Heretical Multiple Importance Sampling

Víctor Elvira, *Member, IEEE,* Luca Martino, David Luengo, *Member, IEEE,*

and Mónica F. Bugallo, *Senior Member, IEEE,*

Abstract

Multiple Importance Sampling (MIS) methods approximate moments of complicated distributions by drawing samples from a set of proposal distributions. Several ways to compute the importance weights assigned to each sample have been recently proposed, with the so-called deterministic mixture (DM) weights providing the best performance in terms of variance, at the expense of an increase in the computational cost. A recent work has shown that it is possible to achieve a trade-off between variance reduction and computational effort by performing an *a priori* random clustering of the proposals (partial DM algorithm). In this paper, we propose a novel "heretical" MIS framework, where the clustering is performed *a posteriori* with the goal of reducing the variance of the importance sampling weights. This approach yields biased estimators with a potentially large reduction in variance. Numerical examples show that heretical MIS estimators can outperform, in terms of mean squared error (MSE), both the standard and the partial MIS estimators, achieving a performance close to that of DM with less computational cost.

Index Terms

Monte Carlo methods, multiple importance sampling, deterministic mixture, biased estimation.

I. INTRODUCTION

Importance sampling (IS) methods are widely used in signal processing to approximate statistical moments of random variables (r.v.) by a set of weighted samples. In particular, they are employed when the targeted probability density function (pdf) of the r.v. is complicated and its moments cannot be analytically computed, and/or drawing samples from the targeted pdf is impossible. In those cases, samples are drawn from a simpler proposal pdf, and a set of weights associated to them are used to measure the discrepancy between the proposal and target densities [1], [2]. While the standard IS approach uses a single proposal density to draw all the samples, multiple importance sampling (MIS) schemes use a set of different proposal distributions to reduce the variance of the estimators [3], [4], [5]. Due to the existence of multiple proposals, there are many different alternatives for the sampling and weighting procedures (see [5] for a thorough description of several valid schemes). In the standard MIS scheme, the weight of a particular sample is proportional to the ratio between the target pdf and the proposal that generated that sample, both evaluated at the sample value (see for instance [1, Section 3.3.1.]). The so-called deterministic mixture (DM) MIS scheme provides the best performance (in terms of variance of the associated estimators) of all the alternatives proposed so far (see [5, Theorems 1 and 2]). Nevertheless, this variance reduction is achieved at the expense of increasing the computational load w.r.t. the standard MIS scheme.

A partial DM MIS scheme, where proposals are randomly clustered *a priori* into disjoint sets and the partial DM weight of each sample is computed by taking into account only the proposals of the specific subset, has recently been proposed [6]. The variance and the computational complexity of this partial DM scheme falls in-between the standard and DM approaches, allowing the practitioner to select the desired computational budget in order to reduce the variance. Note that the effect of using the DM or partial DM weights instead of the standard weights is linked to other approaches that attempt to reduce the mean squared error (MSE) of the IS estimators. It is well known that the IS weights are usually (right) heavy-tailed distributed when the proposal and the target distribution have a mismatch [7], [8], [9], thus resulting in unstable and high variance estimators. Several techniques address this issue by nonlinearly transforming the IS weights. Illustrative examples are the truncated IS algorithm [7], clipping NPMC [8], or Pareto-smoothed IS [9]. All of these techniques attain a large reduction in the variance of the estimators in many useful scenarios at the expense of introducing bias.

In this paper, we present a novel "heretical" MIS methodology that extends the partial DM framework of [6] with the goal of reducing the heavy tails of the distribution of weights. The proposal densities are partitioned into disjoint subsets, as in partial DM. However, unlike [6] (where the partition is fixed *a priori*), here the partitioning is performed *a posteriori* (i.e., once the samples have been drawn), focusing on finding configurations that reduce the largest weights. This approach yields biased estimators with a potentially large reduction in variance, especially for small sample sizes [10], [11]. Numerical examples show that heretical MIS estimators can outperform, in terms of MSE, both the standard and the partial DM MIS estimators.

II. IMPORTANCE SAMPLING

Let us denote the vector of unknown parameters to be inferred as $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$ with pdf

$$\widetilde{\pi}(\mathbf{x}) = \frac{\pi(\mathbf{x})}{Z}, \tag{1}$$

where $\pi(\mathbf{x})$ is the unnormalized target function, and Z is the partition function. The goal is computing some particular moment of $\mathbf{x}$, which can be defined as

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

where $f(\cdot)$ can be any square-integrable function of $\mathbf{x}$ w.r.t. $\pi(\mathbf{x})$. Unfortunately, in many cases an analytical solution of Eq. (2) cannot be computed, and Monte Carlo methods are used to approximate I.

A. *Importance sampling with a single proposal*

Let us consider N samples, $\{\mathbf{x}_n\}_{n=1}^N$, drawn from a single proposal pdf, $q(\mathbf{x})$, with heavier tails than the target, $\pi(\mathbf{x})$. Each sample has an associated importance weight given by

$$w_n = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}, \quad n = 1, \dots, N. \tag{3}$$

Using the samples and weights, the moment of interest can be approximated with a self-normalized estimator as

$$\tilde{I}_{\text{IS}} = \frac{1}{N\hat{Z}} \sum_{n=1}^N w_n f(\mathbf{x}_n), \tag{4}$$

where $\hat{Z} = \frac{1}{N} \sum_{n=1}^N w_n$ is an unbiased estimator of $Z = \int_{\mathcal{D}} \pi(\mathbf{x})d\mathbf{x}$ [1]. If the normalizing constant is known, then it is possible to use the unnormalized estimator, given by

$$\hat{I}_{\text{IS}} = \frac{1}{NZ} \sum_{n=1}^N w_n f(\mathbf{x}_n). \tag{5}$$

Note that $\tilde{I}_{\text{IS}}$ is asymptotically unbiased, whereas $\hat{I}_{\text{IS}}$ is unbiased. Both $\tilde{I}_{\text{IS}}$ and $\hat{I}_{\text{IS}}$ are consistent estimators of I and their variance is directly related to the discrepancy between $\tilde{\pi}(\mathbf{x})|f(\mathbf{x})|$ and $q(\mathbf{x})$ [1], [12]. However, when several different moments f must be estimated or the function f is unknown *a priori*, a common strategy is decreasing the mismatch between the proposal $q(\mathbf{x})$ and the target $\tilde{\pi}(\mathbf{x})$ [13, Section 3.2], which is equivalent to minimizing the variance of the weights (and also the variance of the estimator $\hat{Z}$).

B. Multiple importance sampling (MIS)

Since the target density can only be evaluated point-wise, but cannot be easily characterized in many cases, finding a single good proposal pdf, $q(\mathbf{x})$, is not always possible. A robust alternative is then using a set of proposal pdfs, $\{q_n(\mathbf{x})\}_{n=1}^N$ [3], [4]. This scheme is usually known in the literature as multiple importance sampling (MIS), and it is an important feature of many state-of-the-art adaptive IS algorithms (e.g., [14], [15], [16]).

A general MIS framework has recently been proposed in [5], where several sampling and weighting schemes can be used. Here, we briefly review the most common sampling and two common weighting schemes. Let us draw one sample from each proposal pdf, i.e.,

$$\mathbf{x}_n \sim q_n(\mathbf{x}), \qquad n = 1, ..., N. \tag{6}$$

The most common weighting strategies in the literature are:

1) *Standard MIS* (s-MIS) [14]:

$$w_n = \frac{\pi(\mathbf{x}_n)}{q_n(\mathbf{x}_n)}, \quad n = 1, \ldots, N. \tag{7}$$

2) *Deterministic mixture MIS* (DM) [4]:

$$w_n = \frac{\pi(\mathbf{x}_n)}{\psi(\mathbf{x}_n)} = \frac{\pi(\mathbf{x}_n)}{\frac{1}{N}\sum_{j=1}^N q_j(\mathbf{x}_n)}, \quad n = 1, \ldots, N, \tag{8}$$

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, evaluated at $\mathbf{x}$.

Once more, the estimator of Eq. (4) using any of both sets of weights (i.e., (7) or (8)) is consistent and asymptotically unbiased, whereas the estimator of Eq. (5) is both consistent and unbiased. The superiority of the DM approach w.r.t. the s-MIS, in terms of variance of the estimator $\hat{I}$, has been proved in [5]. However, although both alternatives perform the same number of target evaluations, the DM estimator is computationally more expensive w.r.t. the number of proposal evaluations. In particular, s-MIS and DM require N and N^2 evaluations, respectively. In some scenarios (e.g., with a large number of proposals N) the extra number of proposal evaluations can be a major issue, and alternative efficient solutions have to be devised.

C. Partial Multiple Importance Sampling (p-DM)

A new general framework for building estimators, called partial DM and characterized by a performance and computational complexity in between s-MIS and DM, was introduced in [6]. Let us assume that we draw N samples using Eq. (6), and a partition of all the N proposals into P disjoints subsets of M proposals ($N = PM$) is defined *a priori*. More precisely, the set of indices $\{1, \ldots, N\}$ is partitioned into P disjoint subsets of M indices, $\mathcal{S}_p$ with $p = 1, \ldots, P$, s.t.

$$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \ldots \cup \mathcal{S}_P = \{1, \ldots, N\}, \tag{9}$$

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 $p = 1, \ldots, P$,[1] and then each sample $\mathbf{x}_n$ receives a "partial" DM weight considering at its denominator the mixture of all the proposals of the subset. Following this strategy, the weights of the samples of the p-th mixture are computed as

$$w_n = \frac{\pi(\mathbf{x}_n)}{\psi_p(\mathbf{x}_n)} = \frac{\pi(\mathbf{x}_n)}{\frac{1}{M} \sum_{j \in \mathcal{S}_p} q_j(\mathbf{x}_n)}, \quad n \in \mathcal{S}_p. \tag{10}$$

It can be proved that the p-DM approach yields consistent estimators, with a performance and computational complexity in between s-MIS and DM (see [6, Appendix B]), i.e.,

$$\mathrm{Var}(\hat{I}_{\mathrm{DM}}) \leq \mathrm{Var}(\hat{I}_{\mathrm{p\text{-}DM}}) \leq \mathrm{Var}(\hat{I}_{\mathrm{s\text{-}MIS}}). \tag{11}$$

Moreover, in p-DM the number of evaluations of the proposal pdfs is PM^2. Since $N \leq PM^2 = NM \leq N^2$, the computational cost is larger than that of the s-MIS approach (M times larger), but lower than that of the f-DM approach (since $M \leq N$). The particular cases $P = 1$ and $P = N$ correspond to the DM and the s-MIS approaches, respectively.

III. HERETICAL MIS

In adaptive IS algorithms, the proposals are iteratively adapted in order to reduce the mismatch w.r.t. the target (see for instance PMC [14], AMIS [15], or APIS [16]). However, this does not always occur, and a substantial mismatch can still remain in the first iterations even when the adaptation is successful (see for instance the discussion in [8, Section 4.2.]). This mismatch yields right heavy tailed distributed weights; extremely large weights can be associated to samples drawn in the tails of the proposals for which the target evaluation is high. In this situation, the estimator can be dominated in practice by a single sample. This is a well known fact that can result in a large increase in the variance of the estimators, even yielding estimators with infinite variance (see for instance [4]). This example intuitively explains how DM reduces the variance of the weights: in the aforementioned situation, if there is another proposal in the neighborhood of $\mathbf{x}_n$, the evaluation of the mixture proposal is no longer too small, and the weight does not take an extremely large value. This will be exploited by the proposed algorithm in order to reduce the variance of the MIS weights.

A. Proposed algorithm

In the novel heretical DM (h-DM) MIS algorithm, the proposals are clustered once the samples $\{\mathbf{x}_n\}_{n=1}^N$ have been drawn with the goal of reducing the largest weights, unlike in the p-DM algorithm where the clustering is performed *a priori*. The proposed approach is summarized in Algorithm 1. The sampling (line 1) and the initial weighting (line 2) are performed exactly as in s-MIS. Afterwards, the algorithm selects the largest weight, w_{n^*} (line 5), and the proposal j^* that maximizes $q_{j^*}(\mathbf{x}_{n^*})$. Then, q_{n^*} and q_{j^*} are clustered into the same subset (lines 7–13). This process is repeated until all the proposals have been allocated to one of all the P subsets (or until some other rule is fulfilled, as discussed in Section III-C). Finally, the weights are recomputed according to the p-DM weights of Eq. (10) (lines 17–18).

B. Computational complexity and performance

The h-DM algorithm performs $N^2/P \leq N^2$ proposal evaluations, thus reducing the computational complexity of the full DM approach, which is now comparable to that of the p-DM algorithm. The most costly step (w.r.t. the p-DM) is the search for the proposal j^* that maximizes the evaluation $q_{j^*}(\mathbf{x}_{n^*})$ (line 6). However, let us remark that the computational cost of

---

[1]For the sake of simplicity, and without loss of generality, it is assumed that all the subsets have the same number of proposals.

this stage can be substantially alleviated. Firstly, a smart search can be performed in order to avoid checking all the available proposals. Secondly, suboptimal search strategies are likely to work similarly well; the variance of the weights can still be notably reduced if close enough proposals (not necessarily the closest ones) to the samples with large weights are found. Thirdly, in order to find the closest proposal, q_{j^*}, it is not always necessary to compute the whole evaluation. Finally, the (partial) evaluation of some of the proposals can be reused at the reweighting stage (line 17). Similarly to the p-DM, the performance of h-DM improves when increasing P. In addition, for a specific choice of P, h-DM reduces the variance of the weights at the expense of introducing bias. While a theoretical analysis is still required, the variance reduction can be large, especially for low values of P. This explains why h-DM clearly outperforms p-DM in the experiments of Section IV.

Algorithm 1 Heretical MIS

1: Draw N samples, one from each proposal pdf,

$$\mathbf{x}_n \sim q_n(\mathbf{x}), \qquad n = 1, ..., N.$$

2: Compute the standard IS weights

$$w_n = \frac{\pi(\mathbf{x}_n)}{q_n(\mathbf{x}_n)}, \qquad n = 1, ..., N.$$

3: Initialize the subsets $\mathcal{R} = \{1, ..., N\}$, $\mathcal{A} = \{1, ..., N\}$, and $\mathcal{S}_p = \{\emptyset\}$ with $p = 1, ..., P$.

4: **while** $\mathcal{R} \neq \emptyset$ **do**

5:    Find the index of the maximum weight:

$$n^* = \arg\max_{n \in \mathcal{R}} w_n$$

6:    Find the index of the closest proposal to $\mathbf{x}_{n^*}$:

$$j^* = \arg\max_{j \in \mathcal{A}} q_j(\mathbf{x}_{n^*})$$

7:    **if** j^* belongs already to a subset $\mathcal{S}_{m_{j^*}}$ **then**

8:       cluster n^* into that subset: $\mathcal{S}_{m_{j^*}} = \mathcal{S}_{m_{j^*}} \cup n^*$

9:       save the subset where n^* is stored: $m_{n^*} = m_{j^*}$

10:   **else**

11:      cluster n^* and j^* together:[2] $\mathcal{S}_m = \mathcal{S}_m \cup n^* \cup j^*$

12:      save the subset where n^* is stored: $m_{n^*} = m$

13:      save the subset where j^* is stored: $m_{j^*} = m$

14:   $\mathcal{R} = \mathcal{R} \backslash n^*$

15:   **if** $|\mathcal{S}_{n^*}| = M$ **then**

16:      $\mathcal{A} = \mathcal{A} \backslash \mathcal{S}_{n^*}$.

17: Recompute the weights w_n using Eq. (10)

18: Compute the normalized weights as $\bar{w}_n = \frac{w_n}{\sum_{k=1}^N w_n}$

---

[2] In this case, n^* and j^* are stored in any $\mathcal{S}_m$, with the condition that this subset has two available slots, i.e., $|\mathcal{S}_m| \leq M - 2$. If none of the subsets has two available slots, n^* and j^* are randomly allocated. Note that more advanced strategies can be developed in this situation, but we keep the algorithm simple for the sake of clarity in the explanation.

C. Extensions

The proposed algorithm opens the door to a novel class of MIS algorithms with one feature in common: constructing the weighting functions after drawing the samples (i.e., the weighting function itself is a r.v.). Several aspects of the core algorithm proposed here can be improved: the computational complexity (related to the number of subsets P) might be automatically determined by the algorithm, the subsets might have different number of elements, or the partitioning algorithm could be stopped once the largest weights have been reduced (performing the rest of the allocation randomly). For instance, we propose a simple extension of the clustering method by introducing a parameter ($0 \leq \alpha \leq 1$) that represents the fraction of proposals that are clustered following Algorithm 1. Whenever more than αN proposals have been allocated, the remaining ones are randomly clustered as in p-DM. Hence, $\alpha = 0$ corresponds to the p-DM in [6], while $\alpha = 1$ corresponds to Algorithm 1. The extension of the algorithm proposed in this section largely reduces the complexity of the clustering algorithm (roughly by a fraction of α).

D. Connection with other techniques

Several algorithms in the literature propose alternative ways to reduce the variance of the IS weights. The common approach consists of clipping the largest unnormalized weights to a certain value (typically, the minimum of the set of transformed/clipped weights), and then normalizing the weights. In [7] and [8], a fraction of the largest unnormalized weights are clipped to the minimum of those weights. The method proposed in [9] is an advanced implementation of [7], where the distribution of the weights is fitted by a Pareto distribution.

In h-DM, the largest unnormalized weights are also attenuated, but the approach followed is completely different, since the procedure is based on the construction of partial MIS estimators. The statistical meaning of the h-DM weights is then supported by the DM interpretation, present in DM or p-DM. While [7], [8] and [9] introduce bias due to the nonlinear transformation of the weights, in h-DM the bias appears because the values of the samples are used to perform the clustering. A detailed theoretical study of the bias and consistency of h-DM is thus required. In future works we plan to address this issue by adapting and extending the analysis techniques already developed in [7], [8], [9].

IV. NUMERICAL RESULTS

Example 1: Let us consider a unidimensional target which is itself a mixture of two Gaussian pdfs:

$$\pi(x) = \frac{1}{2}\mathcal{N}(x; a_1, c^2) + \frac{1}{2}\mathcal{N}(x; a_2, c^2), \tag{12}$$

with $a_1 = -3$, $a_2 = 5$, and $c^2 = 1$.

The goal is to approximate, via Monte Carlo, the expected value of $x \sim \pi(x)$, i.e., $E[x] = \int_{\mathbb{R}} x\pi(x)dx$ and $Z = \int_{\mathbb{R}} \pi(x)dx$. We apply the MIS algorithms in a setup with $N = 32$ Gaussian proposal pdfs, $\{q_n(x) = \mathcal{N}(x; \mu_n, \sigma^2)\}_{n=1}^N$, where the means are equidistant in the interval $[-8, 8]$, while their variances are all equal to $\sigma_n^2 = 3$ for $n = 1, ..., N$, and $L = 32 \cdot k$ samples with $k \in \{1, .., 5\}$, i.e., exactly k samples per proposal are drawn. We compare the performance of s-MIS, DM, p-DM, and the proposed h-DM. In the last two algorithms (p-DM and h-DM) we set $P = 16$, i.e., in p-DM the proposals are randomly clustered into 16 subsets, whereas the allocation in h-DM is performed using Algorithm 1. Note that, for one-dimensional Gaussian proposals with equal variance, the maximization in line 6 of Algorithm 1 is equivalent to

$$j^* = \arg\max_{j \in \mathcal{A}} (x_{n^*} - \mu_j)^2 = \arg\max_{j \in \mathcal{A}} |x_{n^*} - \mu_j|,$$

which requires only a partial evaluation of the proposals.

Figure 1 shows the performance of the different algorithms in terms of the MSE of the self-normalized estimator $\tilde{I}$ of the mean of the target distribution. It can be seen that h-DM outperforms p-DM in this scenario, obtaining a performance close to that of DM with a reduced computational cost.
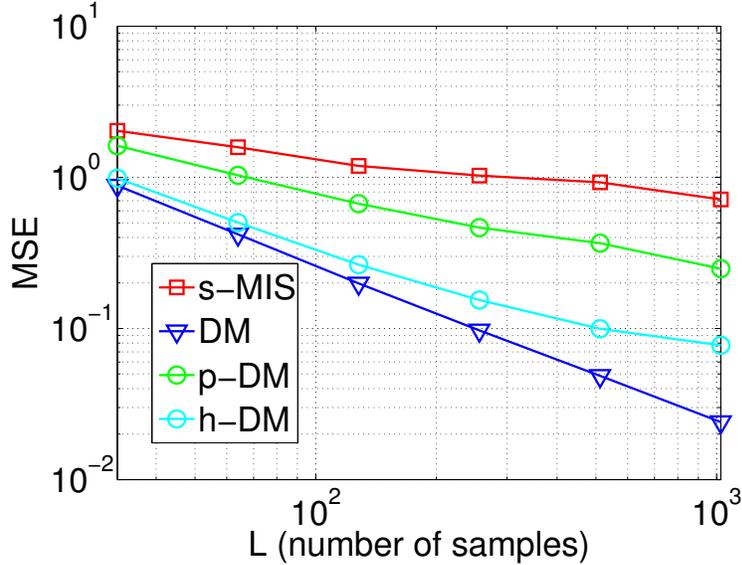

Fig. 1: [Ex. 1] MSE of the self-normalized estimator $\tilde{I}$ of the mean of the target using $L = 32 \cdot k$ samples for $k \in \{1, .., 5\}$.

Example 2: Let us consider a unidimensional target which is a (normalized) mixture of five non-standardized Student's t pdfs with location parameters $a_1 = -3$, $a_2 = -1$, $a_3 = 0$, $a_4 = 3$ and $a_5 = 4$, scale parameters $c_i^2 = 1$ and degrees of freedom $\xi_i = 5$, both for $i = 1, ..., 5$. The goal is to approximate the mean of the target. We apply the MIS algorithm in a setup with $N = 32$ non-standardized Student's t pdfs, where the location parameters μ_n are equidistant in the interval $[-8, 8]$, while their scale parameters and degrees of freedom are equal to $\sigma_n^2 = 3$ and $\nu_n = 4$ for $n = 1, ..., N$, respectively. We compare the p-DM and the proposed h-DM, testing $P \in \{1, 2, 4, 8, 16, 32\}$, and thus $M \in \{32, 16, 8, 4, 2, 1\}$. Note that s-MIS corresponds to $P = 32$ and $M = 1$, whereas DM is obtained with $P = 1$ and $M = 32$. We use the modified algorithm suggested in Section III-C with $\alpha = 0.1$. Figure 2 shows the MSE of the unnormalized estimator $\tilde{I}$ of the mean of the target distribution. Once more, h-DM outperforms p-DM substantially (especially for $M < 8$).

V. CONCLUSIONS

In this paper, we have proposed a heretical MIS framework that allows us to achieve a tradeoff between performance and computational cost by clustering the proposals *a posteriori*. The resulting estimators are biased, but their reduced variance can lead to a substantial decrease in MSE, especially for small sample sizes. Future lines include developing more efficient clustering methods and performing a rigorous theoretical study of the bias and consistency of the estimators.

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

[4] A. Owen and Y. Zhou, "Safe and effective importance sampling," *Journal of the American Statistical Association*, vol. 95, no. 449, pp. 135–143, 2000.
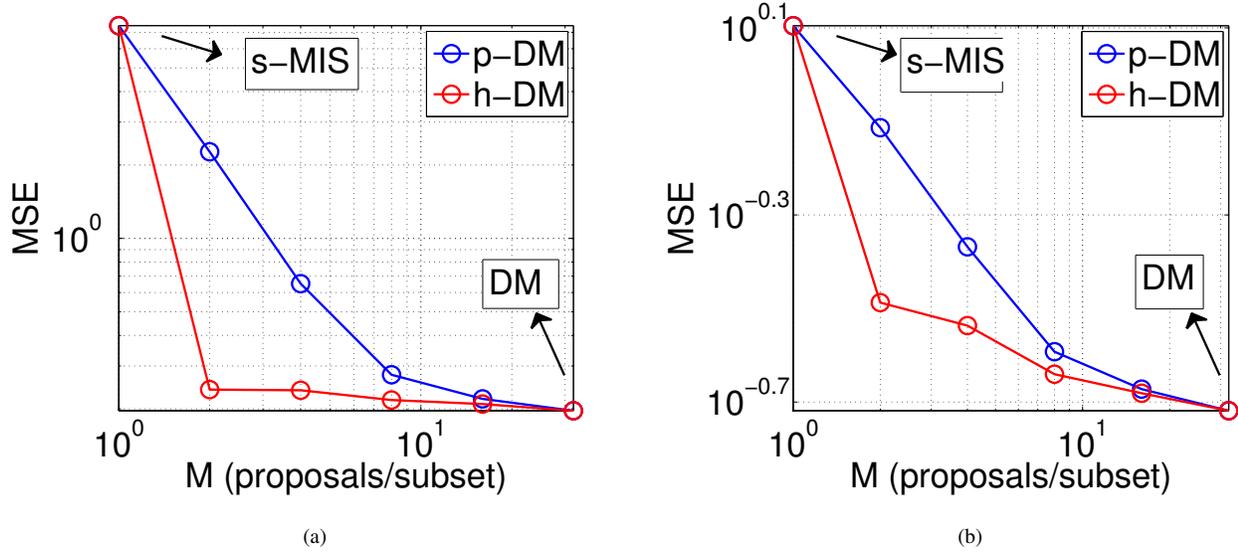
Fig. 2: [Ex. 2] MSE of (a) $\hat{I}$ and (b) $\tilde{I}$.

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

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

[7] E. L. Ionides, "Truncated importance sampling," *Journal of Computational and Graphical Statistics*, vol. 17, no. 2, pp. 295–311, 2008.

[8] E. Koblents and J. Míguez, "A population Monte Carlo scheme with transformed weights and its application to stochastic kinetic models," *Statistics and Computing*, vol. 25, no. 2, pp. 407–425, 2015.

[9] A. Vehtari and A. Gelman, "Pareto smoothed importance sampling," *arXiv preprint arXiv:1507.02646*, 2015.

[10] Y. C. Eldar, "Rethinking biased estimation: Improving maximum likelihood and the cramér–rao bound," *Foundations and Trends in Signal Processing*, vol. 1, no. 4, pp. 305–449, 2008.

[11] S. Kay and Y. C. Eldar, "Rethinking biased estimation," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 133–136, 2008.

[12] H. Kahn and A. W. Marshall, "Methods of reducing sample size in Monte Carlo computations," *Journal of the Operations Research Society of America*, vol. 1, no. 5, pp. 263–278, 1953.

[13] A. Doucet and Adam M A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of Nonlinear Filtering*, vol. 12, no. 656-704, pp. 3, 2009.

[14] 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.

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