On Sample Generation and Weight Calculation in Multiple Importance Sampling

Víctor Elvira*, Luca Martino†, David Luengo‡, Mónica F. Bugallo♯

* Dep. of Signal Theory and Communic., Universidad Carlos III de Madrid, 28911 Leganés (Spain)
† Dep. of Mathematics and Statistics, University of Helsinki, 00014 Helsinki (Finland)
‡ Dep. of Signal Theory and Communic., Universidad Politécnica de Madrid, 28031 Madrid (Spain)
♯ Dep. of Electrical and Computer Eng., Stony Brook University, Stony Brook, NY 11794 (USA)

***Abstract*—Importance sampling is a Monte Carlo technique that approximates moments of target densities by means of weighted samples. These samples are traditionally drawn from a single proposal density. In multiple importance sampling (MIS) a set of different proposal densities is available. In this paper, we propose a formal framework that allows different ways of drawing samples from a set of proposals and different proper weighting functions that can be applied. In particular, we describe three sampling methods and five generic weighting functions. As proper sampling/weighting combinations, six unique MIS schemes (three of them are novel) are discussed throughout the paper. All the methods are analyzed in terms of the variance of the associated estimators, establishing a ranking regarding their performance.**

Keywords*—*Monte Carlo methods, multiple importance sampling, Bayesian inference.

I. Introduction

Importance Sampling (IS) is a well-known Monte Carlo technique to compute integrals involving target probability density functions (pdfs) [1, 2]. The *standard* IS technique draws samples from a single proposal pdf, assigning them weights based on the ratio between the target and the proposal pdfs, both evaluated at the sample value. Although many other proper *random* weight functions can be designed (see [2, Section 2.5] and [1, Section 14.2]), this *deterministic* weight assignment has prevailed in the literature. Within this framework, the choice of a suitable proposal pdf is crucial in order to obtain a good approximation of the desired integral. Therefore, many different strategies have been proposed in the literature to design efficient IS schemes [2, Chapter 2], [3, Chapter 9].

One of the most powerful approaches is based on using a population of proposal pdfs instead of a unique pdf. This approach is usually known in the literature as *multiple* importance sampling (MIS), and several possible implementations have been proposed [4, 5, 6]. In general, MIS strategies provide more robust algorithms, since they do not entrust the performance of the method to a single proposal. Moreover, many adaptive importance sampling (AIS) algorithms have been proposed in order to conveniently update this set of proposals [7, 8, 9].

In this paper, we provide a general framework for sampling and weighting in MIS schemes. First, we extend Liu's notion of properness [2, Section 2.5] from single IS to MIS. Then, three different sampling approaches and five weighting functions are described. Overall, six unique proper sampling/weighting methods (three of them novel), altogether with two simple rules to design additional valid schemes, are developed in the paper. Finally, we establish a ranking of the different MIS schemes in terms of the variance of the associated estimators, showing that one of the discussed combinations outperforms all the approaches available in the literature that we are aware of. An extended version of this work, including all the theoretical derivations (about the variance of the estimators and the effective sample size of the different schemes), detailed discussions on the different sampling/weighting methods, a brief extension to the adaptive MIS scenario, and simulations to quantify the amount of variance reduction, can be found in [10].

This paper is structured as follows. In Section II we introduce the concept of proper IS according to Liu and generalize it for MIS schemes. Then, Sections III and IV describe several valid sampling and weighting schemes, whereas Section V summarizes the different unique sampling/weighting combinations obtained. Finally, Section VI establishes a ranking among the different schemes in terms of the variance of the associated estimators, and Section VII concludes the work.

II. Proper importance sampling

Let us consider a system characterized by a vector of d_x unknown parameters, $\mathbf{x} \in \mathbb{R}^{d_x}$, and a set of d_y observed data related to the hidden parameters through some specific model, $\mathbf{y} \in \mathbb{R}^{d_y}$. Our objective is to approximate the posterior pdf of the parameters $\mathbf{x}$ given the data $\mathbf{y}$,

$$\tilde{\pi}(\mathbf{x}|\mathbf{y}) = \frac{\ell(\mathbf{y}|\mathbf{x})h(\mathbf{x})}{Z(\mathbf{y})} \propto \pi(\mathbf{x}|\mathbf{y}) = \ell(\mathbf{y}|\mathbf{x})h(\mathbf{x}), \qquad (1)$$

where $\ell(\mathbf{y}|\mathbf{x})$ is the likelihood function, $h(\mathbf{x})$ is the prior pdf, and $Z(\mathbf{y})$ is the normalization factor.[1] A common problem is then computing a moment $g(\mathbf{x})$ of $\tilde{\pi}$, i.e., solving the integral

$$I = \int g(\mathbf{x})\tilde{\pi}(\mathbf{x})d\mathbf{x}. \qquad (2)$$

In many cases, I cannot be computed because of the integral itself or because $\tilde{\pi}$ is not fully available (e.g., because $Z = \int \ell(\mathbf{y}|\mathbf{x})h(\mathbf{x})d\mathbf{x}$ cannot be computed either).

[1] In the sequel, in order to simplify the notation, the dependence on $\mathbf{y}$ is removed, e.g., $Z \equiv Z(\mathbf{y})$.

In IS, N samples are drawn from the unique proposal pdf, q, and properly weighted in order to build an estimator of I [1, 2, 3]. Namely, the classical IS estimator is given by

$$\hat{I} = \frac{1}{ZN} \sum_{n=1}^{N} w_n g(\mathbf{x}_n), \qquad (3)$$

where $\mathbf{x}_n \sim q(\mathbf{x})$ and $w_n = w(\mathbf{x}_n) = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}$.

The estimator of Eq. (3) is unbiased and its weights are broadly used in the literature. However, this is not the unique choice of weights. Liu, in [2, Section 2.5], states that a weighted sample $\{\mathbf{x}_n, w_n\}$ drawn from a single proposal q is *proper* if, for any square integrable function g,

$$\frac{E_q[g(\mathbf{x})w(\mathbf{x})]}{E_q[\pi(\mathbf{x})]} = E_{\tilde{\pi}}[g(\mathbf{x})], \qquad (4)$$

i.e., w can be in any form as long as the condition of Eq. (4) is fulfilled. Note that, for a deterministic weight assignment, the only proper weights are the ones considered by the standard IS approach, but many different proper weights could be designed by defining an appropriate probabilistic weight function.

It is well known that the variance of $\hat{I}$ is directly related to the discrepancy between $\tilde{\pi}(\mathbf{x})|g(\mathbf{x})|$ and $q(\mathbf{x})$ [1, 11]. Since the choice of a unique good proposal pdf is critical and can be a very difficult task, in MIS, a set of proposal pdfs, $\{q_n\}_{n=1}^{N}$, are used in order to obtain the samples. In this case, although the MIS estimator is still given by Eq. (3), there is a large number of possibilities for drawing and weighing the set of samples $\{\mathbf{x}_n\}_{n=1}^{N}$, as we discuss in the following sections.

III. Sampling in multiple importance sampling

For the sake of clarity in the explanations, let us consider that we draw N samples from the set of proposals, i.e., the number of samples to be generated is identical to the number of proposal pdfs. However, all the considerations can be automatically extended to the generic case where $M = kN$ samples (with $k \geq 1$ and $k \in \mathbb{N}$) are drawn from each proposal. Note also that the set of available proposal pdfs is, in principle, unweighted, i.e., we do not know which ones provide a better performance for the approximation of I. In this context, we can consider that the samples will be drawn from a unique equal-weighted mixture proposal composed of the N proposal pdfs:

$$\psi(\mathbf{x}) \equiv \frac{1}{N} \sum_{n=1}^{N} q_n(\mathbf{x}). \qquad (5)$$

A. Selection of the proposal pdfs

We consider a sampling mechanism where we select (randomly or not) the sequence of indexes of the proposals $j_{1:N} \equiv j_1, ..., j_N$, and then we draw the samples from the corresponding proposals as $\mathbf{x}_n \sim q_{j_n}(\mathbf{x})$. The full joint distribution of all the samples and indexes is then given by

$$p(\mathbf{x}_{1:N}, j_{1:N}) = P(j_1) \left[\prod_{i=2}^{N} P(j_i | j_{1:i-1}) \right] \left[\prod_{i=1}^{N} p(\mathbf{x}_i | j_i) \right], \qquad (6)$$

and the graphical model associated to the sampling is depicted in Fig. 1.

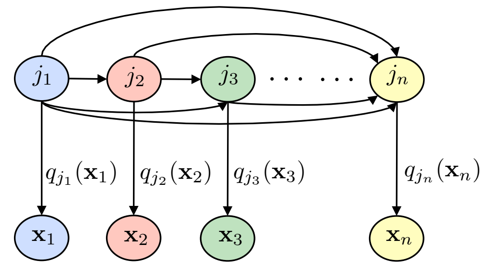


Fig. 1: Graphical model associated to the generic sampling scheme.

In the sequel we describe three different mechanisms for obtaining the sequence of indexes, $j_{1:N}$, of the proposal pdfs: two random mechanisms (with and without replacement) and a deterministic scheme.

$\mathcal{S}_1$: *Random index sampling with replacement*: This is the standard sampling scheme where N indexes are independently drawn from the set $\{1, \ldots, N\}$, i.e., from the multinomial distribution defined by the N possible values, each with probability $1/N$. Therefore, several j_n's may take the same value, i.e., there may ultimately be more than one sample, $\mathbf{x}_n$ for $n = 1, \ldots, N$, generated from the same proposal pdf and there may be proposal pdfs that are not used for generating any samples.

$\mathcal{S}_2$: *Random index sampling without replacement*: In this case, when an index is selected from the set of available values, that particular index is discarded for future generations of indexes. Note that with this strategy, exactly one sample is drawn from each of the proposal pdfs, i.e., the set of sampled indexes $j_{1:N}$ is a permutation of the set $\{1, ..., N\}$.

$\mathcal{S}_3$: *Deterministic index selection (without replacement)*: This sampling is a particular case of sampling $\mathcal{S}_2$ where $j_n = n$. Therefore, the n-th sample is always drawn from q_n, i.e.,

$$\mathbf{x}_n \sim q_{j_n}(\mathbf{x}) = q_n(\mathbf{x}). \qquad (7)$$

This particular index selection procedure has been used by different algorithms in the MIS literature (e.g., APIS [9]), and it is also implicitly present in the particle filtering literature (e.g., bootstrap PF [12]).

While the IS approach focuses just on the distribution of the r.v. $\mathbf{X}_n$ in order to design the weight of the sample $\mathbf{x}_n$, here we are also interested in the distribution of the samples regardless of their index n. The reason is that, in MIS schemes, the N samples can be used jointly regardless of their specific order of appearance. More precisely, we introduce a generic r.v. defined as

$$\mathbf{X} = \mathbf{X}_n \quad \text{with} \quad n \sim \mathcal{U}(\{1, 2, \ldots, N\}), \qquad (8)$$

where $\mathcal{U}(\{1, 2, \ldots, N\})$ is the discrete uniform distribution on the set $\{1, 2, \ldots, N\}$. Namely, the r.v. $\mathbf{X}$ is equal to $\mathbf{X}_n$ chosen uniformly within the set $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$. Therefore, the density of $\mathbf{X}$ is given by the expression[2]

$$f(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N} p_{\mathbf{x}_n}(\mathbf{x}), \quad (9)$$

where $p_{\mathbf{x}_n}(\mathbf{x})$ denotes the marginal pdf of $\mathbf{X}_n$ evaluated at $\mathbf{x}$. Note that, in the sampling schemes with random index selection ($\mathcal{S}_1$ and $\mathcal{S}_2$), we have $f(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N}\psi(\mathbf{x}) = \psi(\mathbf{x})$, since $p_{\mathbf{x}_n}(\mathbf{x}) = \psi(\mathbf{x})$, whereas in the sampling with deterministic index selection ($\mathcal{S}_3$), since $p_{\mathbf{x}_n}(\mathbf{x}) = q_n(\mathbf{x})$, we have $f(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N} q_n(\mathbf{x}) = \psi(\mathbf{x})$. Therefore, in all the sampling schemes $\mathbf{X}$ is distributed according to the mixture $\psi(\mathbf{x})$, as expected. The following remark summarizes the sufficient and necessary condition for any valid sampling scheme within this framework.

Remark 1 (**Sampling**). *In the proposed framework, we consider valid any sequential sampling scheme for generating the set* $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ *such that the pdf of the r.v.* $\mathbf{X}$ *defined in Eq.* (8) *is given by* ψ. *Further considerations about the r.v.* $\mathbf{X}$ *and connections with variance reduction methods [1, 3] are given in the appendix of [10].*

IV. Weighting in multiple importance sampling

A. Proper multiple importance sampling

Our approach is based on analyzing which weighting functions yield *proper* estimators. We consider the definition of properness of Liu, summarized in Eq. (4), and we generalize it to the MIS scenario. This extension is not straightforward, since there are several valid sampling procedures and, in each of them, different conceptual interpretations of what is the proposal pdf of each sample. Therefore, we propose a *generalized properness* condition in the MIS scenario over the whole estimator. Namely, given a specific sampling method, we consider that the set of weighting functions $\{w_n\}_{n=1}^{N}$ is proper if

$$\frac{E_{p(\mathbf{x}_{1:N}, j_{1:N})}\left[\frac{1}{N}\sum_{n=1}^{N} w_n g(\mathbf{x}_n)\right]}{E_{p(\mathbf{x}_{1:N}, j_{1:N})}\left[\frac{1}{N}\sum_{n=1}^{N} w_n\right]} = E_{\tilde{\pi}}[g(\mathbf{x})]. \quad (10)$$

This is equivalent to imposing the restriction

$$\frac{E_{p(\mathbf{x}_{1:N}, j_{1:N})}\left[Z\hat{I}\right]}{E_{p(\mathbf{x}_{1:N}, j_{1:N})}\left[\hat{Z}\right]} = I, \quad (11)$$

which is fulfilled if $E[\hat{I}] = I$ and $E[\hat{Z}] = Z$. Note that the MIS properness is fulfilled by any set of weighting functions $\{w_n\}_{n=1}^{N}$ that yield an unbiased generic estimator $\hat{I}$, i.e., $E[\hat{I}] = I$. Note also that this is a generalization of Liu-properness, in the sense that all the weighting functions that are Liu-proper also fulfill the restriction of Eq. (11), but the opposite is not necessarily true.

In this framework, we impose the weight function to have the (deterministic) structure $w_n = \frac{\pi(\mathbf{x}_n)}{\varphi_{\mathcal{P}_n}(\mathbf{x}_n)}$, where π is the target and $\varphi_{\mathcal{P}_n}$ is a generic function parametrized by a set of parameters $\mathcal{P}_n$,[3] and both terms are evaluated at $\mathbf{x}_n$. Note that $\varphi_{\mathcal{P}_n}$ represents the proposal pdf from which it is interpreted that the n-th sample is drawn. It is on this interpretation of what the proposal pdf used for the generation of the sample is (the evaluation of the denominator in the weight calculation) that different weighting strategies can be devised. Considering the joint pdf of samples and indexes, given by Eq. (6) in the proposed framework, the expectation of the generic estimator $\hat{I}$ is given by

$$E[\hat{I}] = \frac{1}{ZN}\sum_{n=1}^{N}\sum_{j_{1:N}}\int \frac{\pi(\mathbf{x}_n) g(\mathbf{x}_n)}{\varphi_{\mathcal{P}_n}(\mathbf{x}_n)} P(j_{1:N}) p(\mathbf{x}_n|j_n) d\mathbf{x}_n. \quad (12)$$

Then, we consider as valid any weighting scheme that results in an unbiased estimator, as stated by the following remark.

Remark 2 (**Weighting**). *In the proposed framework, we consider valid any weighting scheme (i.e., any function* $\varphi_{\mathcal{P}_n}$ *at the denominator of the weight) that yields* $E[\hat{I}] \equiv I$ *in Eq.* (12).

B. General weighting functions

Here we present five possible generic functions, $\varphi_{\mathcal{P}_n}$, that yield an unbiased estimator $\hat{I}$. The functions $\varphi_{\mathcal{P}_n}$ are said to be generic because they are defined as distributions of $\mathbf{X}_n$ and $\mathbf{X}$, and therefore they yield different specific functions under different valid sampling schemes (see Remark 2). [4]

- $\mathcal{W}_1$: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = p(\mathbf{x}_n|j_{1:n-1})$: It interprets the proposal pdf as the conditional density of $\mathbf{x}_n$ given all the previous proposal indexes of the sequential sampling process.
- $\mathcal{W}_2$: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = p(\mathbf{x}_n|j_n) = q_{j_n}(\mathbf{x}_n)$: It interprets that φ is the proposal q_{j_n} when the index j_n is known.
- $\mathcal{W}_3$: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = p(\mathbf{x}_n)$: It interprets that $\mathbf{x}_n$ is a realization of the marginal $p(\mathbf{x}_n)$. This is probably the most "natural" option (as it does not assume any further knowledge in the generation of $\mathbf{x}_n$) and is a usual choice for the calculation of the weights in some of the existing MIS schemes.
- $\mathcal{W}_4$: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = f(\mathbf{x}_n|j_{1:N}) = \frac{1}{N}\sum_{k=1}^{N} q_{j_k}(\mathbf{x}_n)$: This interpretation makes use of the distribution of the r.v. $\mathbf{X}$ conditioned on the whole set of indexes.
- $\mathcal{W}_5$: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = f(\mathbf{x}_n) = \frac{1}{N}\sum_{k=1}^{N} q_k(\mathbf{x}_n)$
  This option considers that all the $\mathbf{x}_n$ are realizations of the r.v. $\mathbf{X}$ defined in Eq. (8) (see the appendix of [10] for a thorough discussion of this interpretation).

Although some of the selected functions $\varphi_{\mathcal{P}_n}$ may seem more natural than others, all of them yield valid estimators. The proofs can be found in the appendix of [10].

[2] We use a simplified argument-wise notation where $f(\mathbf{x})$ denotes de pdf of the r.v. $\mathbf{X}$ of Eq. (8).

[3] The parametrization in $\varphi_{\mathcal{P}_n}(\mathbf{x}_n)$ always corresponds to a subset (including the empty set) of the sequence of indexes $j_{1:N}$, i.e., $\mathcal{P}_n \subseteq \{j_1, ..., j_N\}$.

[4] From now on, $p(\cdot)$ and $f(\cdot)$, which correspond to the pdfs of $\mathbf{X}_n$ and $\mathbf{X}$ respectively, are used as functions, and the argument represents a functional evaluation.

TABLE I: Specific function, $\varphi_{\mathcal{P}_n}$, at the denominator of weight, $w_n = \frac{\pi(\mathbf{x}_n)}{\varphi_{\mathcal{P}_n}(\mathbf{x}_n)}$, resulting from the combination of the different sampling schemes (Section III) and weighting functions (Section IV).

$\varphi_{\mathcal{P}_n}$	$\mathcal{W}_1$ $p(\mathbf{x}_n \mid j_{1:n-1})$	$\mathcal{W}_2$ $p(\mathbf{x}_n \mid j_n)$	$\mathcal{W}_3$ $p(\mathbf{x}_n)$	$\mathcal{W}_4$ $f(\mathbf{x} \mid j_{1:N})$	$\mathcal{W}_5$ $f(\mathbf{x})$
$\mathcal{S}_1$: **with replacement**	$\psi(\mathbf{x}_n)$ [R3]	$q_{j_n}(\mathbf{x}_n)$ [R1]	$\psi(\mathbf{x}_n)$ [R3]	$\frac{1}{N}\sum_{k=1}^{N} q_{j_k}(\mathbf{x}_n)$ [R2]	$\psi(\mathbf{x}_n)$ [R3]
$\mathcal{S}_2$: **w/o (random)**	$\frac{1}{\lvert\mathcal{I}_n\rvert}\sum_{\forall k\in\mathcal{I}_n} q_k(\mathbf{x}_n)$ [N2]	$q_{j_n}(\mathbf{x}_n)$ [N1]	$\psi(\mathbf{x}_n)$ [N3]	$\psi(\mathbf{x}_n)$ [N3]	$\psi(\mathbf{x}_n)$ [N3]
$\mathcal{S}_3$: **w/o (deterministic)**	$q_n(\mathbf{x}_n)$ [N1]	$q_n(\mathbf{x}_n)$ [N1]	$q_n(\mathbf{x}_n)$ [N1]	$\psi(\mathbf{x}_n)$ [N3]	$\psi(\mathbf{x}_n)$ [N3]

TABLE II: Summary of the sampling procedure and the weighting function of each MIS scheme.

MIS scheme	Sampling	$w(\mathbf{x}_n)$	Used in
R1	$\mathcal{S}_1$	$\frac{\pi(\mathbf{x}_n)}{q_{j_n}(\mathbf{x}_n)}$	Novel
R2	$\mathcal{S}_1$	$\frac{\pi(\mathbf{x}_n)}{\frac{1}{N}\sum_{k=1}^{N} q_{j_k}(\mathbf{x}_n)}$	Novel
R3	$\mathcal{S}_1$	$\frac{\pi(\mathbf{x}_n)}{\psi(\mathbf{x}_n)}$	[8]
N1	$\mathcal{S}_3$	$\frac{\pi(\mathbf{x}_n)}{q_n(\mathbf{x}_n)}$	[7]
N2	$\mathcal{S}_2$	$\frac{\pi(\mathbf{x}_n)}{\frac{1}{\lvert\mathcal{I}_n\rvert}\sum_{\forall k\in\mathcal{I}_n} q_k(\mathbf{x}_n)}$	Novel
N3	$\mathcal{S}_3$	$\frac{\pi(\mathbf{x}_n)}{\psi(\mathbf{x}_n)}$	[9, 13]

V. Multiple importance sampling schemes

In this section, we describe the different possible combinations of the three sampling strategies considered in Section III and the five generic weighting functions devised in Section IV. Note that, although there are fifteen possibilities, they only lead to six unique MIS methods. Table I shows the possible combinations of sampling/weighting, indicating the function $\varphi_{\mathcal{P}_n}$ at the denominator of the weight and the resulting MIS method within brackets. The six MIS methods are labeled either by an R (indicating that the method uses sampling with *replacement*) or with an N (denoting that the method corresponds to a sampling scheme with *no* replacement). We remark that these schemes are examples of proper MIS techniques fulfilling Remarks 1 and 2, and that many other choices could also be valid within this framework. Table II summarizes the six unique schemes, providing the sampling procedure for $\mathbf{x}_n$ and its associated weight w_n. Figs. 2, 3, and 4 show realizations of the sequence of indexes $j_{1:N}$ (with $N = 4$) and the function $\varphi_{\mathcal{P}_n}$ at the denominator of the weight, in the different MIS schemes corresponding to sampling procedures $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$, respectively.

VI. Variance Analysis

The six different MIS schemes that appear in this paper are proper and thus yield an unbiased estimator $\hat{I}$. However, although obtaining an IS estimator with finite variance essentially amounts to having a proposal with heavier tails than the target (see [1, 14] for sufficient conditions that guarantee this finite variance), the performance of estimators based on different sampling/weighting schemes can be dramatically different in terms of variance. In the following, we state two theorems relating the variance of the different MIS schemes.

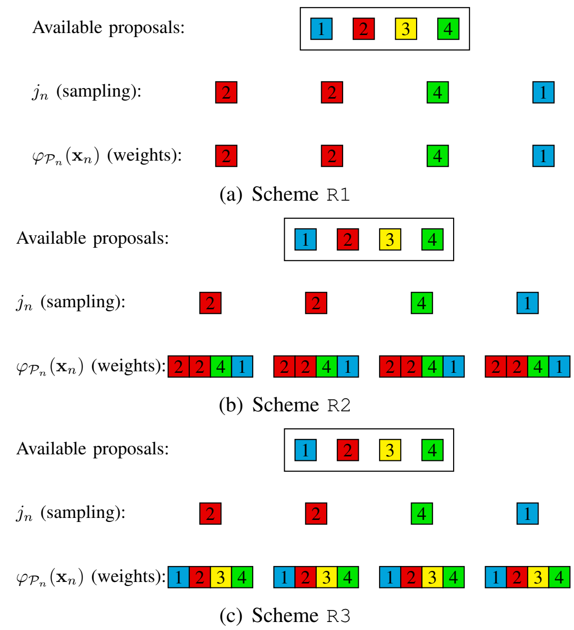


Fig. 2: Example of a realization of the indexes selection ($N = 4$) with the procedure $\mathcal{S}_1$ (with replacement), and the corresponding possible denominators for weighting: (a) R1: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = q_{j_n}(\mathbf{x}_n)$; (b) R2: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = \frac{1}{N}\sum_{k=1}^{N} q_{j_k}(\mathbf{x}_n)$; (c) R3: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = \psi(\mathbf{x}_n) = \frac{1}{N}\sum_{k=1}^{N} q_k(\mathbf{x}_n)$.

Theorem 1. *For any target π, any square integrable function g, and any set of proposal densities $\{q_n\}_{n=1}^{N}$ such that the variance of the corresponding MIS estimators is finite,*

$$Var(\hat{I}_{R1}) = Var(\hat{I}_{N1}) \geq Var(\hat{I}_{R3}) \geq Var(\hat{I}_{N3})$$

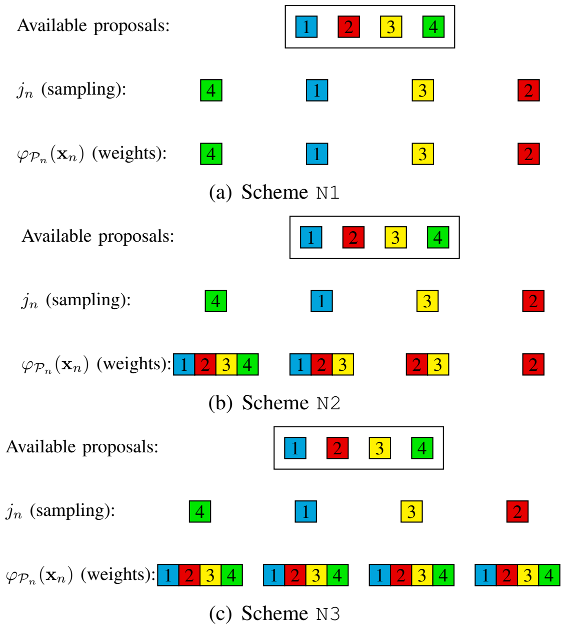
Proof: See the appendix of [10]. □

Theorem 2. *For any target π, any square integrable function g, and any set of proposal densities $\{q_n\}_{n=1}^{2}$ such that the variance of the corresponding MIS estimators is finite,*

$$Var(\hat{I}_{R1}) = Var(\hat{I}_{N1}) \geq Var(\hat{I}_{R2}) = Var(\hat{I}_{N2}) \geq Var(\hat{I}_{N3}) \quad (13)$$

Proof: See the appendix of [10]. □

Note that scheme N3 outperforms (in terms of variance) any other MIS scheme in the literature that we are aware of. Moreover, for $N = 2$, it also outperforms the other novel schemes R2 and N2. While the MIS schemes R2 and N2 do not appear in Theorem 1, we hypothesize that the conclusions of Theorem 2 might be extended to $N > 2$.

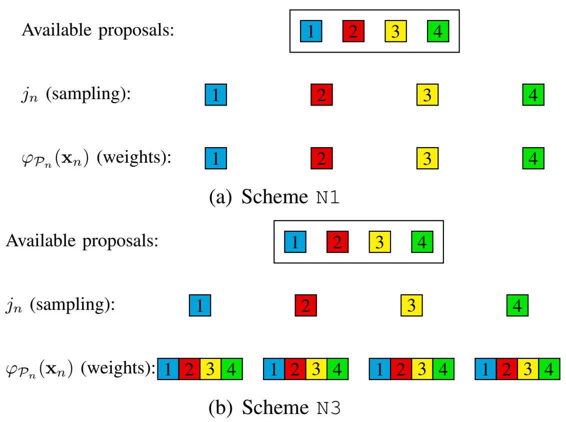


(a) Scheme `N1`

(b) Scheme `N2`

(c) Scheme `N3`

Fig. 3: Example of a realization of the indexes selection ($N = 4$) with the procedure $\mathcal{S}_2$ (without replacement), and the corresponding possible denominators for weighting: (a) `N1`: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = q_{j_n}(\mathbf{x}_n)$; (b) `N2`: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = \frac{1}{|\mathcal{I}_n|}\sum_{\forall k \in \mathcal{I}_n} q_k(\mathbf{x}_n)$; (c) `N3`: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = \frac{1}{N}\sum_{k=1}^{N} q_k(\mathbf{x}_n)$.



(a) Scheme `N1`

(b) Scheme `N3`

Fig. 4: Indexes selection ($N = 4$) with the procedure $\mathcal{S}_3$ (without replacement), and the corresponding possible denominators for weighting: (a) `N1`: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = q_{j_n}(\mathbf{x}_n)$; (b) `N3`: $\varphi_{\mathcal{P}_n}(\mathbf{x}_n) = \psi(\mathbf{x}_n) = \frac{1}{N}\sum_{k=1}^{N} q_k(\mathbf{x}_n)$.

VII. Conclusions

In this work, we have introduced a unified framework for sampling and weighting in multiple importance sampling (MIS). This framework extends Liu's concept of a proper weighted sample, enabling the design of a wide range of valid sampling/weighting combinations. In particular, we have described three specific sampling procedures and we have proposed five types of generic weighting functions, leading to six unique valid schemes (three of them novel). Moreover, we have established a ranking of the different methods in terms of the variance of the associated estimators.

Acknowledgment

This work has been supported by the Spanish government's projects DISSECT (TEC2012-38058-C03-01), AGES (S2010/BMD-2422), ALCIT (TEC2012-38800-C03-01), COMPREHENSION (TEC2012-38883-C02-01), and OTOSiS (TEC2013-41718-R); by the BBVA Foundation through project MG-FIAR ("I Convocatoria de Ayudas Fundación BBVA a Investigadores, Innovadores y Creadores Culturales"); by the National Science Foundation under Award CCF-0953316; and by the European Union's 7th FP through the Marie Curie ITN MLPM2012 (Grant No. 316861).

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] A. Owen, *Monte Carlo theory, methods and examples*, http://statweb.stanford.edu/~owen/mc/, 2013.
[4] E. Veach and L. Guibas, "Optimally combining sampling techniques for Monte Carlo rendering," *In SIGGRAPH 1995 Proceedings*, pp. 419–428, 1995.
[5] A. Owen and Y. Zhou, "Safe and effective importance sampling," *Journal of the American Statistical Association*, vol. 95, no. 449, pp. 135–143, 2000.
[6] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Efficient multiple importance sampling estimators," *Signal Processing Letters, IEEE*, vol. 22, no. 10, pp. 1757–1761, 2015.
[7] 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.
[8] 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.
[9] 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.
[10] V. Elvira, L. Martino, D. Luengo, and M.F. Bugallo, "Generalized multiple importance sampling," *arXiv preprint arXiv:1511.03095*, 2015.
[11] 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.
[12] N. Gordon, D. Salmond, and A. F. M. Smith, "Novel approach to nonlinear and non-Gaussian Bayesian state estimation," *IEE Proceedings-F Radar and Signal Processing*, vol. 140, pp. 107–113, 1993.
[13] 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.
[14] J. Geweke, "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, vol. 24, pp. 1317–1399, 1989.