i = 1, 2, \dots, N$. The standard multiple importance sampling (s-MIS) and the full deterministic mixture IS (f-DM-IS) estimators can be expressed as

$$\hat{I}_{s-MIS} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(\mathbf{x}_i)\pi(\mathbf{x}_i)}{q_i(\mathbf{x}_i)}, \qquad (12a)$$

$$\hat{I}_{f-DM-MIS} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(\mathbf{x}_i)\pi(\mathbf{x}_i)}{\frac{1}{N}\sum_{j=1}^{N} q_j(\mathbf{x}_i)}. \qquad (12b)$$

The variance of the f-DM-IS estimator is always lower or equal than the variance of the s-MIS estimator,

$$Var(\hat{I}_{f-DM-IS}) \leq Var(\hat{I}_{s-MIS}). \qquad (13)$$

1) Proof: See Appendix A of [5], [6]. □

B. Corollary 2

For the standard MIS, full DM-MIS and partial DM-MIS estimators the following inequality holds:

$$Var(\hat{I}_{f-DM-MIS}) \leq Var(\hat{I}_{p-DM-MIS}) \leq Var(\hat{I}_{s-MIS}).$$

1) Proof: For the first inequality, note that the full DM-MIS estimator of (12b) can also be expressed as

$$\hat{I}_{f-DM-MIS} = \frac{1}{N} \sum_{p=1}^{P} \sum_{i \in \mathcal{S}_p} \frac{f(\mathbf{x}_i)\pi(\mathbf{x}_i)}{\frac{1}{P}\sum_{p=1}^{P} \psi_p(\mathbf{x}_i)},$$

where

$$\psi_p(\mathbf{x}_i) = \frac{1}{M} \sum_{j \in \mathcal{S}_p} q_j(\mathbf{x}_i)$$

is the proposal associated to the p-th mixture. Similarly, the partial DM-MIS estimator is given by

$$\hat{I}_{p-DM-MIS} = \frac{1}{N} \sum_{p=1}^{P} \sum_{i \in \mathcal{S}_p} \frac{f(\mathbf{x}_i)\pi(\mathbf{x}_i)}{\psi_p(\mathbf{x}_i)}.$$

Hence, applying Theorem 1 with $\psi_p(\mathbf{x}_i)$ instead of $q_i(\mathbf{x}_i)$ and $\frac{1}{P}\sum_{p=1}^{P}\psi_p(\mathbf{x}_i)$ instead of $\frac{1}{N}\sum_{j=1}^{N}q_j(\mathbf{x}_i)$, we have $Var(\hat{I}_{f-DM-MIS}) \leq Var(\hat{I}_{p-DM-MIS})$.

For the second inequality, following the same approach, the standard MIS estimator of (12a) can be rewritten as

$$\hat{I}_{s-MIS} = \frac{1}{N} \sum_{p=1}^{P} \sum_{i \in \mathcal{S}_p} \frac{f(\mathbf{x}_i)\pi(\mathbf{x}_i)}{q_i(\mathbf{x}_i)}.$$

Applying Theorem 1 with $\psi_p(\mathbf{x}_i) = \frac{1}{P}\sum_{p=1}^{P}\psi_p(\mathbf{x}_i)$ instead of $\frac{1}{N}\sum_{j=1}^{N}q_j(\mathbf{x}_i)$, we have $Var(\hat{I}_{p-DM-MIS}) \leq Var(\hat{I}_{s-MIS})$. □

REFERENCES

[1] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, NY, USA: Springer, 2004.

[2] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. New York, NY, USA: Springer, 2004.

[3] E. Veach and L. Guibas, "Optimally combining sampling techniques for Monte Carlo rendering," in *Proc. SIGGRAPH 1995*, 1995, pp. 419–428.

[4] A. Owen and Y. Zhou, "Safe and effective importance sampling," *J. the Amer. Statist. Assoc.*, vol. 95, no. 449, pp. 135–143, 2000.

[5] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler: Learning from the uncertainty," *viXra:1405.0280*, 2014.

[6] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler: Learning from uncertainty," *IEEE Trans. Signal Process.*, 2015, In Press.

[7] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert, "Population Monte Carlo," *J. Comput. Graph. Statist.*, vol. 13, no. 4, pp. 907–929, 2004.

[8] J. M. Cornuet, J. M. Marin, A. Mira, and C. P. Robert, "Adaptive multiple importance sampling," *Scand. J. Statist.*, vol. 39, no. 4, pp. 798–812, Dec. 2012.

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

[10] H. Kahn and A. W. Marshall, "Methods of reducing sample size in Monte Carlo computations," *J. Oper. Res. Soc. Amer.*, vol. 1, no. 5, pp. 263–278, 1953.

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

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

[13] M. Bugallo, M. Hong, and P. M. Djuric, "Marginalized population Monte Carlo," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2009, pp. 2925–2928, IEEE.

[14] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process, Mag,*, vol. 25, no. 2, pp. 21–30, 2008.

[15] R. T. Ng and J. Han, "CLARANS: A method for clustering objects for spatial data mining," *IEEE Trans. Knowl, Data Eng.*, vol. 14, no. 5, pp. 1003–1016, 2002.

[16] F. Moosmann, E. Nowak, and F. Jurie, "Randomized clustering forests for image classification," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1632–1646, 2008.

[17] D. Luengo, S. Monzón, and A. Artés-Rodríguez, "Novel fast random search clustering algorithm for mixing matrix identification in MIMO linear blind inverse problems with sparse inputs," *Neurocomputing*, vol. 87, pp. 62–78, 2012.