

RESEARCH

Open Access



Multiple importance sampling revisited: breaking the bounds

Mateu Sbert^{1,2*} , Vlastimil Havran³ and Laszlo Szirmay-Kalos⁴

Abstract

We revisit the multiple importance sampling (MIS) estimator and investigate the bound on the efficiency improvement over balance heuristic estimator with equal count of samples established in Veach's thesis. We revise the proof for this and come to the conclusion that there is no such bound and henceforth it makes sense to look for new estimators that improve on balance heuristic estimator with equal count of samples. Next, we examine a recently introduced non-balance heuristic MIS estimator that is provably better than balance heuristic with equal count of samples, and we improve it both in variance and efficiency. We then obtain an equally provably better one-sample balance heuristic estimator, and finally, we introduce a heuristic for the count of samples that can be used when the individual techniques are biased. All in all, we present three new sampling strategies to improve on both variance and efficiency on the balance heuristic using non-equal count of samples.

Our scheme requires the previous knowledge of several quantities, but those can be obtained in an adaptive way. The results also show that by a careful examination of the variance and properties of the estimators, even better estimators could be discovered in the future. We present examples that support our theoretical findings.

Keywords: Global illumination, Rendering equation analysis, Multiple importance sampling, Monte Carlo

AMS Subject Classification: primary Computer Graphics G.3 Mathematics of Computing/PROBABILITY AND STATISTICS probabilistic algorithms; Computer Graphics G.3 Mathematics of Computing / PROBABILITY AND STATISTICS probabilistic algorithms; secondary Computer Graphics G.3 Mathematics of Computing / PROBABILITY AND STATISTICS probabilistic algorithms

1 Introduction

The *multiple importance sampling* (MIS) estimator [1, 2], and in particular *balance heuristic*, which is equivalent to the Monte Carlo estimator with a mixture of probability density functions (pdfs), has been used for many years with a big success, being a reliable and robust estimator that allows an easy and straightforward combination of different sampling techniques. MIS with balance heuristic estimator has been almost exclusively used with equal count of samples for each technique, mainly following the recommendations based on certain quasi-optimality theorems in Veach's thesis. Specifically, these theorems stated that a balance heuristic estimator could not be worse

than n times any other MIS estimator with equal count of samples where n is the number of combined methods. In papers addressing the variance of MIS, strategies were assumed to have equal number of samples [3, 4], and the combination of MIS with jittered sampling was studied in [5].

MIS is often used in rendering applications of computer graphics [6] where the reflected illumination is the product of the intensity in the illumination direction that can be represented by an environment map, and the probability of the reflection from a surface. As it is not feasible to sample with this product of two functions, sampling mimics either directions from where reflection is likely or directions of high intensity. The first approach is called light source sampling and the second approach is called BRDF sampling after the name of the bidirectional reflection distribution function (BRDF).

*Correspondence: mateu@ima.udg.edu

¹School of Computer Science and Technology, Tianjin University, Tianjin, China

²Department of Informatics, Applied Mathematics and Statistics, University of Girona, Campus de Montilivi, Girona, Spain

Full list of author information is available at the end of the article

Lu et al. [7] propose an improvement to balance heuristic estimators for environment map illumination by using a Taylor's second order approximation of the variance around the equal weights $1/2$ to obtain the counts of samples according to the BRDF and the environment map, which is accurate only if the optimal sample numbers are not too far from equal sampling. Heuristics have also been used that prefer certain sampling methods based on the local properties, for example, BRDF sampling is advantageous on highly specular surfaces and light source sampling on diffuse surfaces. Pajot et al. proposed a framework, called representativity, to develop such heuristic [8]. Havran and Sbert [9, 10] introduce a heuristic for unbiased techniques based on assigning a count of samples proportional to the inverse of the variance of each technique that in some cases seemed to put into question the quasi-optimality rules of equal sampling. More recently, based on the analysis of the variance of MIS estimators, Sbert et al. [11] discovered a MIS non-balance heuristic estimator that is provably better than balance heuristic with equal count of samples. The introduction of costs into those schemes allowed an even bigger increase in efficiency. We improve on their work by clarifying the quasi-optimality rules in Veach's thesis by introducing new balance heuristic estimators that are provably better than balance heuristic with equal count of samples, and by introducing a new heuristic valid when the individual estimators are biased. We believe that our work can inspire new provably better estimators or at least sound heuristics for heuristically better estimators.

Putting our results in a wider perspective, we should consider adaptive variance reduction methods that learn some properties of the integrand, e.g., the location of its peak, and refine the sampling strategy on-the-fly by analyzing the samples that have been already generated. In case of mixing different pdfs with which sampling is straightforward, the weights of the individual methods are the target of the adaptation. Adaptive methods can be imagined as a sequence of Monte Carlo quadrature steps and parameter estimation steps. Care should be exercised to implement adaptive methods since making the sampling pdf dependent on the generated samples might make the estimator biased, which can be attacked by freezing the adaptation after certain samples letting the approach be consistent. Another challenge is the control of the adaptation with randomly sampled, i.e., noisy data. To handle this, adaptive MCMC approaches often choose an artificial but robust criterion, like the acceptance rate, instead of the variance of the estimation [12, 13]. In computer graphics, this technique has been used to control the mutation of MCMC methods [14, 15] and to distribute different numbers of samples among the different techniques [16].

Owen and Zhou [17] survey the principles of adaptively mixing different sampling strategies, including defensive importance sampling and the simultaneous application of control variates and importance sampling. They also investigate MIS and point out that there is a room for improvement. Douc et al. [18] investigated the population Monte Carlo method and derived sufficient convergence conditions for adaptive mixtures, and also iterative formulae for the optimal weights [19]. As population Monte Carlo explores the sampling space with a Markov stochastic process, it uses the information of only the current samples directly. Cornuet et al. [20] improve this in their adaptive multiple importance sampling algorithm and present optimal exploitation of previous samples, which automatically stabilizes the process. Marin et al. [21] argue that the Markov property is important to be able to prove the consistency of the estimator, and the stability of the process can be achieved by increasing the sample size in each iteration.

In their recent work, Martino, Elvira et al. proposed adaptive population importance sampling [22] and gradient adaptive population importance sampling [23] that adapt the mean and the covariance of proposal Gaussian pdfs, and reported significant improvements with respect to MIS that keeps the individual proposal pdfs fixed. We note that in computer graphics the not continuous integrands may pose problems to such gradient-based methods. Elvira et al. also examined a discrete weighting scheme [24] where only a subset of available sampling strategies is selected with the aim of reducing the computational complexity of the sampling process.

The rest of the paper is organized as follows: In Section 2, we review the basic formulae for the variance of MIS. In Section 3, we discuss Theorem 9.5 of Veach's thesis and present a new proof for a slightly modified theorem. In Section 4, we present a multi-sample balance heuristic estimator that is provably better than multi-sample balance heuristic with equal count of samples. In Section 5, we obtain a provably better one-sample balance heuristic estimator, and in Section 6, we justify a new heuristic for the count of samples that does not require the individual techniques be unbiased. In Section 7, we discuss the results, and in Section 8, summarize our conclusions. Some proofs are given in Appendices A and B, and numeric 1D examples are presented in Appendix C.

2 MIS variance analysis

We review here the variance of MIS estimators. The naming convention for the estimators is given in Table 1.

2.1 General multi-sample MIS estimator

The MIS estimator introduced by Veach and Guibas [1] to estimate the value of integral $I = \int f(x)dx$ has the following form:

Table 1 Naming convention for the MIS estimators in this paper. We will drop the superindex 1 from primary estimators when not strictly necessary

F	Generic multi-sample MIS estimator
\mathcal{F}	Generic one-sample MIS estimator
\mathcal{F}^1	Generic one-sample MIS primary estimator
\hat{F}	Generic multi-sample MIS balance heuristic estimator
$\hat{\mathcal{F}}$	Generic one-sample MIS balance heuristic estimator
$F_{eq}, \mathcal{F}_{eq}, \mathcal{F}_{eq}^1, \hat{F}_{eq}, \hat{\mathcal{F}}_{eq}$	Equal count of samples estimators

$$F = \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} w_i(X_{ij}) \frac{f(X_{ij})}{p_i(X_{ij})}, \tag{1}$$

where the weights $w_i(X_{ij})$ are such that

$$f(x) \neq 0 \Rightarrow \sum_{i=1}^n w_i(x) = 1 \tag{2}$$

and

$$p_i(x) = 0 \Rightarrow w_i(x) = 0. \tag{3}$$

In this combination scheme, sampling method i uses probability density function $p_i(x)$ to generate N_i number of random samples $\{X_{ij}\}$, ($j = 1, \dots, N_i$). If we have n techniques, the total number of samples is $\sum_{i=1}^n N_i = N$. Integral estimator F is unbiased, as its expected value μ is equal to integral I :

$$\begin{aligned} \mu = E[F] &= \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \int \frac{w_i(x)f(x)}{p_i(x)} p_i(x) dx \\ &= \int \sum_{i=1}^n w_i(x)f(x) dx = \int f(x) dx. \end{aligned} \tag{4}$$

The variance of the estimator is given in the proof of Theorem 9.2 of Veach's thesis [2]. Define F_{ij} as

$$F_{ij} = w_i(X_{ij}) \frac{f(X_{ij})}{p_i(X_{ij})}. \tag{5}$$

For a fixed method i and all j values, the estimators $\{F_{ij}\}$ are independent identically distributed random variables with expected value μ_i :

$$\mu_i = E[F_{ij}] = \int \frac{w_i(x)f(x)}{p_i(x)} p_i(x) dx \tag{6}$$

$$= \int w_i(x)f(x) dx. \tag{7}$$

Observe that

$$\mu = \sum_{i=1}^n \mu_i, \tag{8}$$

and that the variance of F_{ij} is

$$\begin{aligned} \sigma_i^2 &= E[F_{ij}^2] - E^2[F_{ij}] \\ &= \int \left(\frac{w_i(x)f(x)}{p_i(x)} \right)^2 p_i(x) dx - \mu_i^2 \\ &= \int \frac{w_i^2(x)f^2(x)}{p_i(x)} dx - \mu_i^2. \end{aligned} \tag{9}$$

If samples are statistically independent, the variance of the integral estimator is

$$\begin{aligned} V[F] &= \sum_{i=1}^n \frac{1}{N_i^2} \sum_{j=1}^{N_i} \sigma_i^2 = \sum_{i=1}^n \frac{1}{N_i} \sigma_i^2 \\ &= \sum_{i=1}^n \frac{1}{N_i} \left(\int \frac{w_i^2(x)f^2(x)}{p_i(x)} dx - \mu_i^2 \right) \\ &= \sum_{i=1}^n \int \frac{w_i^2(x)f^2(x)}{N_i p_i(x)} dx - \sum_{i=1}^n \frac{1}{N_i} \mu_i^2. \end{aligned} \tag{10}$$

There are two different components of a MIS estimator that determine the variance of the estimates: weighting scheme w_i and the number of samples N_i allocated to different methods. Weighting functions may or may not depend on the number of samples. Additionally, the number of samples can be pre-determined, called the *multi-sample model*, or selected also randomly, which leads to the *one-sample model*.

2.2 Balance heuristic estimator

Veach defined a balance heuristic estimator setting the weights as the following:

$$w_i(x) = \frac{N_i p_i(x)}{\sum_{k=1}^n N_k p_k(x)}. \tag{11}$$

These weights lead to the following estimator:

$$\begin{aligned} \hat{F} &= \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{N_i f(X_{ij})}{\sum_{k=1}^n N_k p_k(X_{ij})} \\ &= \sum_{i=1}^n \sum_{j=1}^{N_i} \frac{f(X_{ij})}{\sum_{k=1}^n N_k p_k(X_{ij})}. \end{aligned} \tag{12}$$

Let us now consider the variance of the combined estimator when method i takes N_i samples and pre-determining the number of samples in advance (multi-sample model):

$$\begin{aligned}
V[\hat{F}] &= \sum_{i=1}^n \frac{1}{N_i} \sigma_i^2 \quad (13) \\
&= \sum_{i=1}^n \frac{1}{N_i} \left(\int \frac{N_i^2 f^2(x) p_i(x)}{(\sum_{k=1}^n N_k p_k(x))^2} dx - \mu_i^2 \right) \\
&= \sum_{i=1}^n \frac{1}{N_i} \int \frac{N_i^2 f^2(x) p_i(x)}{(\sum_{k=1}^n N_k p_k(x))^2} dx - \sum_{i=1}^n \frac{1}{N_i} \mu_i^2 \\
&= \int \frac{f^2(x)}{\sum_{k=1}^n N_k p_k(x)} dx - \sum_{i=1}^n \frac{1}{N_i} \mu_i^2 \\
&= \frac{1}{N} \left(\int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} dx - \sum_{i=1}^n \frac{1}{\alpha_i} \mu_i^2 \right),
\end{aligned}$$

where $\alpha_k = N_k/N$.

2.2.1 Balance heuristic with equal count of samples

Observe that if we take equal count of samples for each technique, i.e., for all i , $N_i = N/n$ or $\alpha_i = 1/n$, then the weights become

$$w_i(x) = \frac{p_i(x)}{\sum_{k=1}^n p_k(x)}, \quad (14)$$

the estimator is

$$\hat{F}_{eq} = \frac{n}{N} \sum_{i=1}^n \sum_{j=1}^{\frac{N}{n}} \frac{f(X_{ij})}{\sum_{k=1}^n p_k(X_{ij})}, \quad (15)$$

and the variance is equal to

$$\begin{aligned}
V[\hat{F}_{eq}] &= \frac{n}{N} \sum_{i=1}^n \sigma_{i,eq}^2 \quad (16) \\
&= \frac{n}{N} \sum_{i=1}^n \left(\int \frac{p_i(x) f^2(x)}{(\sum_{k=1}^n p_k(x))^2} dx - \mu_{i,eq}^2 \right) \\
&= \frac{n}{N} \left(\int \frac{f^2(x)}{\sum_{k=1}^n p_k(x)} dx - \sum_{i=1}^n \mu_{i,eq}^2 \right),
\end{aligned}$$

where $\sigma_{i,eq}^2$ is equal to

$$\sigma_{i,eq}^2 = \int \frac{p_i(x) f^2(x)}{(\sum_{k=1}^n p_k(x))^2} dx - \mu_{i,eq}^2, \quad (17)$$

and $\mu_{i,eq}$ is

$$\mu_{i,eq} = \int \frac{p_i(x) f(x)}{\sum_{k=1}^n p_k(x)} dx. \quad (18)$$

2.3 A provably better non-balance heuristic estimator

Sbert et al. [11] introduced an estimator provably better than balance heuristic with equal count of samples. Taking the weights as in Eq. 14, the estimator F in Eq. 1 is

$$F = \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{f(X_{ij})}{\sum_{k=1}^n p_k(X_{ij})}. \quad (19)$$

The estimators F_{ij} in Eq. 5 become

$$F_{ij} = \frac{f(X_{ij})}{\sum_{k=1}^n p_k(X_{ij})}, \quad (20)$$

and the variances σ_i^2 defined by Eq. 9 are equal to $\sigma_{i,eq}^2$. The variance, Eq. 10, is then

$$\begin{aligned}
V[F] &= \sum_{i=1}^n \frac{1}{N_i} \sigma_{i,eq}^2 \quad (21) \\
&= \sum_{i=1}^n \frac{1}{N_i} \left(\int \frac{p_i(x) f^2(x)}{(\sum_{k=1}^n p_k(x))^2} dx - \mu_{i,eq}^2 \right).
\end{aligned}$$

Sbert et al. showed [11] that taking $N_i \propto \sigma_{i,eq}$ guarantees that we improve on balance heuristic with equal sampling by using estimator Eq. 19, up to the statistical error in estimating the $\sigma_{i,eq}$ values.

Taking into account the cost c_i of sampling with technique i , the sample numbers that guarantee that we improve the efficiency (i.e., the inverse of cost times variance) over balance heuristic with equal sampling are

$$N_i \propto \frac{\sigma_{i,eq}}{\sqrt{c_i}}. \quad (22)$$

3 Breaking the bounds for the relative acceleration

In this section, we investigate the bounds for the improvement on the variance of balance heuristic estimator with equal count of samples.

This problem has been attacked by Veach establishing an inequality (Theorem 9.5 of his thesis [2]) for the variance of the balance heuristic estimator with equal count of sampling \hat{F}_{eq} :

$$V[\hat{F}_{eq}] \leq nV[F] + \frac{n-1}{N} \mu^2, \quad (23)$$

where F is any multiple importance sampling estimator using the same total number of samples N . Veach interpreted Theorem 9.5 as a proof of quasi-optimality of balance heuristic with equal count of samples, saying ‘‘According to this result, changing the N_i can improve the variance by at most a factor of n , plus a small additive term. In contrast, a poor choice of the w_i can increase variance by an arbitrary amount. Thus, the sample allocation is not as important as choosing a good combination strategy.’’

The proof of Eq. 23 is based on the following inequality, which compares a general multiple importance sample estimator F with arbitrary number of samples $\{N_i\}$ with the same estimator (i.e., using the same weights w_i) but with equal count of samples, F_{eq} as follows:

$$V[F] \geq \frac{1}{n} V[F_{eq}]. \quad (24)$$

But this inequality is not valid when the weights $w_i(x)$ depend on the number of samples N_i , see Appendix A for a proof. Just to give a single counter example, let us consider the case when zero variance estimator is possible by properly setting the number of samples, making $V[F]$ zero, but the equal count of samples estimator will not have zero variance. However, we show that Theorem 9.5 can be generalized to such cases as well, but it requires the full reconsideration of the original proof.

The interpretation by Veach of Theorem 9.5 is based on the assumption that additive term $\mu^2(n-1)/N$ is small if the total number of samples, N , gets larger. However, denominator N is implicitly included in the other terms of Eq. 23 as well, thus the considered additive term is, in fact, not negligible. As a result, the selection of N_i sample numbers or weights α_i can make a significant difference in the variance, which is worth examining and opens possibilities to find better estimators.

3.1 One-sample balance heuristic

The general MIS one-sample primary estimator is

$$\mathcal{F}^1 = \frac{w_i(x)f(x)}{\alpha_i p_i(x)}, \quad (25)$$

where technique i is selected with probability α_i . It can be easily shown that it is unbiased, i.e., its expected value is μ . Using the balance heuristic weights,

$$w_i(x) = \frac{\alpha_i p_i(x)}{\sum_{k=1}^n \alpha_k p_k(x)}, \quad (26)$$

the estimator becomes the *one-sample balance heuristic estimator*,

$$\hat{\mathcal{F}}^1 = \frac{f(x)}{\sum_{k=1}^n \alpha_k p_k(x)}. \quad (27)$$

One-sample balance heuristic is the same as the Monte Carlo estimator using the mixture of probabilities $p(x) = \sum_{k=1}^n \alpha_k p_k(x)$, $\sum_{k=1}^n \alpha_k = 1$. The α_i values are called the mixture coefficients and represent the average count of samples from each technique. The variance of this estimator can be obtained by the application of the definition of variance,

$$V[\hat{\mathcal{F}}^1] = \int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} dx - \mu^2. \quad (28)$$

Theorem 1 *If $V[\hat{\mathcal{F}}_{\text{eq}}^1]$ is the variance of the one-sample balance heuristic estimator with equal weights, and $V[\hat{\mathcal{F}}^1]$ the variance of the one-sample balance heuristic estimator with any distribution of weights $\{\alpha_k\}$, then the following inequality holds:*

$$V[\hat{\mathcal{F}}_{\text{eq}}^1] \leq n\alpha_{\max} V[\hat{\mathcal{F}}^1] + (n\alpha_{\max} - 1)\mu^2. \quad (29)$$

Proof The variance of the one-sample balance heuristic with equal weights is

$$V[\hat{\mathcal{F}}_{\text{eq}}^1] = \int \frac{f^2(x)}{\frac{1}{n} \sum_k p_k(x)} dx - \mu^2. \quad (30)$$

As $\alpha_{\max} \sum_{k=1}^n p_k(x) \geq \sum_{k=1}^n \alpha_k p_k(x)$ where $\alpha_{\max} > 0$ is the maximum of the values of α_k and p_k is not negative, we have $\frac{1}{n} \alpha_{\max} \sum_{k=1}^n p_k(x) \geq \frac{1}{n} \sum_{k=1}^n \alpha_k p_k(x)$ and thus:

$$\begin{aligned} V[\hat{\mathcal{F}}_{\text{eq}}^1] &\leq n\alpha_{\max} \int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} dx - \mu^2 \\ &= n\alpha_{\max} \left(\int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} dx - \mu^2 \right) \\ &\quad + (n\alpha_{\max} - 1)\mu^2. \end{aligned} \quad (31)$$

□

When for all i , $\alpha_i = 1/n$, Eq. 29 becomes an equality. Observing that $\alpha_{\max} \leq 1$, the following corollary is immediate.

Corollary 1

$$V[\hat{\mathcal{F}}_{\text{eq}}^1] \leq nV[\hat{\mathcal{F}}^1] + (n-1)\mu^2. \quad (32)$$

Equations 29 and 32 do not imply that the improvement with respect to the equal count of samples is limited by a factor of n , since μ can be large in comparison with the variances. So, it is worth trying to obtain α_i values that can reduce the variance.

Equations 29 and 32 can be extended to the one-sample MIS in general using Theorem 9.4 of Veach's thesis, which states that the variance of the one-sample MIS estimator is minimal when it is a balance heuristic one, i.e., $V[\hat{\mathcal{F}}] \leq V[\mathcal{F}]$. For instance, for Eq. 32

$$V[\hat{\mathcal{F}}_{\text{eq}}^1] \leq nV[\mathcal{F}^1] + (n-1)\mu^2. \quad (33)$$

And for N samples,

$$V[\hat{\mathcal{F}}_{\text{eq}}] \leq nV[\mathcal{F}] + \frac{(n-1)}{N}\mu^2. \quad (34)$$

Observe that Eq. 34 is similar to Eq. 23 except for the fact that it holds for the one-sample estimator. Using $V[\hat{\mathcal{F}}] \leq V[\mathcal{F}]$ (as shown in [11]) we also have:

$$V[\hat{\mathcal{F}}_{\text{eq}}] \leq nV[\mathcal{F}] + \frac{(n-1)}{N}\mu^2. \quad (35)$$

Finally, note that we do not state that Eq. 23 is invalid, but that its proof by Veach in his thesis [2] is wrong and also his interpretation is not correct. On the other hand, we have proved a formally identical relationship, Eq. 34, for the one-sample MIS estimator instead of the multi-sample estimator.

4 A provably better balance-heuristic multi-sample estimator

In Section 1, we reviewed two strategies for the allocation of samples to individual methods, balance heuristic with equal count of samples (Section 2.2.1) and non-balance heuristic with sample numbers proportional to the variance (Section 2.3). The variance of the latter is better than that of the equal count estimator. This section shows that we can do it even better if the sample allocation strategy of Section 2.3 is applied to balance heuristic.

We can use Theorem 9.2 from Veach's thesis, which states that for any MIS estimator F , the balance heuristic estimator \hat{F} with the same count of samples as F has always less variance up to an asymptotically decreasing factor, i.e.,

$$V[\hat{F}] - V[F] \leq \left(\frac{1}{\min_i N_i} - \frac{1}{\sum_{k=1}^n N_k} \right) \mu^2, \quad (36)$$

to obtain a provably better estimator than the one described in Section 2.3. Let us now take the estimator F as Eq. 19 from Section 2.3 where $N_i \propto \sigma_{i,\text{eq}}$. This is provably better than the balance heuristic with equal count of samples, \hat{F}_{eq} , i.e., $V[F] \leq V[\hat{F}_{\text{eq}}]$. Thus,

$$\begin{aligned} V[\hat{F}] &\leq V[F] + \left(\frac{1}{\min_i N_i} - \frac{1}{\sum_{k=1}^n N_k} \right) \mu^2 \\ &\leq V[\hat{F}_{\text{eq}}] + \left(\frac{1}{\min_i N_i} - \frac{1}{\sum_{k=1}^n N_k} \right) \mu^2. \end{aligned} \quad (37)$$

Observe that \hat{F} is the balance heuristic estimator defined in Eq. 12 where the $N_i \propto \sigma_{i,\text{eq}}$. Similar results can be obtained for the case when the sampling cost is also taken into account, i.e. consider now as F the MIS estimator with $N_i \propto \sigma_{i,\text{eq}}/\sqrt{c_i}$.

$$\begin{aligned} \left(\sum_{k=1}^n N_k c_k \right) V[\hat{F}] &\leq \left(\sum_{k=1}^n N_k c_k \right) V[F] + \\ &\quad \left(\sum_{k=1}^n N_k c_k \right) \left(\frac{1}{\min_i N_i} - \frac{1}{\sum_{k=1}^n N_k} \right) \mu^2 \\ &\leq \frac{N}{n} \left(\sum_{k=1}^n c_k \right) V[\hat{F}_{\text{eq}}] + \\ &\quad \left(\sum_{k=1}^n N_k c_k \right) \left(\frac{1}{\min_i N_i} - \frac{1}{\sum_{k=1}^n N_k} \right) \mu^2. \end{aligned} \quad (38)$$

Estimator \hat{F} is now the balance heuristic estimator defined in Eq. 12 with $N_i \propto \sigma_{i,\text{eq}}/\sqrt{c_i}$.

4.1 Discussion

Let us re-write Eq. 37 with variances $V_1[\hat{F}]$, $V_1[F]$, and $V_1[\hat{F}_{\text{eq}}]$ normalized to $N = 1$, i.e., $V[\hat{F}] = V_1[\hat{F}]/N$, $V[F] = V_1[F]/N$, $V[\hat{F}_{\text{eq}}] = V_1[\hat{F}_{\text{eq}}]/N$:

$$V_1[\hat{F}] \leq V_1[\hat{F}_{\text{eq}}] + \left(\frac{1}{\min_i \alpha_i} - 1 \right) \mu^2. \quad (39)$$

The efficiency in this case is

$$\begin{aligned} \left(\sum_{k=1}^n \alpha_k c_k \right) V_1[\hat{F}] &\leq \frac{1}{n} \left(\sum_{k=1}^n c_k \right) V_1[\hat{F}_{\text{eq}}] \\ &\quad + \left(\sum_{k=1}^n \alpha_k c_k \right) \left(\frac{1}{\min_i \alpha_i} - 1 \right) \mu^2. \end{aligned} \quad (40)$$

Thus, the relative difference between variances will be constant disregarding N . This is a similar situation to that encountered in Section 3.

5 A provably better one-sample balance heuristic estimator

The variance of the estimator in Eq. 25 is the second moment minus the square of μ :

$$\begin{aligned} V[\mathcal{F}^1] &= \int \sum_{i=1}^n \left(\frac{w_i(x)f(x)}{\alpha_i p_i(x)} \right)^2 \alpha_i p_i(x) dx - \mu^2 \\ &= \sum_{i=1}^n \int \frac{w_i^2(x)f^2(x)}{\alpha_i p_i(x)} dx - \mu^2 \\ &= \sum_{i=1}^n \frac{1}{\alpha_i} M_i^2 - \mu^2, \end{aligned} \quad (41)$$

where the second moments M_i^2 are defined as

$$M_i^2 = \int \frac{w_i^2(x)f^2(x)}{p_i(x)} dx. \quad (42)$$

Observe that, as μ^2 is a constant, minimizing $V[\mathcal{F}^1]$ means minimizing the first term of the last equality in Eq. 41. When the weights $\{w_i(x)\}$ are independent of the coefficients $\{\alpha_i\}$, we get the following theorem:

Theorem 2 *When the weights $\{w_i(x)\}$ are independent of the $\{\alpha_i\}$, the minimum variance in Eq. 41 is obtained when*

$$\alpha_i \propto M_i. \quad (43)$$

Proof The proof fully given in Appendix B follows from applying the Lagrange multipliers method to the target function:

$$\Lambda(\alpha_i, \lambda) = \sum_{i=1}^n \frac{1}{\alpha_i} M_i^2 - \mu^2 + \lambda \left(\sum_{k=1}^n \alpha_k - 1 \right). \quad \square$$

Let us use the weights of Section 2.3, i.e., $w_i(x) = p_i(x)/\sum_{k=1}^n p_k(x)$. These weights do not depend on the $\{\alpha_i\}$, and thus we can apply Theorem 2. The minimum

variance is then given by substituting the optimal $\{\alpha_i\}$ from Eq. 43 to Eq. 41, resulting in

$$V_{\min}[\mathcal{F}^1] = \left(\sum_{i=1}^n M_{i,\text{eq}} \right)^2 - \mu^2, \quad (44)$$

where the $M_{i,\text{eq}}$ values are the square root of

$$M_{i,\text{eq}}^2 = \int \frac{p_i(x)f^2(x)}{\left(\sum_{k=1}^n p_k(x)\right)^2} dx, \quad (45)$$

obtained by substituting the weights of Section 2.3 into Eq. 42. As they correspond to the minimum variance, the values obtained in Eq. 43 will provide a variance less than using any other count of samples, in particular when taking equal count of samples, thus we get

$$V_{\min}[\mathcal{F}^1] = \left(\sum_{i=1}^n M_{i,\text{eq}} \right)^2 - \mu^2 \leq n \sum_{i=1}^n M_{i,\text{eq}}^2 - \mu^2. \quad (46)$$

Observe that the right side in Eq. 46 also corresponds to the variance of the one-sample balance heuristic estimator with equal count of samples, i.e., for all i , $\alpha_i = 1/n$. Thus, we can guarantee that the one-sample estimator given by Eq. 25 using weights of Section 2.3 is better than the one-sample balance heuristic estimator with equal count of samples. Applying now Theorem 9.4 of Veach's thesis, that reads

$$V[\hat{\mathcal{F}}] \leq V[\mathcal{F}], \quad (47)$$

we have that the one-sample balance heuristic estimator

$$\hat{\mathcal{F}}^1 = \frac{(\sum_k M_{k,\text{eq}})f(x)}{\sum_i M_{i,\text{eq}}p_i(x)}, \quad (48)$$

using $\alpha_i \propto M_{i,\text{eq}}$ weights is guaranteed to have less variance than the one-sample balance heuristic estimator with equal count of samples. If we take into account the costs, we have the following theorem:

Theorem 3 *When the weights $\{w_i(x)\}$ are independent of the $\{\alpha_i\}$, the maximum efficiency of MIS one-sample estimator is obtained when*

$$\alpha_i = \frac{M_i \sqrt{C_T}}{\sqrt{c_i V[\mathcal{F}^1] + C_T \mu^2}}. \quad (49)$$

See Appendix B for a proof. Observe, however, that Eq. 49 is not an explicit expression for α_i since it contains these values in the right hand side, as $C_T = \sum_{k=1}^n \alpha_k c_k$. Thus, some approximation is needed in practice. One possibility is to take

$$\alpha_i \propto \frac{M_i}{\sqrt{c_i}}, \quad (50)$$

another is to substitute in the right side of Eq. 49 the optimal values independent of the costs, i.e., $\alpha_i \propto M_i$. Under certain circumstances, we can guarantee that the one-sample balance heuristic estimator using the values

$$\alpha_i \propto \frac{M_{i,\text{eq}}}{\sqrt{c_i}}, \quad (51)$$

is more efficient than the one-sample balance heuristic with equal count of sampling.

Theorem 4 *The one-sample balance heuristic estimator using the $\{\alpha_i\}$ values in Eq. 51 is more efficient than the one-sample balance heuristic estimator with equal count of sampling when the $\{\alpha_i\}$ values are decreasing with costs $\{c_i\}$.*

A proof can be found in Appendix B. The estimator would be now

$$\hat{\mathcal{F}}^1 = \frac{(\sum_{k=1}^n M_{k,\text{eq}}/c_k)f(x)}{\sum_{i=1}^n \frac{M_{i,\text{eq}}}{c_i} p_i(x)}. \quad (52)$$

When the variation of $M_{i,\text{eq}}$ values is small relatively to the variation of c_i values, then we can consider that $\alpha_i \propto M_{i,\text{eq}}/\sqrt{c_i}$ values are decreasing with c_i , and the efficiency of MIS estimator using $\alpha_i \propto M_{i,\text{eq}}/\sqrt{c_i}$ would be better than taking $\alpha_i = 1/n$.

5.1 Discussion

One might ask why we are interested in the one-sample estimator when its variance is always higher than its multi-sample version due to the additional randomization. In addition, as the costs C_T are the same, as we use on average the same number of samples from each technique for both the one-sample and the multi-sample estimators, we also have $C_T V[F] \leq C_T V[\mathcal{F}]$, this is, given a one-sample estimator, its corresponding multi-sample counterpart will always be more efficient. However, there might be cases where only the one-sample estimator is applicable, e.g., when we do not know a priori the number of samples or we need just a single sample.

6 A new heuristic

In [9, 10] the heuristic $\alpha_i \propto 1/v_i$ was presented, where $v_i = \int \frac{f^2(x)}{p_i(x)} dx - \mu^2$ is the variance of technique i . This heuristic is only valid when all techniques $p_i(x)$ are unbiased. We will present here a generalized heuristic that is valid even when the techniques are biased and based on the following theorems.

Theorem 5 *The variance $V[\hat{\mathcal{F}}^1]$ is bounded by*

$$V[\hat{\mathcal{F}}^1] \leq \sum_{k=1}^n \alpha_k m_k^2 - \mu^2, \quad (53)$$

where $m_k^2 = \int \frac{f^2(x)}{p_k(x)} dx$ is the second moment of technique k .

Proof Using the inequality between harmonic and arithmetic means, we get

$$\begin{aligned} V[\hat{\mathcal{F}}^1] &= \int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} dx - \mu^2 \quad (54) \\ &\leq \int \sum_{k=1}^n \frac{\alpha_k}{p_k(x)} f^2(x) dx - \mu^2 \\ &= \sum_{k=1}^n \alpha_k \int \frac{f^2(x)}{p_k(x)} dx - \mu^2. \end{aligned}$$

Observe that we do not require the $\{p_k\}$ techniques be unbiased. \square

If all techniques are unbiased, then

$$V[\hat{\mathcal{F}}^1] \leq \sum_{k=1}^n \alpha_k v_k, \quad (55)$$

which follows from that μ^2 can be written as $\sum_k \alpha_k \mu^2$. Note that from $V[\hat{F}] \leq V[\hat{\mathcal{F}}]$, the bound in Eqs. 53 and 55 also holds for $V[\hat{F}]$.

Observe also that if $m_{\max}^2 = \max_k \{m_k^2\}$ and $v_{\max} = \max_k \{v_k\}$, then from Eq. 53

$$V[\hat{\mathcal{F}}^1] \leq m_{\max}^2 - \mu^2,$$

and in case of unbiased techniques from Eq. 55

$$V[\hat{\mathcal{F}}^1] \leq v_{\max}.$$

That is, whichever combination $\{\alpha_k\}$ is taken, we cannot make it worse than just using the worst technique.

Let us use Theorem 5 to obtain a heuristic valid for biased techniques.

Theorem 6 Consider a function $f(x)$ and n not necessarily unbiased Monte Carlo integration techniques, $\{p_i(x)\}$, with second moments $\{m_i^2\}$, and the sequence of n positive values $\{\alpha_i\}$, $\sum_i \alpha_i = 1$. The rearrangement that minimizes the bound in Eq. 53 happens when the $\{\alpha_i\}$ are ordered inversely to $\{m_i^2\}$.

Proof As μ is a constant value only depending on $f(x)$, the rearrangement that minimizes the bound in Eq. 53 is the one that minimizes $\sum_{k=1}^n \alpha_k m_k^2$, and by virtue of the rearrangement inequality [25, page 261][26], it happens when the $\{\alpha_i\}$ are ordered inversely to $\{m_i^2\}$. \square

Theorem 6 justifies the introduction of the new heuristic $\alpha_i \propto 1/m_i^2$. We can extend this heuristic to take the

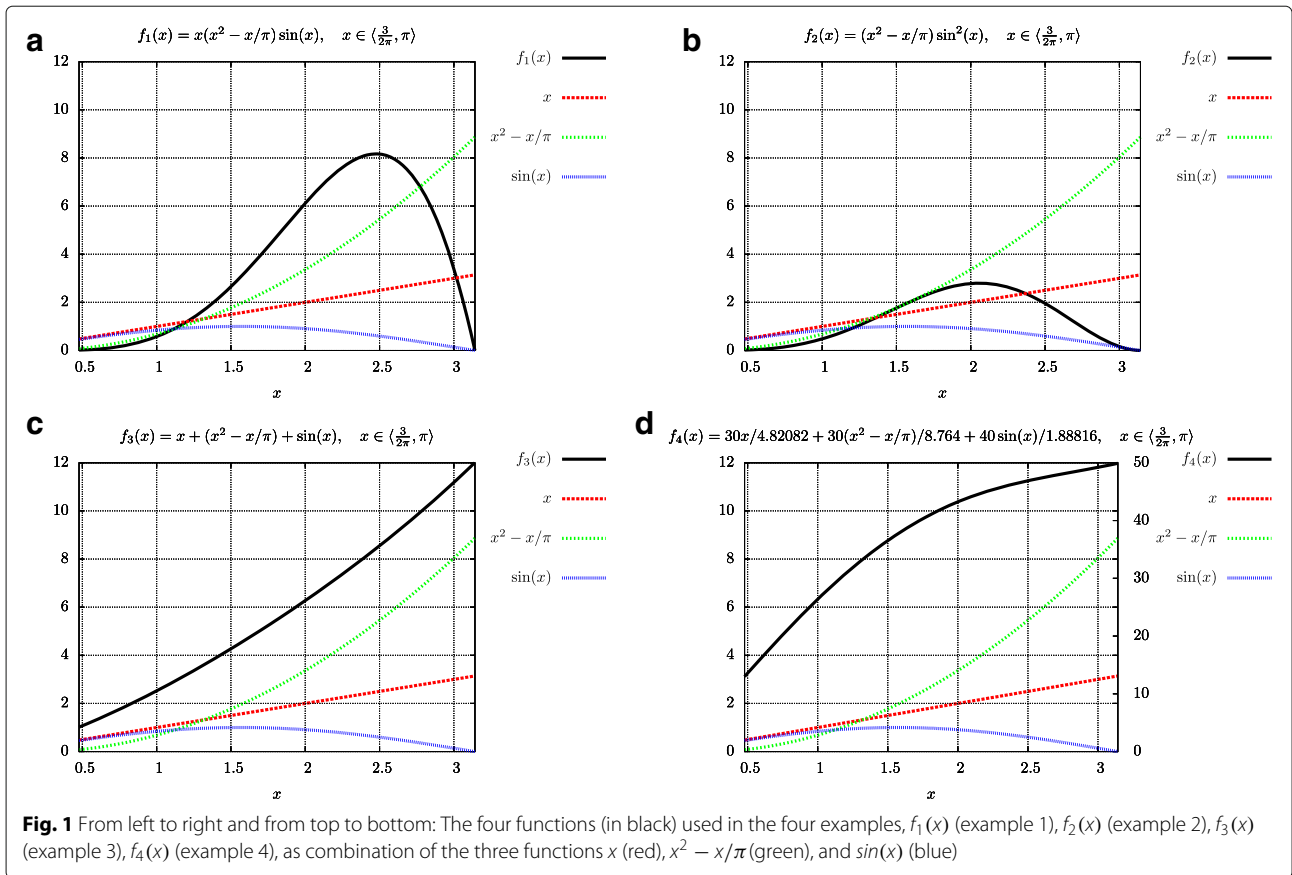
costs of the techniques into account, in the same way as in [9] and [10] by using $\alpha_i \propto 1/(c_i m_i^2)$.

7 Results

7.1 1D functions

We have compared the following estimators: equal count of samples with balance heuristic, count inversely proportional to variances of independent techniques with balance heuristic [9, 10], optimal count non-balance heuristic in [11], and the new balance heuristic estimators defined in Sections 4, 5, and 6 using values $\alpha_k \propto \sigma_{k,eq}$, $\alpha_k \propto M_{k,eq}$, and $\alpha_k \propto \frac{1}{m_k^2}$ for equal cost of sampling, and $\alpha_k \propto \frac{\sigma_{k,eq}}{\sqrt{c_k}}$, $\alpha_k \propto \frac{M_{k,eq}}{\sqrt{c_k}}$, and $\alpha_k \propto \frac{1}{c_k m_k^2}$ for unequal cost of sampling. The new estimators are tried for both the one-sample and the multi-sample case. The results for 1D example functions and pdfs of Fig. 1 are presented in Appendix C and summarized in Tables 2, 3, 4, 5, and 6. We additionally provide in supplemental material the Mathematica code to try any function with any pdfs. In the obtained results, we can see the following patterns:

- As expected [11] multi-sample estimators are always better than the corresponding one-sample estimators with the same distribution of samples
- As expected the multi-sample estimators defined in [11] (Section 2.2.1) are better than balance heuristic with equal count of samples, both in variance and efficiency.
- The new multi-sample estimators defined in Section 4 have similar variance values than balance heuristic with equal number of samples, but have better values for efficiency, both for the one-sample and the multi-sample estimator.
- The new multi-sample estimators defined in Section 4 provide similar variance as the ones in [11] but better efficiency, except in the fourth example where the efficiency is slightly worse.
- The optimal distribution of samples obtained in Section 5 performs always better for both the one-sample and the multi-sample estimator cases than the balance heuristic estimator with equal count of samples, in both variance and efficiency. This improvement is notable, especially when we consider that example 4 is particularly well suited for the equal count of sampling.
- The multi-sample estimators using the optimal distribution of samples obtained in Section 5 perform better than the multi-sample estimators defined in [11], in both variance and efficiency.
- The heuristically based estimators defined in [9] and in Section 6 give the best results for the first three examples, but are by far the worst for the fourth example.



- The heuristically based estimators for unbiased techniques defined in [9] and [10] give better results than the one defined in Section 6 for possibly biased techniques, except for the third example, where the new heuristic performs much better.
- We obtain in one case an acceleration of more than 30 times over the variance of the balance heuristic estimator with equal count of samples.

Table 2 Variances for multi-sample estimator for the four examples in Appendix C and the environment map example ($\times 10^{-3}$) for balance heuristic using equal count of samples, for the count inversely proportional to the variances of independent estimators [9] and [10], for the provably better, non-balance heuristic, estimator defined in [11], and for the three new estimators defined in this paper

	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. EM
$\alpha_k \propto \frac{1}{n}$	29.16	4.91	10.68	28.14	4.63
$\alpha_k \propto \frac{1}{v_k}$ [9, 10]	24.11	4.55	2.02	330.85	1.47
$\alpha_k \propto \sigma_{k,eq}$ [11]	29.09	4.90	10.61	27.94	4.61
$\alpha_k \propto \frac{1}{m_k^2}$	26.55	4.77	0.37	809.28	3.16
$\alpha_k \propto \sigma_{k,eq}$	29.90	4.99	9.48	6.61	4.36
$\alpha_k \propto M_{k,eq}$	28.24	4.83	7.05	18.04	3.87

7.2 Adaptive results for 1D functions

We present also results for the four 1D examples computed with adaptive sampling algorithms. A first batch using equal number of samples from each technique computes order 1 approximation of the different quantities, which in turn allows us to obtain a first approximation of the α values needed for the second batch. The samples of batch $\mathcal{N} + 1$ are obtained with the order \mathcal{N} approximation of the α values, and incrementally update the \mathcal{N} -order quantities. We provide in supplemental material the C++ code to compute all estimators adaptively. The pseudo-code for the estimators

Table 3 Variances for one-sample balance heuristic for the four examples in Appendix C and the environment map example ($\times 10^{-3}$) for balance heuristic using equal count of samples, for the count inversely proportional to the variances of independent estimators [9] and [10], and for the three new estimators defined in this paper

	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. EM
$\alpha_k \propto \frac{1}{n}$	30.16	5.02	13.35	35.16	5.14
$\alpha_k \propto \frac{1}{v_k}$ [9, 10]	24.22	4.60	2.05	332.97	1.50
$\alpha_k \propto \frac{1}{m_k^2}$	27.04	4.85	0.39	863.50	3.46
$\alpha_k \propto \sigma_{k,eq}$	31.03	5.10	11.67	57.38	4.84
$\alpha_k \propto M_{k,eq}$	29.07	4.92	8.72	22.59	4.28

Table 4 Variance times cost (i.e., inefficiency, the smaller the better) for multi-sample estimator for the four examples in Appendix C and the environment map example ($\times 10^{-3}$) for balance heuristic using equal count of samples, for the count inversely proportional to the variances of independent estimators [9] and [10], for the provably better, non-balance heuristic, estimator defined in [11], and for the three new estimators defined in this paper

	Example 1	Example 2	Example 3	Example 4	Example EM
$\alpha_k \propto \frac{1}{n}$	102.26	17.24	37.47	98.68	13.41
$\alpha_k \propto \frac{1}{c_k v_k}$ [9, 10]	40.41	9.28	4.03	300.12	1.31
$\alpha_k \propto \frac{\sigma_{k,eq}}{\sqrt{c_k}}$ [11]	89.40	15.44	31.80	83.78	11.31
$\alpha_k \propto \frac{1}{c_k m_k^2}$	43.06	9.82	3.12	534.37	1.68
$\alpha_k \propto \frac{\sigma_{k,eq}}{\sqrt{c_k}}$	81.43	13.54	28.68	91.01	5.88
$\alpha_k \propto \frac{M_{k,eq}}{\sqrt{c_k}}$	79.73	13.08	25.74	31.77	4.92

defined in [9] and [10] can be found in the supplementary material of [10]. The charts with variances normalized to 1 sample, (i.e., variance times the number of samples), and variances times cost (i.e., inefficiencies) for specific costs as described in the Appendix C, and averaged for 100 runs are shown in Figs. 2 and 3 for samples count from 35 to 5000.

7.3 2D analytical environment map example with Lafortune-Phong BRDF model

Let us consider here the BRDF formula for the physically based variant of the Phong model presented by Lafortune [27], assuming the case when the outgoing direction ω_o is the surface normal:

$$f_r(\omega_o, \omega_i) = \frac{\rho_d}{\pi} + \rho_s \frac{m+2}{2\pi} \cos^m \theta \quad (56)$$

where ρ_d is the diffuse albedo, ρ_s is the specular albedo, m is the shininess, and θ is the angle between incident direction ω_i and the surface normal.

Let us consider also an environment map of intensity $R(\omega_i) = R \cdot \cos \theta$. As $d\omega = \sin \theta d\theta d\phi$ the outgoing radiance is given then by the integral for the hemisphere in spherical parametrization θ, ϕ of the incident direction

$$R \iint_{\theta, \phi} \left(\frac{\rho_d}{\pi} + \rho_s \frac{m+2}{2\pi} \cos^m \theta \right) \cos \theta \cos \theta \sin \theta d\theta d\phi \quad (57)$$

and integrating over ϕ we get

$$2\pi R \int_{\theta=0}^{\pi/2} \left(\frac{\rho_d}{\pi} + \rho_s \frac{m+2}{2\pi} \cos^m \theta \right) \cos \theta \cos \theta \sin \theta d\theta. \quad (58)$$

By the variable substitution of $x = \cos \theta$, we obtain

$$2\pi R \int_{x=0}^1 \left(\left(\frac{\rho_d}{\pi} + \rho_s \frac{m+2}{2\pi} x^m \right) x \right) x dx. \quad (59)$$

We have thus the integral of the product $p_1(x)p_2(x)$, where $p_1(x)$ corresponds to the BRDF times $x = \cos \theta$,

$$p_1(x) = \left(\frac{\rho_d}{\pi} + \rho_s \frac{m+2}{2\pi} x^m \right) x, \quad (60)$$

and $p_2(x) = x$ is the environment map.

We used typical values, $m = 5$, $\rho_d = \rho_s = 0.5$, and compared for equal cost and for the empirically obtained sampling costs, $c_1 = 1$, $c_2 = 4.8$. The detailed results are in the last columns of Tables 2, 3, 4, 5, and in the last two columns of Table 6. The results confirm the patterns found in our 1D examples.

8 Conclusions

This paper discussed an important topic of MIS estimators and the limited acceleration over equal count of

Table 5 Variance times cost (i.e., inefficiency, the smaller the better) for one-sample estimator for the four examples in Appendix C and the environment map example ($\times 10^{-3}$) for balance heuristic using equal count of samples, for the count inversely proportional to the variances of independent estimators [9] and [10], and for the three new estimators defined in this paper

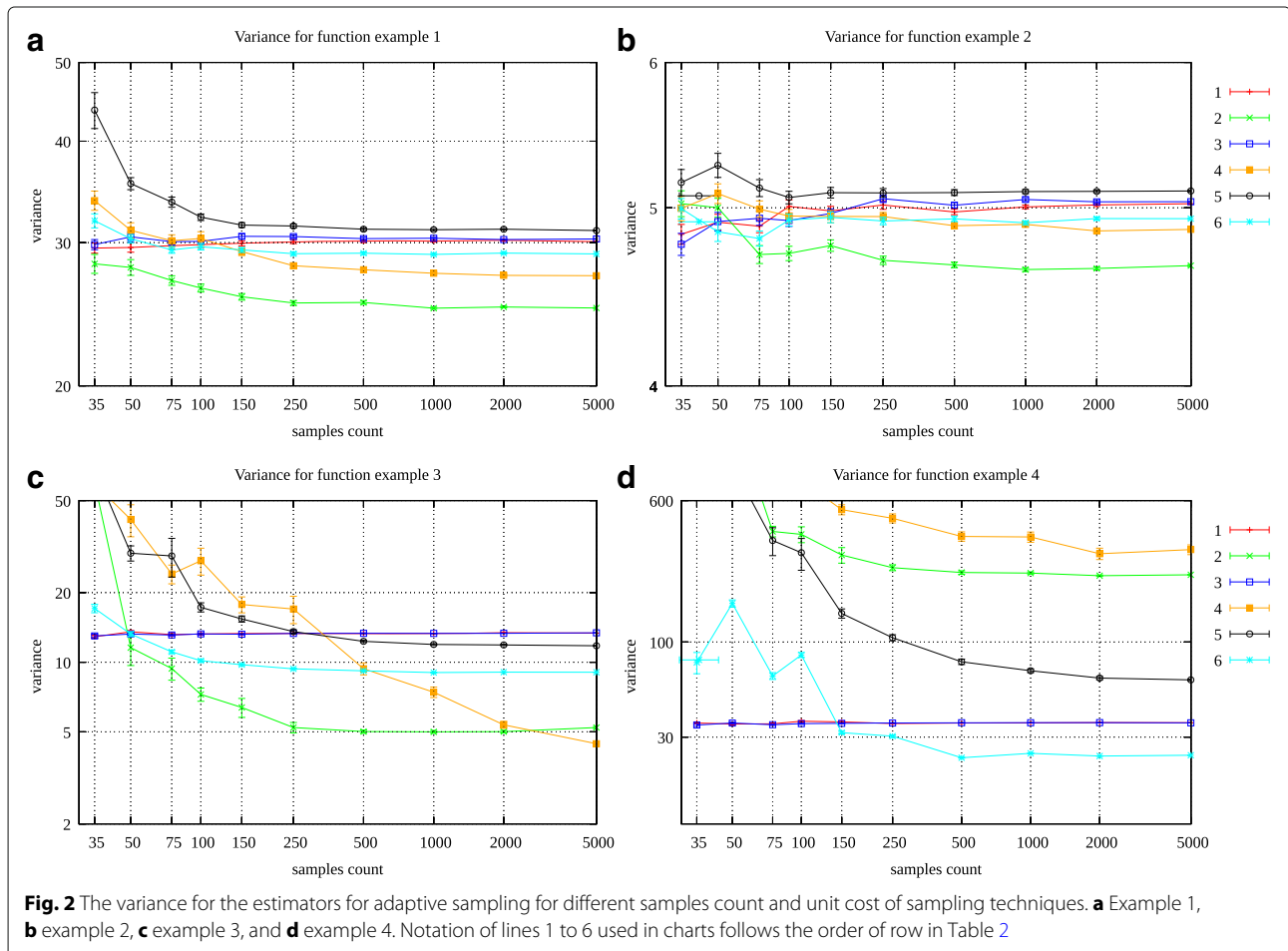
	Example 1	Example 2	Example 3	Example 4	Example EM
$\alpha_k \propto \frac{1}{n}$	105.78	17.60	46.82	123.30	14.90
$\alpha_k \propto \frac{1}{c_k v_k}$ [9, 10]	40.89	9.37	4.04	300.48	1.32
$\alpha_k \propto \frac{1}{c_k m_k^2}$	44.43	9.93	3.18	545.53	2.30
$\alpha_k \propto \frac{\sigma_{k,eq}}{\sqrt{c_k}}$	83.82	13.27	33.36	105.80	6.39
$\alpha_k \propto \frac{M_{k,eq}}{\sqrt{c_k}}$	81.66	13.08	30.11	37.57	5.30

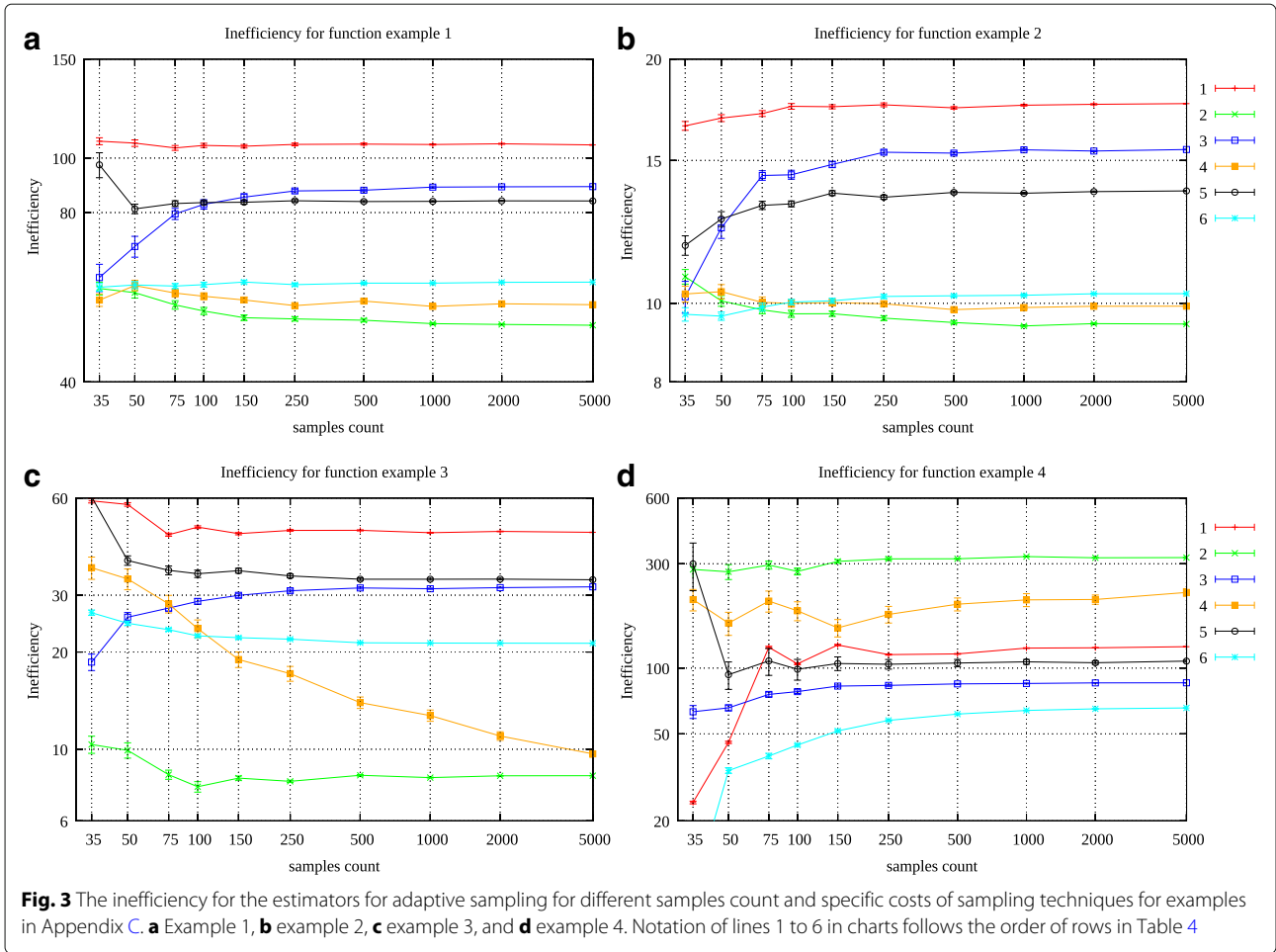
Table 6 α_j^i : α_i values for the j -th 1D example and for the environment map example

	Example 1			Example 2			Example 3			Example 4			Example EM	
	α_1^1	α_2^1	α_3^1	α_1^2	α_2^2	α_3^2	α_1^3	α_2^3	α_3^3	α_1^4	α_2^4	α_3^4	α_1	α_2
$\alpha = \frac{1}{n}$	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.5	0.5
$\alpha_k \propto \frac{1}{v_k}$ [9, 10]	0.42	0.48	0.10	0.35	0.21	0.44	0.90	0.10	0.00	0.95	0.04	0.00	0.93	0.07
$\alpha_k \propto \frac{1}{c_k v_k}$ [9, 10]	0.79	0.14	0.06	0.68	0.07	0.25	0.98	0.02	0.00	0.99	0.01	0.00	0.99	0.01
$\alpha_k \propto \sigma_{k,eq}$ [11]	0.33	0.31	0.35	0.34	0.34	0.31	0.37	0.32	0.30	0.37	0.32	0.30	0.52	0.48
$\alpha_k \propto \frac{\sigma_{k,eq}}{\sqrt{c_k}}$ [11]	0.51	0.19	0.29	0.52	0.21	0.26	0.55	0.19	0.25	0.55	0.19	0.25	0.70	0.30
$\alpha_k \propto \frac{1}{m_k^2}$	0.38	0.39	0.23	0.35	0.29	0.36	0.53	0.46	0.01	0.60	0.36	0.04	0.66	0.34
$\alpha_k \propto \frac{1}{c_k m_k^2}$	0.52	0.16	0.32	0.69	0.09	0.22	0.87	0.12	0.00	0.56	0.19	0.25	0.90	0.10
$\alpha_k \propto M_{k,eq}$	0.33	0.35	0.30	0.32	0.31	0.35	0.34	0.36	0.28	0.33	0.32	0.34	0.57	0.43
$\alpha_k \propto \frac{M_{k,eq}}{\sqrt{c_k}}$	0.52	0.21	0.25	0.50	0.19	0.30	0.52	0.22	0.24	0.50	0.19	0.29	0.75	0.25

samples balance heuristic. We have shown that in reality, this bound on acceleration does not hold, which justifies the search for better and more efficient estimators. We have obtained, both without and with taking into account the cost of sampling, new balance heuristic estimators

that are provably better than balance heuristic with equal count of samples, and new heuristic estimators valid even when the independent techniques are biased. We have analyzed their behavior with 1D examples and with a 2D environment map example.





Appendix A: Inequality in Theorem 9.5 of Veach’s thesis does not hold for balance heuristic

Let us repeat first the proof of Eq. 24 such as it appears in Veach’s thesis

$$\begin{aligned}
 V[F] &= \sum_{i=1}^n \frac{1}{N_i} \left(\int \frac{w_i^2(x) f^2(x)}{p_i(x)} dx - \mu_i^2 \right) \quad (61) \\
 &\geq \sum_{i=1}^n \frac{1}{N} \left(\int \frac{w_i^2(x) f^2(x)}{p_i(x)} dx - \mu_i^2 \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \frac{1}{N/n} \left(\int \frac{w_i^2(x) f^2(x)}{p_i(x)} dx - \mu_i^2 \right) \\
 &= \frac{1}{n} V[F_{eq}],
 \end{aligned}$$

where $V[F_{eq}]$ is computed assuming the same weights as $V[F]$ but with equal count of samples. Let us see that inequality 61 does not hold when the weights $w_i(x)$ depend on the N_i . We will see it for the balance heuristic estimator, where the weights are defined by

$$w_i(x) = \frac{N_i p_i(x)}{\sum_{k=1}^n N_k p_k(x)}. \quad (62)$$

Eq. 61 would be

$$\begin{aligned}
 V[\hat{F}] &= \sum_{i=1}^n \frac{1}{N_i} \left(\int \frac{N_i^2 f^2(x) p_i(x)}{(\sum_{k=1}^n N_k p_k(x))^2} dx - \mu_i^2 \right) \\
 &\geq \sum_{i=1}^n \frac{1}{N} \left(\int \frac{N_i^2 f^2(x) p_i(x)}{(\sum_{k=1}^n N_k p_k(x))^2} dx - \mu_i^2 \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \frac{1}{N/n} \left(\int \frac{N_i^2 f^2(x) p_i(x)}{(\sum_{k=1}^n N_k p_k(x))^2} dx - \mu_i^2 \right) \\
 &\neq \frac{1}{N} \sum_{i=1}^n \left(\int \frac{f^2(x) p_i(x)}{(\sum_{k=1}^n p_k(x))^2} dx - \mu_i^2 \right) \\
 &= \frac{1}{n} V[\hat{F}_{eq}], \quad (63)
 \end{aligned}$$

as the last but two expression is clearly not the variance of a balance heuristic estimator with equal count of samples (Eq. 16). Inequality 61 thus holds only when the weights $w_i(x)$ do not depend on counts N_i , and thus the proof of Theorem 9.5 is not valid in this case.

Appendix B: Section 5 proofs

We present the proof of Theorem 2. This proof is similar to the proof of Theorem 2 in [11].

Proof To optimize the variance $V[\hat{\mathcal{F}}^1] = \sum_{i=1}^n \frac{1}{\alpha_i} M_i^2 - \mu^2$, we use Lagrange multipliers with the constraint $\sum_{i=1}^n \alpha_i = 1$ and objective function

$$\Lambda(\alpha_i, \lambda) = \sum_{i=1}^n \frac{1}{\alpha_i} M_i^2 - \mu^2 + \lambda \left(\sum_{i=1}^n \alpha_i - 1 \right).$$

Computing the partial derivatives and equating them to zero, as the M_i do not depend on the α_i , we get

$$\begin{aligned} \frac{\partial \Lambda(\alpha_i, \lambda)}{\partial \alpha_j} &= \frac{\partial \left(\sum_{i=1}^n \frac{1}{\alpha_i} M_i^2 - \mu^2 \right)}{\partial \alpha_j} \\ &+ \frac{\partial \left(\lambda \left(\sum_{i=1}^n \alpha_i - 1 \right) \right)}{\partial \alpha_j} \\ &= -\frac{M_j^2}{\alpha_j^2} + \lambda = 0. \end{aligned} \quad (64)$$

Thus for all j values $\lambda = M_j^2/\alpha_j^2$, which implies that $\alpha_j \propto M_j$. The Hessian matrix, obtained with the second derivatives of $V[\hat{\mathcal{F}}^1]$, is a diagonal matrix with positive diagonal values

$$\frac{\partial^2 V[\hat{\mathcal{F}}^1]}{\partial \alpha_j \partial \alpha_j} = 2 \frac{M_j^2}{\alpha_j^3} \quad (65)$$

and thus is positive-definite. The variance function is then strictly convex in its convex domain $\sum_{i=1}^n \alpha_i = 1$, where for all i , $0 < \alpha_i < 1$, meaning that the critical point is unique and a minimum.

Substituting the optimal values, i.e., $\alpha_j \propto M_j$, we find the minimum variance,

$$V_{\min}[\hat{\mathcal{F}}^1] = \left(\sum_{i=1}^n M_i \right)^2 - \mu^2. \quad (66)$$

□

We present now the proof of Eq. 49.

Proof Let us consider c_i the cost of sampling with technique i . The average cost (or cost for one total sample) is thus $C_T = \sum_i \alpha_i c_i$. We want to minimize now the cost times variance, i.e., $C_T \times V[\mathcal{F}^1]$, which is the inverse

of efficiency. We use Lagrange multipliers with the constraint $\sum_{i=1}^n \alpha_i = 1$ and objective function

$$\begin{aligned} \Lambda(\alpha_i, \lambda) &= C_T \times V[\mathcal{F}^1] + \lambda \left(\sum_{i=1}^n \alpha_i - 1 \right) \\ &= \left(\sum_i \alpha_i c_i \right) \left(\sum_{i=1}^n \frac{1}{\alpha_i} M_i^2 - \mu^2 \right) \\ &+ \lambda \left(\sum_{i=1}^n \alpha_i - 1 \right). \end{aligned} \quad (67)$$

Taking partial derivatives and equating them to zero, as the M_i do not depend on the α_i ,

$$\begin{aligned} \frac{\partial \Lambda(\alpha_i, \lambda)}{\partial \alpha_j} &= \frac{\partial \left(\sum_{i=1}^n \alpha_i c_i \right)}{\partial \alpha_j} V[\mathcal{F}^1] \\ &+ C_T \frac{\partial \left(\sum_{i=1}^n \frac{1}{\alpha_i} M_i^2 \right)}{\partial \alpha_j} \\ &+ \frac{\partial \left(\lambda \left(\sum_{i=1}^n \alpha_i - 1 \right) \right)}{\partial \alpha_j} \\ &= c_j V[\mathcal{F}^1] - C_T \frac{M_j^2}{\alpha_j^2} + \lambda = 0. \end{aligned} \quad (68)$$

Multiplying by α_j and adding over j , we obtain

$$\begin{aligned} \sum_{j=1}^n \alpha_j c_j V[\mathcal{F}^1] - \sum_{j=1}^n \alpha_j C_T \frac{M_j^2}{\alpha_j^2} + \sum_{j=1}^n \alpha_j \lambda \\ = V[\mathcal{F}^1] C_T - C_T (V[\mathcal{F}^1] + \mu^2) + \lambda \\ = -C_T \mu^2 + \lambda \\ = 0 \end{aligned} \quad (69)$$

and thus $\lambda = C_T \mu^2$. Substituting this value in Eq. 68 we find that the optimal sampling counts are the ones in Eq. 49. □

We present here the proof of Theorem 4.

Proof We compare first the efficiency of the general one-sample MIS estimator with weights given by Eq. 14 and $\alpha_i \propto M_{i,\text{eq}}/\sqrt{c_i}$, with the balance heuristic estimator with equal count of samples. The inverse of efficiency, i.e., cost times variance, is given in both cases by Eq. 70:

$$\begin{aligned} \left(\sum_{i=1}^n \alpha_i c_i \right) \left(\sum_{i=1}^n \frac{M_{i,\text{eq}}^2}{\alpha_i} - \mu^2 \right) \\ = \left(\sum_{i=1}^n \alpha_i c_i \right) \left(\sum_{i=1}^n \frac{M_{i,\text{eq}}^2}{\alpha_i} \right) - \left(\sum_{i=1}^n \alpha_i c_i \right) \mu^2 \end{aligned} \quad (70)$$

Let us compare first the terms corresponding to the second moments, i.e., the term $\left(\sum_{i=1}^n \alpha_i c_i \right) \left(\sum_{i=1}^n \frac{M_{i,\text{eq}}^2}{\alpha_i} \right)$, for $\alpha_i = 1/n$ and $\alpha_i \propto M_{i,\text{eq}}/\sqrt{c_i}$. Using Cauchy-Schwartz

inequality, we can easily check that Eq. 71 is true, where the first term is the second moment for $\alpha_i \propto M_{i,eq}/\sqrt{c_i}$ and the second term for $\alpha_i = 1/n$, because it holds that

$$\left(\sum_{i=1}^n \sqrt{c_i} M_{i,eq}\right)^2 \leq \left(\sum_{i=1}^n c_i\right) \left(\sum_{i=1}^n M_{i,eq}^2\right). \quad (71)$$

Let us consider now the term of Eq. 70 that contains μ^2 , i.e., $(\sum_{i=1}^n \alpha_i c_i) \mu^2$. Whenever the α_i values are decreasing with c_i , the following inequality can be proved [26, 28].

$$\frac{1}{n} \sum_{i=1}^n c_i \geq \sum_{i=1}^n \alpha_i c_i. \quad (72)$$

Thus, we can guarantee that the general MIS estimator with weights $w_i(x)$ given by Eq. 14 and $\alpha_i \propto M_{i,eq}/\sqrt{c_i}$ is more efficient than the one-sample balance heuristic estimator with equal count of samples whenever the α_i values are decreasing with c_i . Considering now the one-sample balance heuristic estimator with $\alpha_i \propto M_{i,eq}/\sqrt{c_i}$, as the costs are the same and the variance is less than the general MIS estimator considered by virtue of Theorem 9.4 in Veach's thesis, its efficiency is higher, and in turn, higher than the balance heuristic estimator with equal count of sampling. \square

Appendix C: 1D examples

We compare the results for the functions and pdfs in Fig. 1 with the following estimators: equal count of samples, count inversely proportional to variances of independent techniques [9], the new heuristic defined in Section 6 of this paper, optimal count in [11], and the two new balance heuristic provably better estimators defined in Sections 4 and 5. These two new estimators are tried for both one-sample and multi-sample.

Example 1

Suppose we want to solve the integral

$$\mu = I = \int_{\frac{3}{2\pi}}^{\pi} x \left(x^2 - \frac{x}{\pi}\right) \sin(x) dx = 10.29 \quad (73)$$

by MIS sampling on functions x , $(x^2 - \frac{x}{\pi})$, and $\sin(x)$, respectively. We first find the normalization constants: $\int_{\frac{3}{2\pi}}^{\pi} x dx = 4.82$, $\int_{\frac{3}{2\pi}}^{\pi} (x^2 - \frac{x}{\pi}) dx = 8.76$, $\int_{\frac{3}{2\pi}}^{\pi} \sin(x) dx = 1.89$. We take into account the sampling costs given in [9], i.e., $c_1 = 1$, $c_2 = 6.24$, $c_3 = 3.28$.

Example 2

Let us solve the integral

$$\mu = I = \int_{\frac{3}{2\pi}}^{\pi} \left(x^2 - \frac{x}{\pi}\right) \sin^2(x) dx = 3.60 \quad (74)$$

using the same functions x , $(x^2 - \frac{x}{\pi})$, and $\sin(x)$ as before.

Example 3

As the third example, let us solve the integral

$$\mu = I = \int_{\frac{3}{2\pi}}^{\pi} x + \left(x^2 - \frac{x}{\pi}\right) + \sin(x) dx = 15.47 \quad (75)$$

using the same functions as before.

Example 4

As the last example, consider the integral of the sum of the three pdfs

$$\begin{aligned} \mu &= I \\ &= \int_{\frac{3}{2\pi}}^{\pi} 30 \frac{x}{4.82} + 30 \frac{(x^2 - \frac{x}{\pi})}{8.76} + 40 \frac{\sin(x)}{1.89} dx \\ &= 100. \end{aligned} \quad (76)$$

In this case, we know the optimal (zero variance) α values: (0.3, 0.3, 0.4). This case should be most favorable to equal count of samples.

Results are summarized in Tables 2, 3, 4, 5, and 6.

Acknowledgements

The authors acknowledge the comments by anonymous reviewers that helped to improve a preliminary version of the paper.

Funding

The authors are funded in part by Czech Science Foundation research program GA14-19213S by grant TIN2016-75866-C3-3-R from the Spanish Government and by grant OTKA K-124124 and VKSZ-14 PET/MRI 7T.

Authors' contributions

Equal contribution from all authors. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹School of Computer Science and Technology, Tianjin University, Tianjin, China. ²Department of Informatics, Applied Mathematics and Statistics, University of Girona, Campus de Montilivi, Girona, Spain. ³Czech Technical University at Prague, Faculty of Electrical Engineering, Prague, Czech Republic. ⁴Budapest Technical University, Budapest, Hungary.

Received: 11 July 2017 Accepted: 23 January 2018

Published online: 27 February 2018

References

1. E Veach, LJ Guibas, in *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '95*. Optimally Combining Sampling Techniques for Monte Carlo Rendering (ACM, New York, 1995), pp. 419–428. <https://doi.org/10.1145/218380.218498>. <http://doi.acm.org/10.1145/218380.218498>
2. E Veach, Robust Monte Carlo Methods for light transport simulation. PhD thesis, Stanford University (1997)
3. V Elvira, L Martino, D Luengo, MF Bugallo, Efficient multiple importance sampling estimators. *IEEE Signal Process. Lett.* **22**(10), 1757–1761 (2015). <https://doi.org/10.1109/LSP.2015.2432078>
4. V Elvira, L Martino, D Luengo, MF Bugallo, Generalized multiple importance sampling. ArXiv e-prints (2015). <http://arxiv.org/abs/1511.03095>

5. K Subr, D Nowrouzezahrai, W Jarosz, J Kautz, K Mitchell, Error analysis of estimators that use combinations of stochastic sampling strategies for direct illumination. *Comput. Graph. Forum Proc. EGSR*. **33**(4), 93–102 (2014). <https://doi.org/10.1111/cgf.12416>
6. M Pharr, G Humphreys, *Physically based rendering, Second Edition: From Theory To Implementation*, 2nd edn. (Morgan Kaufmann Publishers Inc., San Francisco, 2010)
7. H Lu, R Pacanowski, X Granier, Second-order approximation for variance reduction in multiple importance sampling. *Comput. Graph. Forum*. **32**(7), 131–136 (2013). <https://doi.org/10.1111/cgf.12220>
8. A Pajot, L Barthe, M Paulin, P Poulin, Representativity for robust and adaptive multiple importance sampling. *IEEE Trans. Vis. Comput. Graph.* **17**(8), 1108–1121 (2011). <https://doi.org/10.1109/TVCG.2010.230>
9. V Havran, M Sbert, in *Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, VRCAI '14. Optimal Combination of Techniques in Multiple Importance Sampling* (ACM, New York, 2014), pp. 141–150. <https://doi.org/10.1145/2670473.2670496>
10. M Sbert, V Havran, Adaptive multiple importance sampling for general functions. *Vis. Comput.* 1–11 (2017). <https://doi.org/10.1007/s00371-017-1398-1>
11. M Sbert, V Havran, L Szirmay-Kalos, Variance analysis of multi-sample and one-sample multiple importance sampling. *Comput. Graph. Forum*. **35**(7), 451–460 (2016). <https://doi.org/10.1111/cgf.13042>
12. GO Roberts, A Gelman, WR Gilks, Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7**, 110–120 (1997)
13. C Andrieu, J Thoms, A tutorial on adaptive MCMC. *Stat. Comput.* **18**(4), 343–373 (2008)
14. T Hachisuka, HW Jensen, Robust adaptive photon tracing using photon path visibility. *ACM Trans. Graph.* **30**(5), 114–111411 (2011). <https://doi.org/10.1145/2019627.2019633>
15. L Szirmay-Kalos, L Szécsi, Improved stratification for metropolis light transport. *Comput. Graph.* **68**(Supplement C), 11–20 (2017). <https://doi.org/10.1016/j.cag.2017.07.032>
16. T Hachisuka, AS Kaplanyan, C Dachsbacher, Multiplexed metropolis light transport. *ACM Trans. Graph.* **33**(4), 100–110010 (2014). <https://doi.org/10.1145/2601097.2601138>
17. A Owen, Y Zhou, Safe and Effective Importance Sampling. *J. Am. Stat. Assoc.* **95**, 135–143 (2000)
18. R Douc, A Guillin, J-M Marin, CP Robert, Convergence of Adaptive Mixtures of Importance Sampling Schemes. *Ann. Stat.* **35**(1), 420–448 (2007). <https://doi.org/10.1214/009053606000001154>
19. R Douc, A Guillin, JM Marin, CP Robert, Minimum variance importance sampling via population monte carlo. *ESAIM: Probab. Stat.* **11**, 424–447 (2007)
20. J-M Cornuet, J-M Marin, A Mira, CP Robert, Adaptive multiple importance sampling. *Scand. J. Stat.* **39**(4), 798–812 (2012). <https://doi.org/10.1111/j.1467-9469.2011.00756.x>
21. J-M Marin, P Pudlo, M Sedki, Consistency of the adaptive multiple importance sampling. Preprint arXiv:1211.2548 (2012). <http://arxiv.org/abs/1211.2548>
22. L Martino, V Elvira, D Luengo, J Corander, An adaptive population importance sampler: Learning from uncertainty. *IEEE Trans. Signal Process.* **63**(16), 4422–4437 (2015). <https://doi.org/10.1109/TSP.2015.2440215>
23. V Elvira, L Martino, D Luengo, J Corander, in *Proceedings of ICASSP 2015. A gradient adaptive population importance sampler* (IEEE, Brisbane, 2015), pp. 4075–4079
24. V Elvira, L Martino, D Luengo, MF Bugallo, Efficient multiple importance sampling estimators. *IEEE Signal Process. Lett.* **22**(10), 1757–1761 (2015). <https://doi.org/10.1109/LSP.2015.2432078>
25. GH Hardy, JE Littlewood, G Pólya, *Inequalities*. (Cambridge University Press, Cambridge, 1952)
26. M Sbert, J Poch, A necessary and sufficient condition for the inequality of generalized weighted means. *J. Inequalities Appl.* **2016**(1), 292 (2016). <https://doi.org/10.1186/s13660-016-1233-7>
27. EP Lafortune, YD Willems, Using the modified Phong reflectance model for physically based rendering. Technical Report report CW197, Dept. of Computer Science, K.U.Leuven (1994)
28. MM Marjanović, Z Kadelburg, A proof of chebyshev's inequality. *Teach. Math.* **X**(2), 107–108 (2007)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com