DOI: 10.1111/cgf.13042

Pacific Graphics 2016
E. Grinspun, B. Bickel, and Y. Dobashi
(Guest Editors)

*Volume 35 (2016), Number 7*

# Variance Analysis of Multi-sample and One-sample Multiple Importance Sampling

M. Sbert[1,2] & V. Havran[3] & L. Szirmay-Kalos[4]

[1] Tianjin University, China
[2] University of Girona, Spain
[3] Faculty of Electrical Engineering, Czech Technical University, Czech Republic
[4] Budapest University of Technology and Economics, Hungary

**Abstract**

*We reexamine in this paper the variance for the Multiple Importance Sampling (MIS) estimator for multi-sample and one-sample model. As a result of our analysis we can obtain the optimal estimator for the multi-sample model for the case where the weights do not depend on the count of samples. We extend the analysis to include the cost of sampling. With these results in hand we find a better estimator than balance heuristic with equal count of samples. Further, we show that the variance for the one-sample model is larger or equal than for the multi-sample model, and that there are only two cases where the variance is the same. Finally, we study on four examples the difference of variances for equal count as used by Veach, our new estimator, and a recently introduced heuristic.*

Categories and Subject Descriptors (according to ACM CCS): G.3 [Computer Graphics]: Mathematics of Computing / PROBABILITY AND STATISTICS—Probabilistic algorithms

**Keywords:** global illumination, rendering equation analysis, multiple importance sampling, Monte Carlo

## 1. Introduction

Many computer graphics algorithms need efficient sampling schemes that lead to low variance integral estimators. The *Multiple Importance Sampling* (MIS) estimator [VG95, Vea97], 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, as it is a reliable and robust estimator that allows an easy and straightforward combination of different sampling techniques.

MIS can be applied with pre-determining the number of samples generated by the different methods (multi-sample model). Alternatively, the decision on the actually used sample method can also be made randomly (one-sample model). This random decision is essential when very few samples need to be generated by multiple methods, e.g. a single sample.

Despite the wide scale application of MIS, there are still open questions about their benefit and optimal use. Among others, for an efficient application, we need to know how many samples are worth taking with different techniques and where the trade-off is between random decisions and deterministic, but non-optimal selections. The determination of the relative number of samples allocated to each of the sampling methods or their selection probabilities is non-trivial, it should take into account the variance and sampling cost of individual methods and their contribution to the overall variance and cost of the MIS approach. It is intuitive that strategies of high variance and/or significant cost should be given a lower weight or sample number, but this requirement is difficult to formalize and meet. On the one hand, this is a typical chicken and egg problem since we wish to optimize a sampling strategy before applying it and gathering information about its properties in a particular neighborhood of our scene. Heuristics can be 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 [PBPP11].

In other papers addressing the variance of MIS, strategies were assumed to have equal number of samples [EMLB15] and the combination of MIS with jittered sampling was studied [SNJ*14]. Lu et al. [LPG13] used a Taylor's second order approximation of the variance around the equal weights 1/2 to obtain the counts of samples from BRDF and environment map, which is accurate only if the optimal sample numbers are not too far from equal sampling. Havran and Sbert in [HS14] used a heuristic for the count of samples based on the inverse of the variance of each technique, and taking into account the cost of sampling.

This paper revisits the variance of MIS estimators and addresses the open problems of optimal sample allocation. In addition to variance and cost analysis, we also propose a new estimator in Section 4, which is provably better than the equal count of samples estimator. This new estimator belongs to the category of multi-sample estimators and thus can be used when we can control the number of samples allocated to different methods.

The rest of the paper is organized as follows. In the next section we review the variance of the MIS estimators where the weights are independent of the samples taken and address the problem of optimal number of samples allocated to the individual methods. In Section 3, a similar analysis is carried out for the balance heuristic, where the weights also involve the sample numbers. Section 4 presents our new estimator that belongs to the category where weights are independent of the number of samples. We study the effect of further randomization used in the one-sample estimator in Section 5. Numeric examples are presented in the appendix.

## 2. Variance of MIS estimator when weights are independent of the number of samples

The MIS estimator introduced by Veach and Guibas [VG95] to estimate the value of integral $I = \int f(x) d\mu(x)$ has the following expression:

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

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 $X_{ij}$, $(j = 1, \ldots, n_i)$. 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 $\mu$ is equal to integral $I$:

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

The variance of the estimator is given in the proof of Theorem 9.2 of Veach's thesis [Vea97]. 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 $j$, the estimators $F_{ij}$ are independent identically distributed random variables with expected value $\mu_i$:

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

Observe that

$$\mu = \sum_{i=1}^{n} \mu_i, \tag{7}$$

and that the variance of $F_{ij}$ is

$$\begin{aligned} \sigma_i^2 &= E[F_{ij}^2] - E^2[F_{ij}] \qquad (8) \\ &= \int \left( \frac{w_i(x) f(x)}{p_i(x)} \right)^2 p_i(x) d\mu(x) - \mu_i^2 \\ &= \int \frac{w_i^2(x) f^2(x)}{p_i(x)} d\mu(x) - \mu_i^2. \end{aligned}$$

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

$$\begin{aligned} V[F] &= V[\sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} F_{ij}] = \sum_{i=1}^{n} \frac{1}{n_i^2} \sum_{j=1}^{n_i} V[F_{ij}] \qquad (9) \\ &= \sum_{i=1}^{n} \frac{1}{n_i^2} \sum_{j=1}^{n_i} \sigma_i^2 = \sum_{i=1}^{n} \frac{1}{n_i} \sigma_i^2 \\ &= \sum_{i=1}^{n} \frac{1}{n_i} \left( \int \frac{w_i^2(x) f^2(x)}{p_i(x)} d\mu(x) - \mu_i^2 \right) \\ &= \sum_{i=1}^{n} \int \frac{w_i^2(x) f^2(x)}{n_i p_i(x)} d\mu(x) - \sum_{i=1}^{n} \frac{1}{n_i} \mu_i^2. \end{aligned}$$

The variance of the integral estimator depends on how the total number of samples $N$ are distributed among the different techniques. We can state the following theorem:

**Theorem 1:** For given sequences $\{n_i\}$, $\{\sigma_i^2\}$, $1 \leq i \leq n$, if variances $\{\sigma_i^2\}$ do not depend on sample numbers $\{n_i\}$, then the arrangement that minimizes (maximizes) $V[F]$ in Eq. 9 is given by the pairing of the $\{n_i\}$, $\{\sigma_i^2\}$ sequences in same order (inverse order) respectively.

*Proof* It is enough to apply the *rearrangement inequality* [HLP52, page 261] to the sequences $\{n_i\}$, $\{\sigma_i^2\}$ in the expression for the variance $V[F] = \sum_{i=1}^{n} \frac{1}{n_i} \sigma_i^2$. □

This theorem states that methods of higher variance should be given more samples than methods of lower variance. For example, a good strategy to choose the count of samples would be

$$n_i \propto \sigma_i^2. \tag{10}$$

But we can do better, as we will show in the following theorems, where the proofs can be found in the Appendix B.

**Theorem 2:** If variances $\{\sigma_i^2\}$ are independent of the number of samples, then the distribution $\{n_i\}, \sum_{i=1}^{n} n_i = N$, that minimizes Eq. 9 is

$$n_i \propto \sigma_i. \tag{11}$$

Substituting the normalized values of Eq. 11 into the expression for the variance of the integral estimator, we have the minimum variance

$$V[F] = \frac{1}{N} \sum_{i=1}^{n} \frac{\sum_k \sigma_k}{\sigma_i} \sigma_i^2 = \frac{1}{N} \left( \sum_{i=1}^{n} \sigma_i \right)^2. \tag{12}$$

While if we use equal count of samples for each technique, i.e., for all $i$, $n_i = \frac{N}{n}$, the variance is

$$V[F] = \frac{n}{N}\sum_{i=1}^{n}\sigma_i^2. \tag{13}$$

Equality between Eq. 12 and Eq. 13 will only happen when all variances $\sigma_i^2$ are equal.

### 2.1. Cost and efficiency

Let us consider now the cost of sampling. We want to minimize now the cost times variance, i.e. $C_T \times V[F]$, which is the inverse of efficiency. Using Lagrange multipliers and Cauchy-Schwartz inequality we can prove (see Appendix B) the following theorem:

**Theorem 3:** Let the variances $\{\sigma_i^2\}$ be independent of the number of samples, and $c_i$ be the cost of sampling technique $i$, making the total cost be $C_T = \sum_i n_i c_i$. Then the product $C_T \times V[F]$ takes its minimum at the distribution of samples

$$n_i \propto \frac{\sigma_i}{\sqrt{c_i}}. \tag{14}$$

Using these values into the expression $(C_T \times V[F])$ we find

$$(C_T \times V[F])_{\min} = \left(\sum_i \sigma_i\sqrt{c_i}\right)^2. \tag{15}$$

## 3. Variance of balance heuristic estimator

Veach defined balance heuristic estimator setting the weights as:

$$w_i(x) = \frac{n_i p_i(x)}{\sum_{k=1}^{n} n_k p_k(x)}. \tag{16}$$

The estimator of method $i$ becomes

$$F_{ij} = \frac{n_i f(X_{ij})}{\sum_{k=1}^{n} n_k p_k(X_{ij})}. \tag{17}$$

The expected value is

$$\mu_i = \int \frac{n_i f(x) p_i(x)}{\sum_{k=1}^{n} n_k p_k(x)} \mathrm{d}\mu(x). \tag{18}$$

The variance becomes

$$\sigma_i^2 = \int \frac{n_i^2 f^2(x) p_i(x)}{(\sum_k n_k p_k(x))^2} \mathrm{d}\mu(x) - \mu_i^2. \tag{19}$$

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

$$\begin{aligned} V[F] &= \sum_{i=n}^{n}\frac{1}{n_i}\sigma_i^2 \qquad (20)\\ &= \sum_{i=n}^{n}\frac{1}{n_i}\left(\int \frac{n_i^2 f^2(x) p_i(x)}{(\sum_k n_k p_k(x))^2}\mathrm{d}\mu(x) - \mu_i^2\right)\\ &= \sum_{i=n}^{n}\frac{1}{n_i}\int \frac{n_i^2 f^2(x) p_i(x)}{(\sum_k n_k p_k(x))^2}\mathrm{d}\mu(x) - \sum_{i=n}^{n}\frac{1}{n_i}\mu_i^2\\ &= \int \frac{f^2(x)}{\sum_{k=1}^{n} n_k p_k(x)}\mathrm{d}\mu(x) - \sum_{i=1}^{n}\frac{1}{n_i}\mu_i^2\\ &= \frac{1}{N}\left(\int \frac{f^2(x)}{\sum_{k=1}^{n}\alpha_k p_k(x)}\mathrm{d}\mu(x) - \sum_{i=1}^{n}\frac{1}{\alpha_i}\mu_i^2\right), \end{aligned}$$

where $\alpha_k = n_k/N$.

Let us set $N = 1$, i.e. take just a single sample, and call the estimator $F^1$. With abuse of language we call it the primary estimator, although in fact primary estimators will only exist for each of the techniques used, i.e., only each separated technique can be used with a minimum of 1 sample. Then the variance for $F^1$ is

$$\begin{aligned} V[F^1] &= \int \frac{f^2(x)}{\sum_{k=1}^{n}\alpha_k p_k(x)}\mathrm{d}\mu(x) - \sum_{i=1}^{n}\frac{1}{\alpha_i}\mu_i^2 \qquad (21)\\ &= \sum_i \frac{1}{\alpha_i}\left(\int \frac{\alpha_i^2 f^2(x)}{\sum_{k=1}^{n}\alpha_k p_k(x)}\mathrm{d}\mu(x) - \mu_i^2\right)\\ &= \sum_i \frac{1}{\alpha_i}\sigma_i^2. \end{aligned}$$

Eqs. 18 and 19 can be written in terms of the $\alpha_i$

$$\mu_i = \int \frac{\alpha_i p_i(x)}{\sum_{k=1}^{n}\alpha_k p_k(x)} f(x)\mathrm{d}\mu(x), \tag{22}$$

and

$$\sigma_i^2 = \int \frac{\alpha_i^2 f^2(x) p_i(x)}{(\sum_k \alpha_k p_k(x))^2}\mathrm{d}\mu(x) - \mu_i^2. \tag{23}$$

The task would be again the determination of optimal mixture $\alpha_1, \ldots, \alpha_n$. Note, however, that we cannot apply the results of the previous section, as $\sigma_i^2$ depends on $n_i$ (or $\alpha_i$) both in the numerator and the denominator of the integrand.

Without a criterion for optimal mixture, different sampling methods are usually given the same number of samples. The weighting defined by Eq. 16 is then

$$w_i(x) = \frac{p_i(x)}{\sum_{k=1}^{n} p_k(x)}. \tag{24}$$

## 4. A new estimator

We introduce a new estimator that belongs to the category of multi-sample methods when the weights are independent of the number of samples. The weighting scheme is similar to that of balance heuristic taking equal number of samples, i.e. Eq. 24, but we allow arbitrary number of samples. So, let us consider in Eq. 1 the weights

$$w_i(x) = \frac{p_i(x)}{\sum_{k=1}^{n} p_k(x)}. \tag{25}$$

The estimator $F$ in Eq. 1 becomes

$$F = \sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\frac{f(X_{ij})}{\sum_{k=1}^{n} p_k(X_{ij})}. \tag{26}$$

The estimators $F_{ij}$ in Eq. 5 become

$$F_{ij} = \frac{f(X_{ij})}{\sum_{k=1}^{n} p_k(X_{ij})}. \tag{27}$$

and the variance Eq. 8 is

$$\sigma_i^2 = \int \frac{p_i(x) f^2(x)}{\left(\sum_{k=1}^n p_k(x)\right)^2} \mathrm{d}\mu(x) - \mu_i^2. \tag{28}$$

The variance Eq. 9 of the integral estimator is

$$\begin{aligned} V[F] &= \sum_{i=1}^n \frac{1}{n_i} \sigma_i^2 \\ &= \sum_{i=1}^n \frac{1}{n_i} \left( \int \frac{p_i(x) f^2(x)}{\left(\sum_{k=1}^n p_k(x)\right)^2} \mathrm{d}\mu(x) - \mu_i^2 \right). \end{aligned} \tag{29}$$

Observe two things. First, that we can apply Theorems 1&2, as now the variances $\sigma_i^2$ do not depend anymore on the count of samples $n_i$, and second, that this estimator, when all $n_i$ are equal, is the balance heuristic with equal count of sampling. However, we can use now the results of Theorems 2&3 and prove that by the careful selection of sample numbers, the new estimator is better than balance heuristic with equal count of samples.

The application of Theorem 2, i.e., taking $n_i \propto \sigma_i$, guarantees that we always improve on balance heuristic with equal weight when using estimator Eq. 26, up to the statistical error in estimating the $\sigma_i$ values. The following inequality shows, on the left side, the optimal variance expressed in Eq. 12, on the right side, the variance for balance heuristic with equal count

$$\frac{1}{N} \left( \sum_{i=1}^n \sigma_i \right)^2 \leq \frac{n}{N} \left( \sum_{i=1}^n \sigma_i^2 \right). \tag{30}$$

It is easy to see that there is a significant gain only when the differences between the $\sigma_i$ values are high, and that the maximum attainable gain is up to $n$ times. Observe also that the same optimization can be applied to cut-off, power, and maximum heuristic [Vea97] when those heuristics use equal count of samples for each technique.

Suppose we take into account the cost $c_i$ of sampling each technique, then the total cost is $C_T = \sum_i n_i c_i$. We want to minimize now the cost times variance, i.e., $C_T \times V[F]$, which is the inverse of efficiency. Applying Theorem 3, we know that the optimal values are

$$n_i \propto \frac{\sigma_i}{\sqrt{c_i}}, \tag{31}$$

and then the optimal $C_T \times V[F]$ value is the left hand side of Eq. 32, while the right hand side is the value corresponding to taking equal count of sampling.

$$\left( \sum_{i=1}^n \sigma_i \sqrt{c_i} \right)^2 \leq \left( \sum_{i=1}^n c_i \right) \left( \sum_{i=1}^n \sigma_i^2 \right). \tag{32}$$

The increase in efficiency is not anymore limited to $n$ times.

## 5. One-sample balance heuristic estimator

We can define now the following primary estimator, i.e., using only one sample (in this case it is fully legitimate to use this term) $\mathcal{F}^1$ for $\mu = \int f(x) \mathrm{d}\mu(x)$ as

$$\mathcal{F}^1 = \frac{f(x)}{\sum_k \alpha_k p_k(x)}. \tag{33}$$

Indeed it is an unbiased estimator

$$\begin{aligned} E[\mathcal{F}^1] &= \int \frac{f(x) \sum_k \alpha_k p_k(x)}{\sum_k \alpha_k p_k(x)} \mathrm{d}\mu(x) \\ &= \int f(x) \mathrm{d}\mu(x) = \mu. \end{aligned} \tag{34}$$

The estimator $\mathcal{F}^1$ was introduced by Veach as the *one-sample balance heuristic estimator*. We first sort out with the $\alpha_i$ distribution which technique to use and then we sample this technique. 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 $\alpha_i$ are called the mixture coefficients, and represent the average count of samples from each technique. The variance of this estimator can be obtained by simple application of the definition of variance,

$$\begin{aligned} V[\mathcal{F}^1] &= E[(\mathcal{F}^1)^2] - E^2[\mathcal{F}^1] \\ &= \int \left( \frac{f(x)}{\sum_k \alpha_k p_k(x)} \right)^2 \left( \sum_k \alpha_k p_k(x) \right) \mathrm{d}\mu(x) - \mu^2 \\ &= \int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} \mathrm{d}\mu(x) - \mu^2. \end{aligned} \tag{35}$$

Observe that, if we write the second moment of $\mathcal{F}^1$ with respect to technique $i$ as $E_i[(\mathcal{F}^1)^2]$, we can write

$$E[(\mathcal{F}^1)^2] = \sum_{1=1}^n \alpha_i E_i[(\mathcal{F}^1)^2]. \tag{36}$$

The expected value $\mu_i'$ of $\mathcal{F}^1$ with respect to technique $i$, which we write as $E_i[\mathcal{F}^1]$, is

$$\mu_i' = E_i[\mathcal{F}^1] = \int \frac{f(x) p_i(x)}{\sum_k \alpha_k p_k(x)} \mathrm{d}\mu(x) = \frac{\mu_i}{\alpha_i}, \tag{37}$$

and its variance $\sigma'^2_i$ with respect to technique $i$, which we write as $V_i[\mathcal{F}^1]$, is

$$\begin{aligned} \sigma'^2_i = V_i[\mathcal{F}^1] &= \int \frac{f^2(x) p_i(x)}{\left(\sum_k \alpha_k p_k(x)\right)^2} \mathrm{d}\mu(x) - \left( \frac{\mu_i}{\alpha_i} \right)^2 \\ &= \frac{1}{\alpha_i^2} \left( \int \frac{\alpha_i^2 f^2(x) p_i(x)}{\left(\sum_k \alpha_k p_k(x)\right)^2} \mathrm{d}\mu(x) - \mu_i^2 \right) \\ &= \frac{\sigma_i^2}{\alpha_i^2}. \end{aligned} \tag{38}$$

Using Eq. 38, $V[F^1]$ in Eq. 21 can be written as

$$V[F^1] = \sum_{i=1}^n \alpha_i \sigma'^2_i. \tag{39}$$

By isolating the second moment integral in the first equality in Eq. 38, weighting by $\alpha_i$, and taking into account Eq. 35, the variance of estimator $V[\mathcal{F}^1]$ can be written as a function of the $\sigma'^2_i$

$$\begin{aligned} V[\mathcal{F}^1] &= \sum_{i=1}^n \alpha_i \left( \sigma'^2_i + \mu'^2_i \right) - \mu^2 \\ &= \sum_{i=1}^n \alpha_i \left( \sigma'^2_i + \mu'^2_i - \mu^2 \right). \end{aligned} \tag{40}$$

### 5.1. Minimum variance

Observe that we cannot apply the rearrangement inequality, neither Theorem 2, to Eq. 39, as ${\sigma'}^2_i$ depends on $\alpha_i$. Should we be able to apply the rearrangement inequality, then $\alpha_i$ would be now inversely proportional to the variances ${\sigma'}^2_i$, i.e., the higher the variance, the less the count of samples. Havran and Sbert found in [HS14] the condition for the minimum variance of the one-sample estimator, and which happens when all $E_i[(\mathcal{F}^1)^2]$ values are equal, i.e., for all $i, j$,

$$E_i[(\mathcal{F}^1)^2] = E_j[(\mathcal{F}^1)^2]. \tag{41}$$

This can be written as

$$\sigma_i'^2 + {\mu'}^2_i = \sigma_j'^2 + {\mu'}^2_j. \tag{42}$$

Observe that the $\{\alpha_i\}$ values are implicit in the Eqs. 41 and 42. A similar relationship exists for the multi-sample estimator, i.e., the minimum variance happens when (see proof in Appendix C), for all $i, j$,

$$\begin{aligned} & {\sigma'}^2_i + 2\mu \int \frac{f(x)p_i(x)}{(\sum_{k=1}^n \alpha_k p_k(x))^2} \mathrm{d}\mu(x) \\ = \; & {\sigma'}^2_j + 2\mu \int \frac{f(x)p_j(x)}{(\sum_{k=1}^n \alpha_k p_k(x))^2} \mathrm{d}\mu(x). \end{aligned} \tag{43}$$

For the sake of completeness, let us give the minimum variance condition in terms of ${\sigma'}_i$ values when Theorem 2 can be applied. In that case the optimal values are for all $i$ $\alpha_i = \frac{\sigma_i}{\sum_k \sigma_k}$, and as $\sigma_i = \alpha_i {\sigma'}_i$, we have $\sigma_i = \frac{\sigma_i}{\sum_k \sigma_k} {\sigma'}_i$, thus ${\sigma'}_i$ values are the same for all $i$.

### 5.2. Variance of multi-sample balance heuristic against variance of one-sample balance heuristic

Observe that the expressions for the variance, Eq. 35 and Eq. 21 differ in

$$V[\mathcal{F}^1] - V[F^1] = \sum_{i=1}^n \frac{1}{\alpha_i} \mu_i^2 - \mu^2. \tag{44}$$

We can state the following theorem:

**Theorem 4:** The expression in Eq. 44 is always positive, and becomes zero only in two cases: 1) when for all $i$ all values $\frac{\mu_i}{\alpha_i}$ are equal, implying that all techniques used are the same and, 2) when $f(x) \propto \sum_p \alpha_k p_k(x)$.

*Proof* Consider $\mu_i' = E_i[\mathcal{F}^1]$, then by Eq. 37:

$$\mu_i' = \frac{\mu_i}{\alpha_i}. \tag{45}$$

Using Eq. 22 we can write

$$\sum_{i=1}^n \frac{1}{\alpha_i} \mu_i^2 - \mu^2 = \sum_{i=1}^n \alpha_i {\mu'}^2_i - \left( \sum_{i=1}^n \alpha_i {\mu'}_i \right)^2. \tag{46}$$

Applying now Jensen's inequality [HLP52] to the convex function $y = x^2$,

$$\left( \sum_{i=1}^n \alpha_i {\mu'}_i \right)^2 \le \sum_{i=1}^n \alpha_i {\mu'}^2_i \tag{47}$$

and thus the difference in variances $V[\mathcal{F}^1] - V[F^1]$ is always positive or zero.
Equality in Eq. 47 will happen when all $\mu_i'$ are equal, which implies that either
1) for all $i$

$$\alpha_i = \frac{\mu_i}{\mu}, \tag{48}$$

and all techniques used are the same, as

$$\frac{\mu_i}{\alpha_i} = \int \frac{f(x)p_i(x)}{\sum_{k=1}^n \alpha_k p_k(x)} \mathrm{d}\mu(x), \tag{49}$$

which only happens when for all $i$, $p_i(x)$ are equal, or
2) the zero variance case,

$$f(x) = C \times \sum_{i=1}^n \alpha_i p_i(x), \tag{50}$$

with $C$ constant, because in this case for all $i$, $\mu_i' = E_i[\mathcal{F}^1] = C$. □

### 5.3. Zero variance

Both multi-sample and one-sample balance heuristic can have zero variance, where (excluding the trivial case of $f(x)$ constant) for some mixture $\{\alpha_i^\star\}$, $f(x) \propto \sum_k \alpha_k^\star p_k(x)$. By choosing the values $\{\alpha_i^\star\}$ we obtain zero variance. Observe that in this case the difference in Eq. 46 is zero and thus $\sum_i \frac{\mu_i^2}{\alpha_i} = \mu^2$. This case of zero variance was not considered in the derivation of Theorem 9.5 of Veach's thesis [Vea97].

Observe also that for a function $f(x)$ equal to a product of functions (such as in the rendering equation) and taking these functions as sampling techniques, the zero variance case can never happen, except in the trivial case where all functions are equal.

## 6. Discussion

Theorem 1 and Theorem 2 tell us how to take the samples when the variances of the multiple importance sampling estimators do not depend on the weights. Theorem 3 extends Theorem 2 to take into account the cost of sampling. We are able to exploit Theorems 2&3 to obtain a provably better estimator than balance heuristic with equal count of samples. This new estimator requires the knowledge of the $\sigma_i^2$ quantities, and it is worth using when there is a big difference between these quantities and also between the costs $c_i$.

Theorems 1-3 are not applicable in general to balance heuristic, so we cannot state that methods of higher $\sigma_i^2$ value deserve more samples. In fact, supposing that the rearrangement inequality were applicable, we would assign rather less samples to higher $\sigma_i^2$ value methods because of Eq. 39. Veach recommended the use of equal count of samples for all techniques, but there are better strategies and it can even happen that excluding one or more high variance methods from the sampling increases the accuracy.

In Table 1 we show the conditions for minimum variance that have to be met by the different estimators.

In Appendix A we give four examples of equal count of samples versus the count of samples computed by our new estimator and the heuristic in [HS14].

We have shown by Theorem 4 that it always pays off to use the multi-sample balance heuristic against the one-sample balance heuristic. Intuitively it means that the random decision always adds some variance. Again, Theorems 1-3 are not applicable, so our objective should be to mimic the integrand with the linear combination of the probability densities of the methods. In an ideal linear combination, zero variance can be achieved. The formulae for the variance of the two balance heuristic estimators are shown in Table 2. Observe that, although the first terms of both variances are formally equal, the first term of the multi-sample heuristic does not represent the expected value of the second moment of an estimator of the integral $I$, as it is the case for the one-sample heuristic. Also, in the multi-sample case, the $\alpha_i$ values do not represent any probabilities, as they do in the one-sample case, but only the proportion of samples for each technique.

## 7. Results

We show below the results for the new estimator described in Section 4 for environment map illumination, using four sampling strategies, in terms of mean square error (MSE). The rendered images and charts as dependence of MSE on the number of samples taken are shown in Fig. 1. The first sampling strategy samples according environment map, the second according to BRDF.cos($\theta$), the third one uses balance heuristic as described by Veach and Guibas [VG95, Vea97], the fourth one according to Lu et al. [LPG13]. The last fifth line in charts gives the results for the proposed heuristic in Section 4.

We have modified the algorithm in adaptive way. Each pixel is sampled independently, taking $N$ samples for each pixels. The adaptive sampling algorithm has 2 stages, a pilot one and a main stage. During pilot stage, which contains $0.2N$ samples, we estimate $\sigma_i^2$ given by Eq. 28 for both sampling strategies, taking equal number of samples from the two sampling techniques. Then the main sampling stage is subdivided into 8 sampling substages, each taking $0.1N$ samples. Values $\sigma_i^2$ are incrementally updated in each substage. Parameter $\alpha$ defined by Eq. 31 is computed at the end of the sampling substage, which is then used for the next sampling stage. Note that for arbitrary $\alpha$ value, the sampling scheme is unbiased, so the fact that this parameter is estimated from the samples of the previous stages does not compromise the unbiasedness of the multi-stage sampling scheme. Via modifying $\alpha$, earlier substages affect the variance of the later substages but not their mean.

We have used two objects (sphere and dragon), various BRDFs including varying one on the objects's surface and 8 different environment maps to create 10 test cases. The occluder as a grid is put around the objects to make the illumination more complicated and less predictable. This is why the images with objects contain visible shadow structures on their surfaces.

It is possible to see that the new algorithm described in Section 4 reaches faster convergence for all 10 scenes tested. The improvement against the balance heuristic is from moderate to even significant, depending on the setting of BRDF and environment map.

## 8. Conclusions

We have reexamined the variance for the Multiple Importance Sampling estimator and obtained the optimal sampling, taking also into account the cost of sampling, in case the weights do not depend on the number of samples taken from the different techniques. This has allowed us to obtain a new estimator, provably better than balance heuristic with equal count of sampling. For balance heuristic estimator, where the weights are defined as a function of the number of samples, we have compared the variances of multi-sample and one-sample balance heuristic, and given the expression of their difference. We have given the value of this difference for examples with different distribution of the count of samples, namely equal count and a recently introduced heuristic.

### Acknowledgements

Mateu Sbert has been funded by projects 61471261, 61179067 from the Natural Science Foundation of China, and by project TIN2013-47276-C6-1-R from Spanish Government. Vlastimil Havran acknowledges the partial support by Czech Science Foundation under project no. GA14-19213S. Laszlo Szirmay-Kalos has been funded by project OTKA K-104476 from Hungarian Academy of Sciences.

### References

[DGMR07] Douc R., Guillin A., Marin J. M., Robert C. P.: Minimum variance importance sampling via population monte carlo. *ESAIM: Probability and Statistics 11* (2007), 424–447. 10

[EMLB15] Elvira V., Martino L., Luengo D., Bugallo M. F.: Efficient multiple importance sampling estimators. *IEEE Signal Process. Lett. 22*, 10 (2015), 1757–1761. URL: http://dx.doi.org/10.1109/LSP.2015.2432078, doi:10.1109/LSP.2015.2432078. 1

[HLP52] Hardy G., Littlewood J., Pólya G.: *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, 1952. URL: https://books.google.es/books?id=t1RCSP8YKt8C. 2, 5

[HS14] Havran V., Sbert M.: Optimal Combination of Techniques in Multiple Importance Sampling. In *Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (New York, NY, USA, Nov. 2014), VRCAI '14, ACM, pp. 141–150. URL: http://doi.acm.org/10.1145/2670473.2670496, doi:10.1145/2670473.2670496. 1, 5, 7

[LPG13] Lu H., Pacanowski R., Granier X.: Second-Order Approximation for Variance Reduction in Multiple Importance Sampling. *Computer Graphics Forum 32*, 7 (2013), 131–136. URL: http://dx.doi.org/10.1111/cgf.12220, doi:10.1111/cgf.12220. 1, 6, 8

[PBPP11] Pajot A., Barthe L., Paulin M., Poulin P.: Representativity for robust and adaptive multiple importance sampling. *IEEE Trans. Vis. Comput. Graph. 17*, 8 (2011), 1108–1121. URL: http://dx.doi.org/10.1109/TVCG.2010.230, doi:10.1109/TVCG.2010.230. 1

[SNJ*14] Subr K., Nowrouzezahrai D., Jarosz W., Kautz J., Mitchell K.: Error analysis of estimators that use combinations of stochastic sampling strategies for direct illumination. *Computer Graphics Forum (Proceedings of EGSR) 33*, 4 (June 2014), 93–102. doi:10.1111/cgf.12416. 1

[Vea97] Veach E.: *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, 1997. 1, 2, 4, 5, 6

| Model | Condition: $\forall i, j$ |
|---|---|
| $\{\sigma_i^2\}$ independent of $\{\alpha_i\}$ | $\sigma'^2_i = \sigma'^2_j$ |
| one-sample | $\sigma'^2_i + \mu'^2_i = \sigma'^2_j + \mu'^2_j$ |
| multi-sample | $\sigma'^2_i + 2\mu \int \frac{f(x)p_i(x)}{(\sum_{k=1}^n \alpha_k p_k(x))^2} d\mu(x) = \sigma'^2_j + 2\mu \int \frac{f(x)p_j(x)}{(\sum_{k=1}^n \alpha_k p_k(x))^2} d\mu(x)$ |

**Table 1:** *Minimum variance conditions for MIS when $\{\sigma_i^2\}$ are independent of the $\{\alpha_i\}$, Theorem 2, first row, and balance heuristic one-sample estimator, second row, and multiple-sample estimator, third row. $\mu'_i$ and $\sigma'_i$ are defined in Eq. 37 and Eq. 39, respectively.*

| Model | Variance as a sum | Variance |
|---|---|---|
| one-sample | $\sum_{i=1}^n \alpha_i \left(\sigma'^2_i + \mu'^2_i - \mu^2\right)$ | $\int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} d\mu(x) - \mu^2$ |
| multi-sample | $\sum_{i=1}^n \frac{1}{\alpha_i}\sigma_i^2 = \sum_{i=1}^n \alpha_i \sigma'^2_i$ | $\int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} d\mu(x) - \sum_{i=1}^n \frac{1}{\alpha_i}\mu_i^2$ |

**Table 2:** *Formulae for the variance of balance heuristic estimators, for $N = 1$, of one-sample model, $V[\mathcal{F}^1]$ (first row) and multi-sample $V[F^1]$ (second row), second column as a sum of variances of independent estimators. $\mu_i$ and $\sigma_i$ are defined in Eq. 22 and Eq. 23, respectively. $\mu'_i$ and $\sigma'_i$ are defined in Eq. 37 and Eq. 39, respectively.*

[VG95] VEACH E., GUIBAS L. J.: Optimally Combining Sampling Techniques for Monte Carlo Rendering. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1995), SIGGRAPH '95, ACM, pp. 419–428. URL: http://doi.acm.org/10.1145/218380.218498, doi:10.1145/218380.218498. 1, 2, 6

## Appendix A: 1D examples

Below, we provide four 1D examples for the two heuristics on the number of samples for balance heuristic. First, equal count of samples and second, the one in [HS14], where the count of samples was taken inversely proportional to the variances of the estimators of each technique taken independently. We also provide the values of the difference in Eq. 44, $V[\mathcal{F}^1] - V[F^1]$.

We will extend here the example from [HS14]. We additionally provide in supplemental material Mathematica code to try any function with any pdfs.

### Example 1

Suppose we want to solve the integral

$$\mu = I = \int_{\frac{3}{2\pi}}^{\pi} x\left(x^2 - \frac{x}{\pi}\right)\sin(x)dx = 10.2876 \tag{51}$$

by MIS sampling on functions $x$, $(x^2 - \frac{x}{\pi})$, and $\sin(x)$ respectively. We first find the normalization constants: $\int_{\frac{3}{2\pi}}^{\pi} x dx = 4.82082$, $\int_{\frac{3}{2\pi}}^{\pi} (x^2 - \frac{x}{\pi})dx = 8.76463$, $\int_{\frac{3}{2\pi}}^{\pi} \sin(x)dx = 1.88816$.

Then we find the three variances when doing importance sampling with all three pdfs respectively. If $G_1, G_2$, and $G_3$ are the three independent estimators of $I$ for the three techniques, then:

$$\begin{aligned} V[G_1] &= 4.82082 \int_{\frac{3}{2\pi}}^{\pi} x\left(x^2 - \frac{x}{\pi}\right)^2 \sin^2(x)dx - I^2 \\ &= 26.6759 \\ V[G_2] &= 8.76463 \int_{\frac{3}{2\pi}}^{\pi} x^2(x^2 - \frac{x}{\pi})\sin^2(x)dx - I^2 \\ &= 23.507 \\ V[G_3] &= 1.88816 \int_{\frac{3}{2\pi}}^{\pi} x^2(x^2 - \frac{x}{\pi})^2 \sin(x)dx - I^2 \\ &= 111.065 \end{aligned}$$

Taking $\alpha_i \propto \frac{1}{V[G_i]}$, we obtain

$$(\alpha_1, \alpha_2, \alpha_3) = (0.42105, 0.47782, 0.10113)$$

Using Eq. 35 to compute $V[\mathcal{F}^1]$:

$$V(\alpha_1, \alpha_2, \alpha_3) = \int_{\frac{3}{2\pi}}^{\pi} \frac{x^2(x^2 - \frac{x}{\pi})^2 \sin^2(x)}{\frac{2\alpha_1 x}{\pi^2} + \frac{\alpha_2}{8.764}(x^2 - \frac{x}{\pi}) + \frac{1}{2}\alpha_3 \sin(x)} dx - \mu^2 \tag{52}$$

Substituting the $(\alpha_1, \alpha_2, \alpha_3)$ values found above, we have that $V(0.42105, 0.47782, 0.10113) = 24.2211$. On the other hand, $V(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) = 30.1676$, and there is a gain of 24.55%.

We take now into account the sampling costs given in [HS14], i.e., $\alpha_i \propto \frac{1}{c_i V[G_i]}$, with $c_1 = 1$, $c_2 = 6.24$, $c_3 = 3.28$. The relative efficiency increases from 1.24 for equal sampling costs to 2.12 when taking into account the different sampling costs.

### Example 2

Now let us solve the integral

$$\mu = I = \int_{\frac{3}{2\pi}}^{\pi} \left(x^2 - \frac{x}{\pi}\right)\sin^2(x)dx = 3.59615 \tag{53}$$

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

We find $(V[G_1], V[G_2], V[G_3])$=(5.6334, 9.41988, 4.54464), and thus $(\alpha_1, \alpha_2, \alpha_3)$=(0.35241, 0.21075, 0.43684), and $V(5.6334, 9.41988, 4.54464) = 4.6041$. On the other hand, $V(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) = 5.01917$, and there is a small gain of 9%.

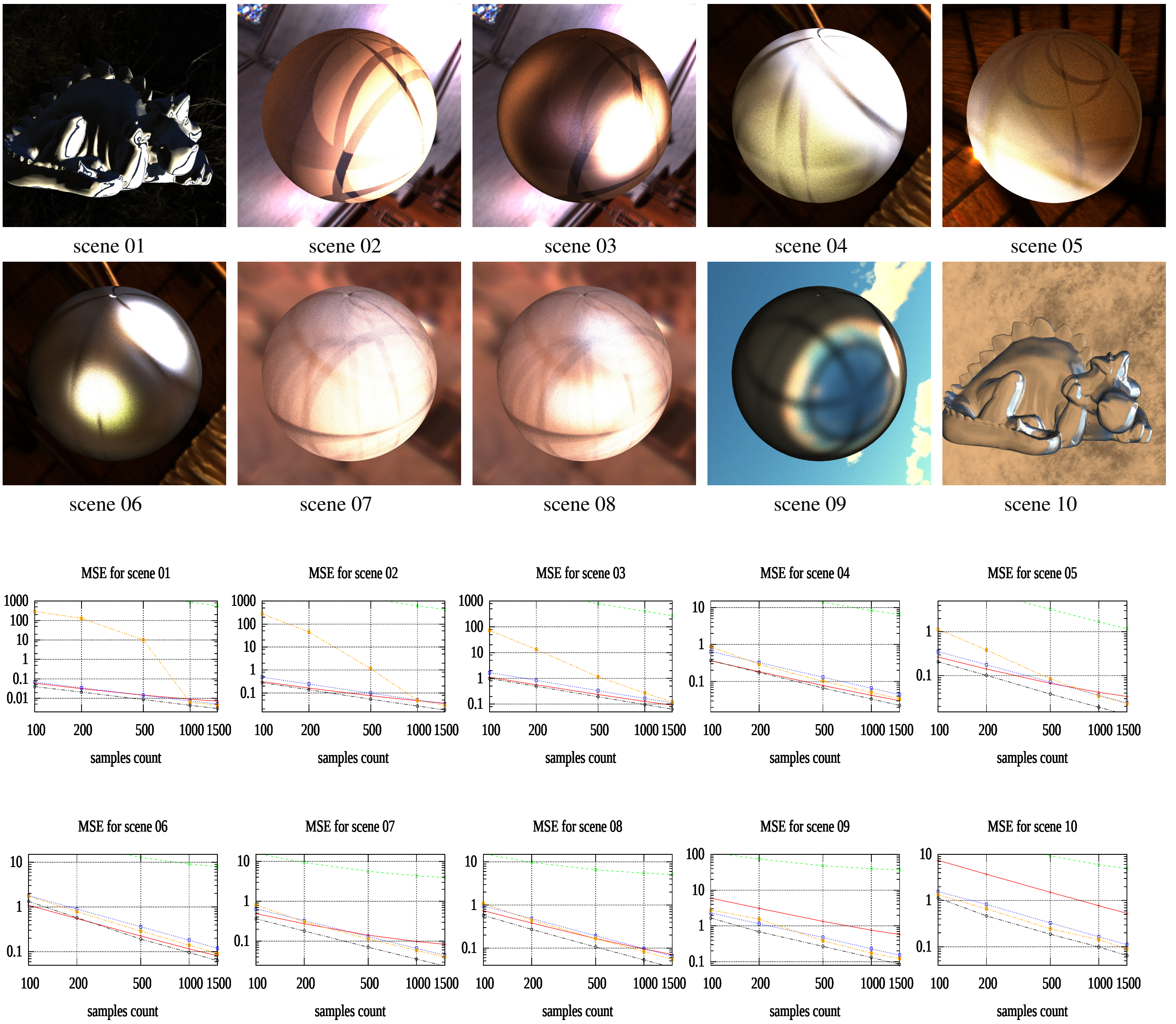


**Figure 1:** *Rendered images and MSE errors for 10 scenes. The images were rendered for 200 samples per pixel for the representative cases and multiple importance sampling with balance heuristic $\alpha_1 = \alpha_2 = 0.5$ and new estimator as described in Section 4. The proposed heuristic has pilot stage sampling $0.2N$ samples from all $N$ samples taken, $N \in \{100, 200, 500, 1000, 1500\}$. The rest of computation is organized in sampling stages $0.1N$ samples each, evaluating the number of samples for the next sampling stage as described Eq. 31. Legend for charts with MSE errors. On x-axis is number of samples, on y-axis is MSE for 5 importance sampling algorithms. Lines: red color – sampling according to environment map, green color – sampling according to BRDF.$\cos(\theta)$, blue – sampling according to balance heuristic, orange – sampling method according to Lu et al. [LPG13], black – sampling according to the newly proposed heuristic described by Eq. 31.*

Considering now the cost of sampling each technique, for the heuristics inversely proportional to variance times cost we have a quotient of efficiencies of 1.87.

**Example 3**

As third example let us solve the integral

$$\mu = I = \int_{\frac{3}{2\pi}}^{\pi} x + \left(x^2 - \frac{x}{\pi}\right) + \sin(x)\mathrm{d}x = 15.4736 \qquad (54)$$

using the same functions as before. In this case we find $(V[G_1], V[G_2], V[G_3]) = (4.09963, 35.3278, 9586.73)$, and thus $(\alpha_1, \alpha_2, \alpha_3) = (0.89568, 0.10394, 0.00038)$, so that

| | $\alpha_k \propto \frac{1}{n}$ | $\alpha_k \propto \frac{1}{V[G_i]}$ | Eq. 44($\frac{1}{n}$) | Eq. 44($\frac{1}{V[G_i]}$) |
|---|---|---|---|---|
| 1 | 30.167 | **24.221** | 1.004 | 0.10946 |
| 2 | 5.01917 | **4.6041** | 0.102 | 0.05131 |
| 3 | 13.354 | **2.050** | 2.666 | 0.02897 |
| 4 | **0** | 0.121 | 0 | 0.00039 |

**Table 3:** *In the two first columns, the variances, $V[\mathcal{F}^1]$, for the 4 examples for balance heuristic, and for the two heuristics on the number of samples taken from each sampling technique using one-sample estimator. The best results are in bold. In the two last columns, we show the values for Eq. 44, $V[\mathcal{F}^1]-V[F^1]$, for the two heuristics considered.*

| | $\alpha_k \propto \frac{1}{n}$ | $\alpha_k \propto \frac{1}{c_i V[G_i]}$ |
|---|---|---|
| 1 | 105.788 | **49.750** |
| 2 | 19.470 | **10.466** |
| 3 | 46.827 | **4.047** |
| 4 | **0** | 0.110 |

**Table 4:** *Variance times cost (i.e. inefficiency, the smaller the better) for the 4 examples in this appendix for balance heuristic using the two heuristics on the number of samples taken and for the one-sample estimator (second column equal number of samples for each technique, third column - variance based heuristic for the number of samples). The best results are in bold.*

$V(0.89567, 0.10393, 0.00038) = 2.04964$. On the other hand, $V(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) = 13.354$, and there is a gain of 551.21%. Considering now the cost of sampling each technique, for the heuristics inversely proportional to variance times cost we have a quotient of efficiencies of 11.57.

**Example 4**

As a last example, and to see the limitations of the heuristic, consider the integral of the sum of the three pdfs used

$$\mu = I = \int_{\frac{3}{2\pi}}^{\pi} \frac{x}{4.82082} + \frac{\left(x^2 - \frac{x}{\pi}\right)}{8.764} + \frac{\sin(x)}{1.88816} dx = 3 \quad (55)$$

It is clear that the variance is zero for equal count of samples, as in this case $g(x) = f(x)$, while it can not indeed be zero for the other heuristic considered. The variance for inverse of variances heuristic is 0.121359. Introducing the costs, the value of variance times cost is 0.110. Results are summarized in Tables 3 and 4.

In Table 5 we compare balance heuristic with equal number of samples with the estimator defined in Section 4 with the optimal number of samples obtained from Eq. 14. We see that if we know in advance the $\sigma_i$ values, or estimate them with enough precision, we always improve using the new estimator, and can even get the zero optimal zero variance. If we do not use the cost of sampling, the improvement for our first three examples is very small, but we are still able to get the zero variance solution.

| | $\alpha_k \propto \frac{1}{n}$ | $\alpha_k \propto \frac{\sigma_i}{\sqrt{c_i}}$ |
|---|---|---|
| 1 | 102.26 | **89.40** |
| 2 | 17.244 | **15.441** |
| 3 | 37.478 | **31.08** |
| 4 | **0** | **0** |

**Table 5:** *Variance times cost (i.e. inefficiency, the smaller the better) for the 4 examples in this appendix for the estimator with optimal number of samples defined in Section 4 and for balance heuristic with equal number of samples for multiple-sample estimator (second column equal number of samples for each technique, third column - Eq. 14 for the number of samples). The best results are in bold.*

## Appendix B: Proofs of Theorems 2&3

We present here the proofs of Theorems 2&3.

*Proof* To optimize the variance in Eq. 9 we take as the expression for the variance $V[F] = \sum_{i=1}^{n} \frac{1}{n_i}\sigma_i^2$, and we use Lagrange multipliers with the constraint $\sum_{i=1}^{n} n_i = N$ and objective function

$$\Lambda(n_i,\lambda) = \sum_{i=1}^{n} \frac{1}{n_i}\sigma_i^2 + \lambda(\sum_{i=1}^{n} n_i - N).$$

Taking partial derivatives and equating to zero, as the $\sigma_i$ do not depend on the $n_i$,

$$\begin{aligned} \frac{\partial\Lambda(n_i,\lambda)}{\partial n_j} &= \frac{\partial\left(\sum_{i=1}^{n}\frac{1}{n_i}\sigma_i^2\right)}{\partial n_j} + \frac{\partial\left(\lambda(\sum_{i=1}^{n} n_i - N)\right)}{\partial n_j} \\ &= -\frac{\sigma_j^2}{n_j^2} + \lambda = 0 \end{aligned} \quad (56)$$

Thus for all $j$

$$\lambda = \frac{\sigma_j^2}{n_j^2} \quad (57)$$

and the second term in Eq. 57 has to be equal for all $j$, which implies that $n_j \propto \sigma_j$.
The Hessian Matrix, obtained with the second derivatives of $V[F]$, is a diagonal matrix with positive diagonal values

$$\frac{\partial V[F]}{\partial_{n_j}\partial_{n_j}} = 2\frac{\sigma_j^2}{n_j^3} \quad (58)$$

and thus is positive-definite. The variance function is then strictly convex in its convex domain $\sum_{i=1}^{n} n_i = N$, where for all $i$, $0 < n_i < n$, meaning that the critical point is unique and a minimum.
Substituting the optimal values, i.e., $n_j \propto \sigma_j$, we find the minimum variance,

$$V_{min}[F] = \frac{1}{N}\left(\sum_{i=1}^{n}\sigma_i\right)^2 \quad (59)$$

Let us extend Theorem 2 now to include the cost of sampling. Let us consider $c_i$ the of sampling each technique, and the total cost is thus $C_T = \sum_i n_i c_i$. We want to minimize now the cost times variance, i.e., $C_T \times V[F]$, which is the inverse of efficiency. We use

Lagrange multipliers with the constraint $\sum_{i=1}^n n_i = N$ and objective function

$$\begin{aligned}\Lambda(n_i,\lambda) &= C_T \times V[F] + \lambda(\sum_{i=1}^n n_i - N) \\ &= \left(\sum_i n_i c_i\right)\left(\sum_{i=1}^n \frac{1}{n_i}\sigma_i^2\right) + \lambda(\sum_{i=1}^n n_i - N).\end{aligned} \tag{60}$$

Taking partial derivatives and equating to zero, as the $\sigma_i$ do not depend on the $n_i$,

$$\begin{aligned}\frac{\partial\Lambda(n_i,\lambda)}{\partial_{n_j}} &= \frac{\partial(\sum_i n_i c_i)}{\partial_{n_j}} V[F] + C_T \frac{\partial\left(\sum_{i=1}^n \frac{1}{n_i}\sigma_i^2\right)}{\partial_{n_j}} \\ &+ \frac{\partial(\lambda(\sum_{i=1}^n n_i - N))}{\partial_{n_j}} \\ &= c_j V[F] - C_T \frac{\sigma_j^2}{n_j^2} + \lambda = 0\end{aligned} \tag{61}$$

Multiplying by $n_j$ and adding over $j$

$$\begin{aligned}&\sum_j n_j c_j V[F] - \sum_j n_j C_T \frac{\sigma_j^2}{n_j^2} + \sum_j n_j \lambda \\ &= V[F]C_T - C_T V[F] + N\lambda = 0\end{aligned} \tag{62}$$

and thus $\lambda = 0$. Substituting this value in Eq. 61 we find that the optimal sampling counts are

$$n_j = \sqrt{\frac{C_T}{V[F]}}\frac{\sigma_j}{\sqrt{c_j}} \propto \frac{\sigma_j}{\sqrt{c_j}} \tag{63}$$

Using these values into the expression $(C_T \times V[F])$ we find

$$(C_T \times V[F])_{min} = \left(\sum_i \sigma_i \sqrt{c_i}\right)^2. \tag{64}$$

Let us show that this value is indeed a global minimum. Instead of using the Hessian matrix of $C_T \times V[F]$, let us prove it by applying Cauchy–Schwartz inequality to the sequences $\{\frac{\sigma_i}{\sqrt{n_i}}\}$ and $\{\sqrt{c_i n_i}\}$. Cauchy–Schwartz inequality states that for any two sequences of $n$ real numbers $\{x_i\}$ and $\{y_i\}$, the inequality holds $(\sum_i x_i y_i)^2 \le \sum_i x_i^2 \sum_i y_i^2$. We have then

$$\begin{aligned}&\left(\sum_i \sigma_i\sqrt{c_i}\right)^2 = \left(\sum_i \sqrt{c_i n_i}\frac{\sigma_i}{\sqrt{n_i}}\right)^2 \\ &\le \sum_i (\sqrt{c_i n_i})^2 \sum_i \left(\frac{\sigma_i}{\sqrt{n_i}}\right)^2 = \left(\sum_i c_i n_i\right)\left(\sum_i \frac{\sigma_i^2}{n_i}\right)\end{aligned} \tag{65}$$

□

## Appendix C: Proof of minimum variance condition for multi-sample estimator

Consider the variance Eq. 21, in the form

$$\begin{aligned}V[F^1] &= \int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} d\mu(x) - \sum_{i=1}^n \frac{1}{\alpha_i}\mu_i^2 \\ &= \int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} d\mu(x) - \sum_{i=1}^n \alpha_i {\mu'}_i^2\end{aligned} \tag{66}$$

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

$$\Lambda(\alpha_i,\lambda) = \int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} d\mu(x) - \sum_{i=1}^n \alpha_i {\mu'}_i^2 + \lambda(\sum_{i=1}^n \alpha_i - 1).$$

Taking partial derivatives and equating to zero,

$$\begin{aligned}\frac{\partial\Lambda(\alpha_i,\lambda)}{\partial\alpha_j} &= \frac{\partial\int \frac{f^2(x)}{\sum_{k=1}^n \alpha_k p_k(x)} d\mu(x)}{\partial\alpha_j} \\ &- \frac{\partial\sum_{i=1}^n \alpha_i {\mu'}_i^2}{\partial\alpha_j} + \lambda \\ &= -\int \frac{f^2(x)p_j(x)}{(\sum_{k=1}^n \alpha_k p_k(x))^2} d\mu(x) + {\mu'}_j^2 \\ &+ 2\sum_{i=1}^n \alpha_i {\mu'}_i \frac{\partial {\mu'}_i}{\partial\alpha_j} + \lambda \\ &= -{\sigma'}_j^2 + {\mu'}_j^2 \\ &- 2\sum_{i=1}^n \alpha_i {\mu'}_i \int \frac{f(x)p_j(x)}{(\sum_{k=1}^n \alpha_k p_k(x))^2} d\mu(x) + \lambda \\ &= -{\sigma'}_j^2 + {\mu'}_j^2 \\ &- 2(\sum_{i=1}^n \alpha_i {\mu'}_i)\int \frac{f(x)p_j(x)}{(\sum_{k=1}^n \alpha_k p_k(x))^2} d\mu(x) + \lambda \\ &= -{\sigma'}_j^2 + {\mu'}_j^2 \\ &- 2\mu\int \frac{f(x)p_j(x)}{(\sum_{k=1}^n \alpha_k p_k(x))^2} d\mu(x) + \lambda = 0\end{aligned} \tag{67}$$

Eq. 67 means that for all $i, j$

$$\begin{aligned}&{\sigma'}_i^2 + 2\mu\int \frac{f(x)p_i(x)}{(\sum_{k=1}^n \alpha_k p_k(x))^2} d\mu(x) \\ &= {\sigma'}_j^2 + 2\mu\int \frac{f(x)p_j(x)}{(\sum_{k=1}^n \alpha_k p_k(x))^2} d\mu(x) = \lambda\end{aligned} \tag{68}$$

Thus, the $\{\alpha_i\}$ values implicit in Eq. 68 correspond to a critical point in the variance of $F^1$, which is convex as stated by Douc et al. [DGMR07] (in fact, Douc et al. stated the convexity of the variance of one-sample $\mathcal{F}^1$ estimator, but the difference in variances of $\mathcal{F}^1$ and $F^1$ estimators amounts to a convex function), and thus the critical point is a global minimum. Multiplying the final equality of Eq. 67 by $\alpha_j$ and adding over $j$, we can obtain the value of the minimum variance $V_{min}[F^1]$ by isolating it in the following equation together with Eq. 68

$$\begin{aligned}&\sum_j \alpha_j {\sigma'}_j^2 \\ &+ \sum_j \alpha_j 2\mu\int \frac{f(x)p_j(x)}{(\sum_{k=1}^n \alpha_k p_k(x))^2} d\mu(x) \\ &= V_{min}[F^1] + 2\mu\int \frac{f(x)}{\sum_{k=1}^n \alpha_k p_k(x)} d\mu(x) = \lambda.\end{aligned} \tag{69}$$