

Robust Multiple Importance Sampling with Tsallis φ -Divergences

Mateu Sbert ^{1,*}  and László Szirmay-Kalos ²¹ Institute of Informatics and Applications, University of Girona, 17071 Girona, Spain² Department of Control Engineering and Information Technology, Budapest University of Technology and Economics, 1111 Budapest, Hungary

* Correspondence: mateu@ima.udg.edu; Tel.: +34-972-418419

Abstract: Multiple Importance Sampling (MIS) combines the probability density functions (pdf) of several sampling techniques. The combination weights depend on the proportion of samples used for the particular techniques. Weights can be found by optimization of the variance, but this approach is costly and numerically unstable. We show in this paper that MIS can be represented as a divergence problem between the integrand and the pdf, which leads to simpler computations and more robust solutions. The proposed idea is validated with 1D numerical examples and with the illumination problem of computer graphics.

Keywords: multiple importance sampling; Monte Carlo integration; φ -divergence; f-divergence; Tsallis divergence; Kullback–Leibler divergence; chi-square divergence; image synthesis

1. Introduction and Previous Work

Multiple Importance Sampling (MIS) [1,2] has been proven efficient in Monte Carlo integration. It is able to preserve the advantages of the combined techniques and requires only the calculation of the pdfs of all methods when a sample is generated with one particular method. The weighting scheme applied in MIS depends on the pdfs of the individual techniques and also on the number of samples generated with each of them.

MIS has been applied in a number of rendering algorithms, and its variance is extensively studied [1]. Several estimators have been proposed that are better than balance heuristics with equal sample budgets [3–6]. Lu et al. [7] proposed an adaptive algorithm for environment map illumination, which used the Taylor series approximation of the variance around an equal sample budget case. In [8,9], different equal sample number strategies were analyzed. Sbert et al. [10] considered the cost associated with the sampling strategies and obtained an adaptive solution by optimizing the variance using the Newton–Raphson method [11]. In [12], the Kullback–Leibler divergence was optimized instead of the variance. Several authors have shown that the variance of an importance sampling estimator is equal to a chi-square divergence [13–16], which, in turn, can be approximated by the Kullback–Leibler divergence up to the second order [17]. The optimal sample budget has also been targeted with neural networks [15,18]. Recently, a theoretical formula has been elaborated for the weighting functions [19]. In [16], the balance heuristic estimator was generalized by decoupling the weights from the sampling rates, and implicit solutions for the optimal case were given.

These techniques offer lower variance and, therefore, theoretically outperform MIS with equal number of samples. However, equations determining the optimal weighting and sample budget require the knowledge of the integrand and numerical solution methods. In computer graphics, for example, this integrand is not analytically available, so previous discrete samples should be used for the approximation, which introduces errors in the computation. Thus, it is not guaranteed that a theoretically superior estimator also performs better in practice.



Citation: Sbert, M.; Szirmay-Kalos, L. Robust Multiple Importance Sampling with Tsallis φ -Divergences. *Entropy* **2022**, *24*, 1240. <https://doi.org/10.3390/e24091240>

Academic Editors: Chun-Hung Liu, Jwo-Yuh Wu and Peter Y. Hong

Received: 26 July 2022

Accepted: 31 August 2022

Published: 3 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

This paper proposes an adaptive approach to automatically determine the sampling budgets of the combined methods based on the statistics of previous samples. To improve the robustness, instead of directly optimizing the variance, we make the combined pdf mimic the integrand by minimizing the divergence between them. From the possible alternatives, the Tsallis φ -divergence is used since this leads to a general and stable approach. We show that finding the optimum in all cases, including also Hellinger, chi-square, Kullback–Leibler divergences, and the variance, boils down to a non-linear equation stating that the γ -moments of the different combined techniques must be equal. This equation has unique root that can be found by Newton–Raphson iteration.

The organization of the paper is the following. In Section 2, the multi-sample version of the MIS estimator is reviewed where the number of samples allocated to each combined technique is fixed deterministically. Section 3 summarizes the one-sample MIS estimator where the particular sampling method is also selected randomly. In Section 4, the adaptation of the MIS weighting functions is discussed and we argue that the adaptation should be controlled not by the estimated variance but by an appropriately chosen divergence. Section 5 reformulates the MIS problem as the optimization of divergences and introduces the γ -moments. In Section 6, the details of the numerical optimization are discussed. Finally, the results are demonstrated with numerical examples and the direct lighting solution of computer graphics.

2. Multi-Sample Balance Heuristic Mis

Suppose we wish to estimate the value of integral $\int f(x)dx$ and have m proposal pdfs $p_i(x)$ to generate the j th sample X_{ij} of method i in the domain of the integral. With method i , n_i independent samples are drawn, so the total number of samples is $\sum_{i=1}^m n_i = N$. When the sample numbers n_i are fixed deterministically, the approach is called *multi-sample*.

The multi-sample MIS estimator [1] has the following expression:

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

where the weights satisfy the normalization condition:

$$\sum_{i=1}^m w_i(x) = 1.$$

Integral estimator F is unbiased, as its expected value μ is equal to the integral:

$$\mu = E[F] = \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \int \frac{w_i(x)f(x)}{p_i(x)} p_i(x) dx = \sum_{i=1}^m \mu_i = \int f(x) dx \quad (2)$$

where $\mu_i = \int w_i(x)f(x)dx$. The variance of the integral estimator is

$$V[F] = \sum_{i=1}^m \int \frac{w_i^2(x)f^2(x)}{n_i p_i(x)} dx - \sum_{i=1}^m \frac{\mu_i^2}{n_i}. \quad (3)$$

The variance of the estimator depends on the combination weights $w_i(x)$. The first step of their definition is to propose an algebraic form. The *balance heuristic* states that the weight of method i at domain point x should be proportional to the density of samples generated by method i in this point:

$$w_i(x) = \frac{\alpha_i p_i(x)}{p(\alpha, x)} \quad (4)$$

where $\alpha_i = n_i/N$ is the fraction of the samples allocated to method i , and $p(\alpha, x)$ is the mixture pdf:

$$p(\alpha, x) = \sum_{k=1}^m \alpha_k p_k(x). \quad (5)$$

Substituting this weighting function into the MIS estimator formulas, we obtain the multi-sample balance heuristics estimator:

$$F = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_i} \frac{f(X_{i,j})}{p(\alpha, X_{i,j})}. \quad (6)$$

The variance of the balance heuristics estimator is

$$V[F] = \frac{1}{N} \left(\int \frac{f^2(x)}{p(\alpha, x)} dx - \sum_{i=1}^m \alpha_i \mu_i'^2 \right), \quad (7)$$

where $\mu_i' = \mu_i / \alpha_i$.

3. One-Sample Balance Heuristic MIS

In the *one-sample* version of MIS, the numbers of samples used by particular techniques are not determined before starting the sampling process, but for each sample, the sampling method itself is selected randomly with the probability of fraction parameter α_i . As this approach introduces additional randomization, its variance is larger than that of the multi-sample version, but can also be used in cases when the number of total samples is less than the number of combined methods.

The one-sample balance heuristic estimator is given by

$$\mathcal{F} = \frac{f(X)}{p(\alpha, X)} \quad (8)$$

with variance

$$V[\mathcal{F}] = \int \frac{f^2(x)}{p(\alpha, x)} dx - \mu^2. \quad (9)$$

4. Adaptive MIS

Having the algebraic form of the weight function, the task is to find the optimal fractions α_i with the constraint that the sum of sample numbers must be equal to the total sample budget, i.e., $\sum_{i=1}^m \alpha_i = 1$. One possibility is to use some heuristics before starting the sampling process. A simple example is the *equal sample count* MIS that sets all fractions to be equal.

Adaptive techniques, on the other hand, try to minimize the error by controlling the fractions on-the-fly as new samples introduce additional information. During this, the following issues need to be considered:

- The integrals in the variance cannot be computed analytically, but must be estimated from the available finite number of samples drawn from the sampling distribution. This uncertainty may significantly affect the goodness of the final results.
- The optimization process should be fast and should not introduce too high overhead.

Unfortunately, the direct optimization for the variance does not meet these requirements. The integral of the variance can be estimated from the available samples X_i drawn from mixture distribution $p(\alpha, x)$ as:

$$\int \frac{f^2(x)}{p(\alpha, x)} dx = E_p \left[\frac{f^2(x)}{p^2(\alpha, x)} \right] \approx \frac{1}{N} \sum_{i=1}^N \left(\frac{f(X_i)}{p(\alpha, X_i)} \right)^2 \quad (10)$$

where $E_p[\xi]$ is the expectation of random variable ξ of pdf p . In this estimation, the ratios of the integrand and the pdf are squared causing high fluctuation and making the optimization process unstable.

Thus, instead of the variance, we need other optimization targets with more robust estimators in order to find the fractions α_k during adaptation.

5. MIS as Divergence between Distributions

MIS consists in looking for a distribution $p(\alpha, x)$ that mimics $f(x)$ as much as possible. If we have non-negative integrand $f(x) \geq 0$ with integral $\mu = \int f(x)dx$, then the integrand scaled down by the integral $g(x) = f(x)/\mu$ is also a distribution. The MIS problem can be stated as finding mixture pdf $p(\alpha, x)$ that minimizes the divergence between two distributions.

5.1. φ -Divergences

A φ -divergence $\mathcal{D}_\varphi(q, r)$ [20,21] is a measure of the dissimilarity between two probability distributions $q(x)$ and $r(x)$. It is defined by a convex function $\varphi(t)$ satisfying $\varphi(1) = 0$:

$$\mathcal{D}_\varphi(q, r) = \int r(x)\varphi\left(\frac{q(x)}{r(x)}\right)dx. \tag{11}$$

φ -divergences are always positive (by Jensen inequality) and are zero iff $q \equiv r$ (for strict convexity) [21].

For instance, for $\varphi(t) = t \log t$, we obtain the Kullback–Leibler divergence:

$$\mathcal{D}_\varphi(q, r) = \int q(x) \log \frac{q(x)}{r(x)} dx = KL(q, r). \tag{12}$$

Note that φ -divergence is not symmetric, thus we have in principle two options to measure the dissimilarity of probability densities q and r , using either $\mathcal{D}_\varphi(q, r)$ or $\mathcal{D}_\varphi(r, q)$. Any of these two options can be used in practice, although they are not independent. If we define $\varphi^*(t) = t\varphi(\frac{1}{t})$ then $\mathcal{D}_{\varphi^*}(r, q) = \mathcal{D}_\varphi(q, r)$ [22]. For $\varphi(t) = t \log t$, we have $\varphi^*(t) = -\log t$.

The objective of importance sampling is to make probability density $p(\alpha, x)$ similar to scaled integrand $g(x) = f(x)/\mu$, i.e., to minimize φ -divergence $\mathcal{D}_\varphi(g, p(\alpha))$. A φ -divergence is a convex function in both arguments and $p(\alpha)$ is a linear function of α , thus the functional is convex in α . This will ensure the existence of a minimum. With choosing $\varphi(t) = t^2 - 1$, we can get the variance back as a special case of this more general approach:

$$V[\mathcal{F}] = \int \frac{f^2(x)}{p(\alpha, x)} dx - \mu^2 = \mu^2 \int p(\alpha, x) \left(\left(\frac{g(x)}{p(\alpha, x)} \right)^2 - 1 \right) dx = \mu^2 \mathcal{D}_\varphi(g(x), p(\alpha, x)). \tag{13}$$

The optimal MIS problem is reduced to find the minimum of $\mathcal{D}_\varphi(g, p(\alpha))$ with the constraints $\sum_{i=1}^m \alpha_i = 1$ and $\alpha_i \geq 0$. Using Lagrange multipliers, the solution is

$$\frac{\partial \mathcal{D}_\varphi(g, p(\alpha))}{\partial \alpha_i} - \lambda = 0, \tag{14}$$

which means that the derivatives of the divergence with respect to fractions α_i must be the same for each sampling method i .

The derivative can be expressed as

$$\frac{\partial \mathcal{D}_\varphi(g, p(\alpha))}{\partial \alpha_i} = \int p_i(x) \left[\varphi\left(\frac{g(x)}{p(\alpha, x)}\right) - \frac{g(x)}{p(\alpha, x)} \varphi'\left(\frac{g(x)}{p(\alpha, x)}\right) \right] dx. \tag{15}$$

This integral is the expectation when samples are generated with pdf $p_i(x)$, thus we can write

$$E_{p_i} \left[\varphi \left(\frac{g}{p(\alpha)} \right) - \frac{g}{p(\alpha)} \varphi' \left(\frac{g}{p(\alpha)} \right) \right] = E_{p(\alpha)} \left[\varphi \left(\frac{g}{p(\alpha)} \right) - \frac{g}{p(\alpha)} \varphi' \left(\frac{g}{p(\alpha)} \right) \right] = \lambda. \tag{16}$$

Equation (16) only holds when $\alpha_i > 0$, i.e., in the interior of the simplex domain, not on the border.

Swapping the two distributions in the φ -divergence, we have another measure for the dissimilarity between $p(\alpha)$ and g , $\mathcal{D}_\varphi(p(\alpha), g)$. Using Lagrange multipliers again, this leads to the following equation for unknown fractions α_i

$$\frac{\partial \mathcal{D}_\varphi(p(\alpha), g)}{\partial \alpha_i} = \int p_i \varphi' \left(\frac{p(\alpha)}{g} \right) dx = \lambda^*. \tag{17}$$

Using expected values, this option is expressed as

$$E_{p_i} \left[\varphi' \left(\frac{p(\alpha)}{g} \right) \right] = E_{p(\alpha)} \left[\varphi' \left(\frac{p(\alpha)}{g} \right) \right] = \lambda \tag{18}$$

The problem with the optimization of the variance was that in the resulting equation the ratios of the sampled integrand and the pdf are squared (see Equation (10)), making the equation sensitive to non-optimal pdfs. With the proper definition of φ , we can solve this issue. For example, if $\varphi(t) = -\log t$, then $\varphi' = -1/t$, and the condition of optimality is

$$E_{p_i} \left[\frac{g}{p(\alpha)} \right] = E_p \left[\frac{g}{p(\alpha)} \right] = -\lambda^*. \tag{19}$$

Remembering that $g(x) = f(x)/\mu$, the last equation can also be written as

$$E_{p_i} \left[\frac{f}{p(\alpha)} \right] = E_p \left[\frac{f}{p(\alpha)} \right], \tag{20}$$

which contains the ratios without squaring. This gives back the solution in [23], which is also conjectured in [16]. To further exploit this idea, we take the parametric family of Tsallis divergences. With its parameter the compromise between the robustness of the optimization equation and the similarity to the variance can be controlled.

5.2. Tsallis Divergences

Tsallis divergence [24] is associated with the Tsallis entropy [25]. Let us consider the one-parameter Tsallis divergence between distributions $q(x)$ and $r(x)$:

$$\begin{aligned} \mathcal{D}_\gamma^T(q, r) &= \frac{1}{\gamma - 1} \left(\int \frac{q^\gamma(x)}{r^{\gamma-1}(x)} dx - 1 \right) \\ &= \frac{1}{\gamma - 1} \int \left(\frac{q^\gamma(x)}{r^{\gamma-1}(x)} - r(x) \right) dx = \frac{1}{\gamma - 1} \int r(x) \left(\frac{q^\gamma(x)}{r^\gamma(x)} - 1 \right) dx. \end{aligned} \tag{21}$$

Tsallis divergence $\mathcal{D}_\gamma^T(q, r)$ is a φ -divergence defined by

$$\varphi(t) = \frac{t^\gamma - 1}{\gamma - 1}.$$

Parameter γ should satisfy $\gamma > 0$ and $\gamma \neq 1$, which guarantees that $\varphi(t)$ exists and is convex.

It can be shown that $\lim_{\gamma \rightarrow 1} \mathcal{D}_\gamma^T(q, r) = KL(q, r)$ is the Kullback–Leibler divergence. When $\gamma = 2$, Tsallis divergence turns to be the chi-square divergence:

$$\chi^2(q, r) = \int \frac{(q(x) - r(x))^2}{r(x)} dx \tag{22}$$

$$= \int \frac{q^2(x)}{r(x)} dx - 2 \int q(x) dx + \int r(x) dx \tag{23}$$

$$= \int \frac{q^2(x)}{r(x)} dx - 1 = \mathcal{D}_2^T(q, r).$$

For $\gamma = 1/2$, Tsallis divergence $\mathcal{D}_{1/2}^T(q, r)$ becomes the squared Hellinger distance $H^2(q, r)$:

$$\mathcal{D}_{1/2}^T(q, r) = -2 \int (q(x)^{1/2} r(x)^{1/2} - q(x)) dx \tag{24}$$

$$= 2 \left(1 - \int q(x)^{1/2} r(x)^{1/2} dx \right) = 2H^2(q, r).$$

Note that $H(q, r)$ is a true metric.

5.3. Optimal MIS Solution with Tsallis Divergence

Optimal MIS needs to find fractions α_i that minimize $\mathcal{D}_\gamma^T(g(x), p(\alpha, x))$ with the constraint $\sum_{i=1}^m \alpha_i = 1$. Substituting function $\varphi(t)$ of Tsallis divergence into Equation (16), we obtain that for all i , the quantities, called γ -moments,

$$M_{\gamma,i} = \int \left(\frac{f(x)}{p(\alpha, x)} \right)^\gamma p_i(x) dx = E_{p_i}[\mathcal{F}^\gamma], \tag{25}$$

have to be equal. Equation (25) guarantees a global minimum, as the function $\mathcal{D}_\gamma^T(g(x), p(\alpha, x))$ is convex for all $\gamma > 0$. Considering the case when $g(x)$ and $p(\alpha, x)$ are swapped in the divergence, i.e., substituting function $\varphi(t)$ of Tsallis into Equation (17), the requirement of the equality of the γ -moments is established again, but now γ should be replaced by $1 - \gamma$. Examining Equation (20) we can realize that optimizing for $\gamma = 1$ is the same as minimizing the cross entropy or the Kullback–Leibler divergence. Thus, the criterion of the equality of the γ -moments covers all cases.

5.4. Other φ -Divergences: Total Variation Divergence, χ_P^k and $|\chi|_P^k$ Divergences

The total variation can also be interpreted as a φ -divergence defined by $\varphi(t) = \frac{1}{2}|t - 1|$:

$$TV(q, r) = \frac{1}{2} \int |q(x) - r(x)| dx$$

This is the only φ -divergence that is a true metric [26]. The condition of optimality is

$$\int \text{sign}(g(x) - p(\alpha, x)) p_i(x) dx = \int \text{sign}(g(x) - p(\alpha, x)) p(\alpha, x) dx \tag{26}$$

for all i , although we can hardly use Equation (26) to approximate a solution. φ -divergences also include the Pearson–Vajda χ_P^k and $|\chi|_P^k$ divergences:

$$\chi_P^k(q, r) = \int \frac{(q(x) - r(x))^k}{q^{k-1}(x)} dx, \quad |\chi|_P^k(q, r) = \int \frac{|q(x) - r(x)|^k}{q^{k-1}(x)} dx$$

defined by the functions $\varphi(t) = (t - 1)^k$ and $\varphi(t) = |t - 1|^k$, respectively [17]. For $k = 2$, we have $\chi_P^2(q, r) = |\chi|_P^2(q, r) = \chi^2(r, q)$, and for $k = 1$, identities $\chi_P^1(q, r) = 0$ and $|\chi|_P^1(q, r) = 2TV(q, r)$ hold.

6. Finding the Optimal Solution

For the sake of simplicity, we assume that two sampling methods of pdfs $p_1(x)$ and $p_2(x)$, associated with fractions α and $1 - \alpha$ are combined. The mixture density is

$$p(\alpha, x) = \alpha p_1(x) + (1 - \alpha) p_2(x).$$

From the optimality condition of Equation (25), we obtain

$$\int \left(\frac{f(x)}{p(\alpha, x)} \right)^\gamma (p_1(x) - p_2(x)) dx = \zeta(\alpha) = 0. \tag{27}$$

This is a non-linear equation for unknown fraction α , which is solved with Newton–Raphson iteration. The derivative of $\zeta(\alpha)$ is

$$\zeta'(\alpha) = \frac{\partial \zeta(\alpha)}{\partial \alpha} = -\gamma \int \frac{f^\gamma(x)}{p^{\gamma+1}(\alpha, x)} (p_1(x) - p_2(x))^2 dx, \tag{28}$$

as $\partial p(\alpha, x) / \partial \alpha = p_1(x) - p_2(x)$.

The value of $\zeta(\alpha)$ and its derivative can be estimated in the following way. Let us re-write $\zeta(\alpha)$ as an expectation:

$$\zeta(\alpha) = \int \frac{f^\gamma(x)}{p^{\gamma+1}(\alpha, x)} (p_1(x) - p_2(x)) p(\alpha, x) dx = E_{p(\alpha)} \left[\frac{f^\gamma(X)}{p^{\gamma+1}(\alpha, X)} (p_1(X) - p_2(X)) \right]. \tag{29}$$

Taking N samples $\{X_1, X_2, \dots, X_N\}$ according to pdf $p(\alpha, x)$, we have the estimator

$$\widehat{\zeta(\alpha)} = \frac{1}{N} \sum_{i=1}^N \frac{f^\gamma(X_i)}{p^{\gamma+1}(\alpha, X_i)} (p_1(X_i) - p_2(X_i)), \tag{30}$$

and in the same way,

$$\widehat{\zeta'(\alpha)} = -\frac{\gamma}{N} \sum_{i=1}^N \frac{f^\gamma(X_i)}{p^{\gamma+2}(\alpha, X_i)} (p_1(X_i) - p_2(X_i))^2. \tag{31}$$

The integrals are estimated with an iterative algorithm updating α after each iteration. Starting with $\alpha^{(0)} = 1/2$, in iteration k we draw N samples according to $p(\alpha^{(k-1)}, x)$, compute $\widehat{\zeta(\alpha^{(k-1)})}$ and $\widehat{\zeta'(\alpha^{(k-1)})}$, and use the Newton–Raphson formula to obtain updated fraction $\alpha^{(k)}$:

$$\alpha^{(k)} = \alpha^{(k-1)} - \frac{\widehat{\zeta(\alpha^{(k-1)})}}{\widehat{\zeta'(\alpha^{(k-1)})}}. \tag{32}$$

7. Numerical Examples

We present here three 1D examples. We used the Mathematica package for the computations.

7.1. Example 1

Suppose we want to evaluate the integral (see Figure 1)

$$\mu = \int_{0.01}^{3.5\pi} (\sqrt{x} + \sin x) dx \approx 25.3065 \tag{33}$$

by MIS combining sampling methods of pdfs $Gauss(2, 1)$ and $Gauss(8, 2)$, where $Gauss(m, \sigma)$ stands for the normal distribution of mean m and standard deviation σ . For this example, equal sample number MIS has variance $V[F] = 24.1152$, and the minimum variance is $V[F] = 13.4788$.

In Figure 2, we show the variances $V[F]$ and $V[\mathcal{F}]$ for the optimal solution of Tsallis divergences from $\gamma = 0.1$ to $\gamma = 8$ in steps of 0.1. Dots depict the values of $V[F]$ for the optimal α fractions for each parameter γ using 5 Newton–Raphson iterations with 100 total samples in each iteration.

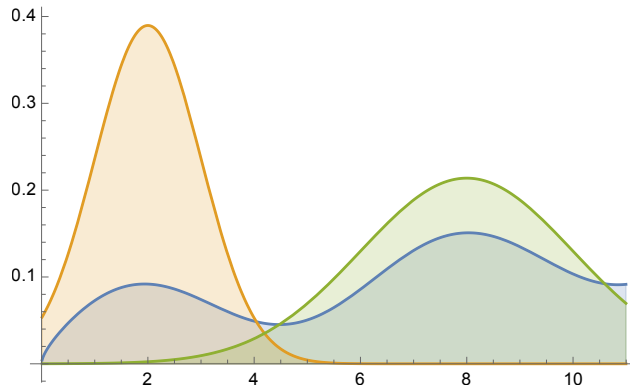


Figure 1. Example 1: $f(x)/\mu$ (in blue) superimposed on the two pdfs used for MIS integration.

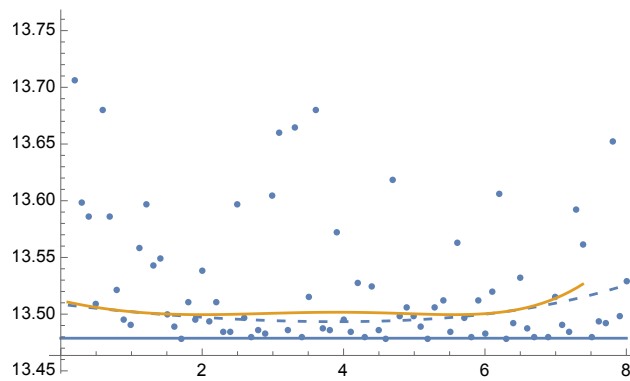


Figure 2. Example 1: Values of $V[F]$ (dashed line) and $V[\mathcal{F}]$ (continuous line) for the solution of Equation (25) for each value of parameter γ from 0.1 to 8, in the horizontal axis. Variance $V[\mathcal{F}]$ (continuous line) is minimal when $\gamma = 2$, as $V[\mathcal{F}]$ corresponds to χ^2 divergence (Equation (13)). Dots show the variance $V[F]$ for the optimal α values for each parameter γ using 5 Newton–Raphson iterations with 100 total samples in each iteration. Horizontal line corresponds to the minimum variance $V[F] = 13.4788$. A zoom-out is shown in Figure 3.

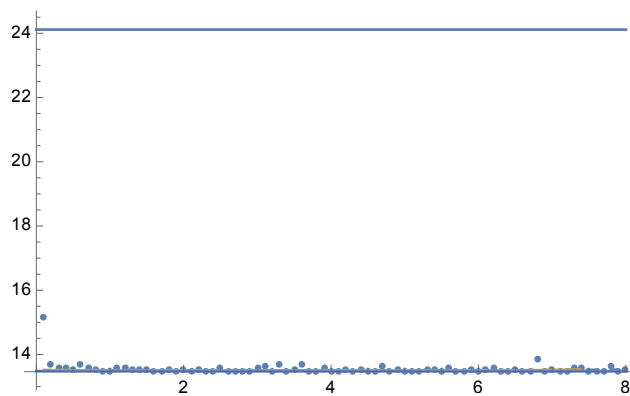


Figure 3. Example 1. Zoom-out of Figure 2. Observe that $V[\mathcal{F}] < V[F]$ except for $\gamma = 1$ where $V[\mathcal{F}] = V[F]$. For the equal sample budget case, $V[F] = 24.1152$, while the minimum is $V[F] = 13.4788$.

7.2. Example 2

Let us consider integral (see Figure 4)

$$\mu = \int_{-4}^4 \text{Gauss}(-1.5, 1) + 2\text{Gauss}(1.5, 0.75) dx \approx 2.9929 \tag{34}$$

by MIS using functions $\text{Gauss}(-1.5, 1)$ and $\text{Gauss}(1.5, 0.75)$. For this example, equal sample number MIS has a variance of $V[F] = 0.1134$, and the minimum variance is $V[F] = 0$. Figures 5 and 6 present the optimal solutions and the results after 5 Newton–Raphson iterations with 100 total samples in each iteration.

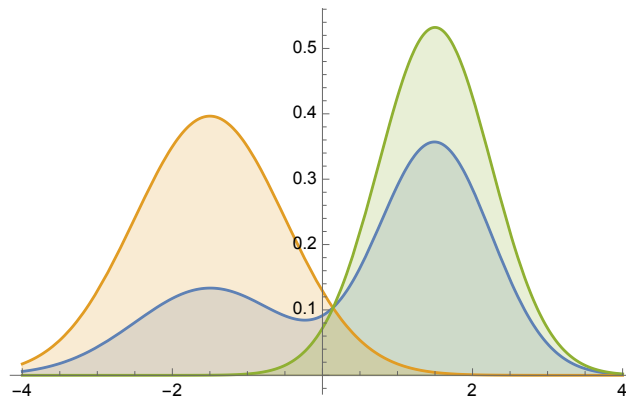


Figure 4. Example 2: $f(x)/\mu$ (in blue) superimposed on the two pdfs used for MIS integration.

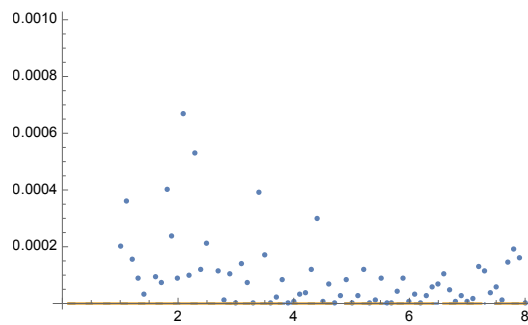


Figure 5. Example 2: $V[F]$ values corresponding to the Newton–Raphson solutions for the optimal α values (vertical axis), from $\gamma = 0.1$ to $\gamma = 8$ in steps of 0.1 (horizontal axis) for each parameter γ . Dots show the variance $V[F]$ for the optimal α values for each parameter γ using 5 Newton–Raphson iterations with 100 total samples in each iteration. Minimum variance is $V[F] = 0$. A zoom-out is shown in Figure 6.

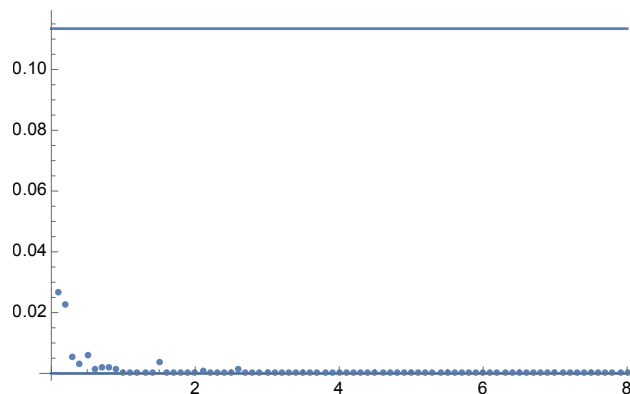


Figure 6. Example 2: Zoom-out of Figure 5. The minimum of $V[F]$ and $V[\mathcal{F}]$ is 0. For equal sample budget MIS, the variance is $V[F] = 0.1134$.

7.3. Example 3

Finally, consider the approximation of the following integral (see Figure 7)

$$\mu = \int_{0.01}^{\pi/2} (\sqrt{x} + \sin x) dx \approx 2.3118. \tag{35}$$

by MIS using functions $2 - x$, and $\sin^2(x)$. For this example, equal sample budget MIS has a variance of $V[F] = 0.2772$, and the minimum variance is $V[F] = 0.09032$. In Figures 8 and 9 we show the variances $V[F]$ and $V[\mathcal{F}]$ for the optimal solution for parameters from $\gamma = 0.1$ to $\gamma = 8$. Dots are the results of $V[F]$ using 5 Newton–Raphson iterations with 100 total samples in each iteration.

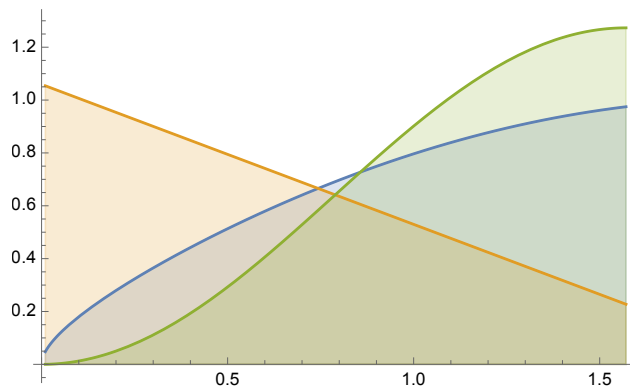


Figure 7. Example 3: $f(x)/\mu$ (in blue) superimposed on the two pdfs used for MIS integration.

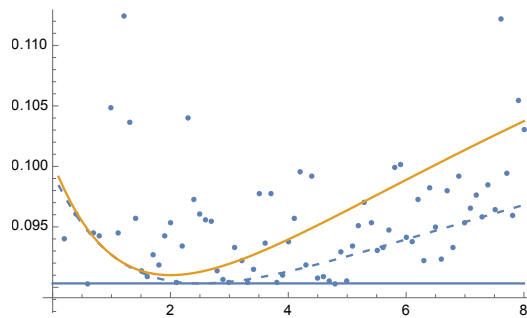


Figure 8. Example 3: $V[F]$ (dashed line) and $V[\mathcal{F}]$ (continuous line), for the solution of Equation (25) for each value of the parameter γ between 0.1 and 8, in the horizontal axis. Variance $V[\mathcal{F}]$ (continuous line) is minimal when $\gamma = 2$, as $V[\mathcal{F}]$ corresponds to χ^2 divergence (Equation (13)). Dots are the $V[F]$ values after 5 Newton–Raphson iterations taking 100 samples in total at each iteration. Horizontal line corresponds to minimum variance, $V[F] = 0.09032$. A zoom-out is shown in Figure 9.

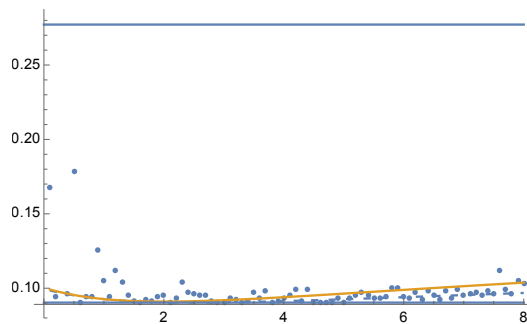


Figure 9. Example 3: Zoom-out of Figure 8. For equal sample budget $V[F] = 0.2772$, and minimum is $V[F] = 0.09032$.

8. Combination of Light Source Sampling and BRDF Sampling in Computer Graphics

Here we consider a classical problem of computer graphics, the combination of light source sampling and BRDF (Bidirectional Reflectance Distribution Function) sampling. To compute the reflected radiance $L^r(\vec{p}, \vec{\omega})$ of a surface point \vec{p} at direction $\vec{\omega}$, the following integral should be evaluated in the hemispherical domain of incident directions Ω :

$$L^r(\vec{p}, \vec{\omega}) = \int_{\Omega} L(\vec{p}', \vec{\omega}') f_r(\vec{\omega}', \vec{p}, \vec{\omega}) \cos \theta' d\omega' \quad (36)$$

where $L(\vec{p}', \vec{\omega}')$ is the radiance of point \vec{p}' visible from point \vec{p} in incident direction $-\vec{\omega}'$, $f_r(\vec{\omega}', \vec{p}, \vec{\omega})$ is the BRDF expressing the portion of the light beam that is reflected from direction $\vec{\omega}'$ to $\vec{\omega}$ at point \vec{p} and θ' is the angle between the surface normal at point \vec{p} and incident direction $-\vec{\omega}'$. While solving such problems, we have two sampling methods, $p_1(\vec{\omega}')$ approximately mimicking the $f_r(\vec{\omega}', \vec{p}, \vec{\omega}) \cos \theta'$ factor and $p_2(\vec{\omega}')$ mimicking the incident radiance $L(\vec{p}', \vec{\omega}')$.

MIS would use a combined pdf in the following form

$$p(\alpha, \vec{\omega}') = \alpha p_1(\vec{\omega}') + (1 - \alpha) p_2(\vec{\omega}'). \quad (37)$$

where the task is the optimal determination of weight α .

In order to test the proposed method, we render the classic scene of Veach [1] with combined light source and BRDF sampling. The shiny rectangles have max-Phong BRDF [27] with shininess parameters 50, 100, 500, and 1000, respectively. The four spherical light sources emit the same power.

For each pixel, we used 50 samples in total organized in 5 iterations. The process starts with 5 BRDF and 5 light source samples per pixel, and the per-pixel α weights are updated at the end of each iteration. Figure 10 shows the rendered images together with the α -maps, and we compare the original sampling techniques, equal count MIS, variance minimization, and Tsallis divergence. The reference obtained by high number of samples is shown by Figure 11. Table 1 depicts the Root Mean Square Error (RMSE) values. Figure 12 shows the convergence plots.

Finally, we note that φ -divergences were used for global illumination as oracles for adaptive sampling in [28].

Table 1. RMSE of different MIS methods.

Method	RMSE
BRDF sampling	613
Light source sampling	295
Equal count MIS	244
Tsallis ($\gamma = 0.01$)	216
Tsallis ($\gamma = 0.1$)	213
Tsallis ($\gamma = 0.5$) (Hellinger)	215
Tsallis ($\gamma = 1$) (Kullback-Leibler)	218
Tsallis ($\gamma = 1.5$)	218
Tsallis ($\gamma = 2$) (chi-square, variance)	221
Tsallis ($\gamma = 2.5$)	254

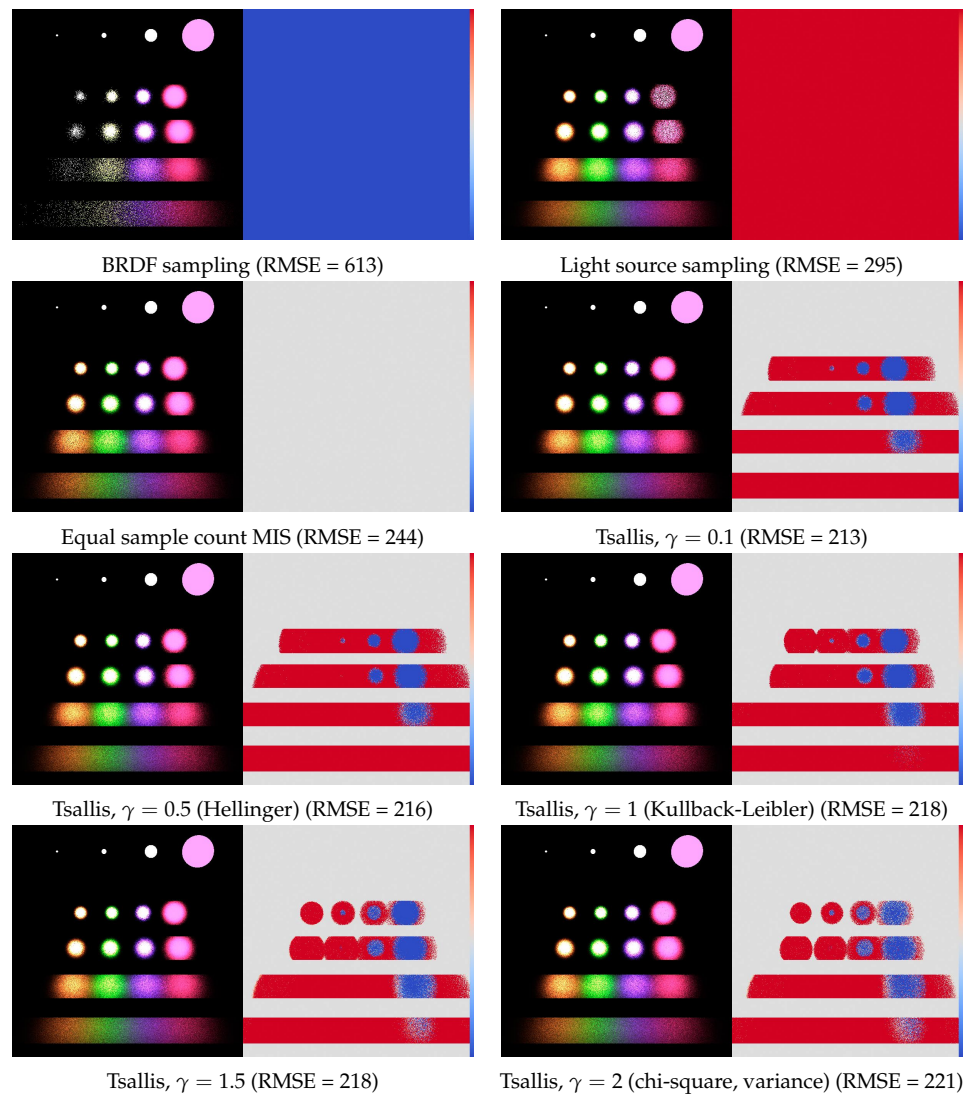


Figure 10. Comparison of MIS weighting schemes for the direct lighting problem of computer graphics. The left part is the image rendered with 100 rays per pixel, the right part is weight α of the light source sampling. The RMSE is computed as the average of 30 independent executions. The (0,1) interval of possible α values is visualized by the color bar.

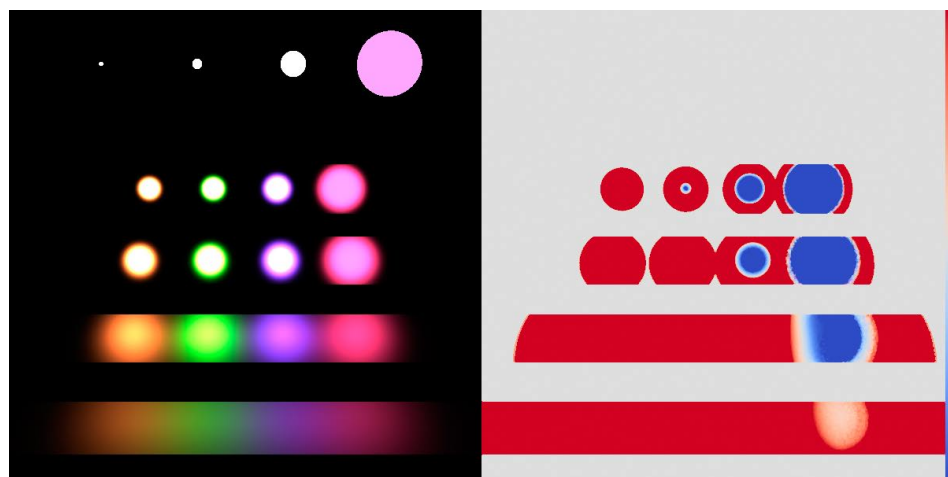


Figure 11. Reference image and α -map obtained with 50,000 samples per pixel organized in 5 iterations.

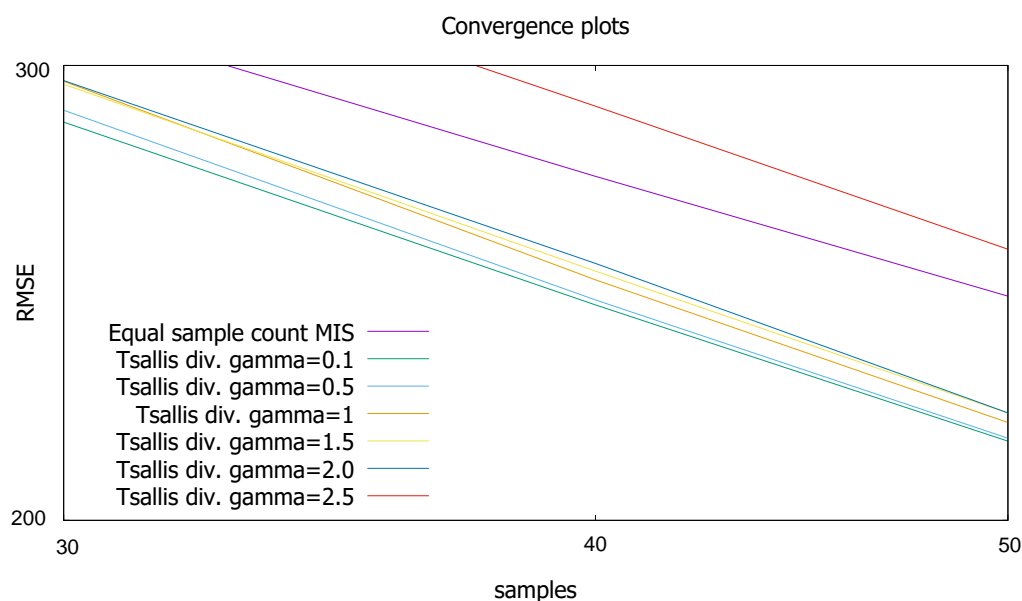


Figure 12. RMSE as functions of the number of samples per pixel.

9. Discussion

The goal of adaptive MIS is to find the weights of given pdfs that eventually lead to minimal integration error. The definition of the error may be application dependent, in this paper we chose the RMSE, which describes the variance of the estimator. We argued that instead of the direct optimization of the estimated variance, it is better to use other measures for the optimization target since the variance estimation may strongly fluctuate if the sample number is moderate, which prevents the adaptation process from converging to the real optimum. The alternative criterion is the Tsallis divergence, which includes important particular divergence types and by setting its parameter, more robust targets can be defined.

In order to demonstrate the proposed method, we took three numerical examples and a fundamental problem of computer graphics. We can observe in all examples that the adaptive MIS can significantly outperform the equal sample count MIS, thus examining adaptation strategies has theoretical and practical relevance. In the first numerical example, none of the combinations of the two sampling pdfs could well mimic the integrand, thus the computation errors are quite high and the influence of parameter γ is small. In the second numerical example, the integrand is proportional to an appropriate linear combination of the two sampling pdfs, thus zero variance estimator is possible. If parameter γ is not too small, the method could indeed compute the integral with high accuracy. This means that in easy integration problems, Tsallis divergence can be safely used. The third numerical example is a more difficult, realistic case when sufficiently high number of samples are used in each iteration step to estimate the optimization target. As expected, the optimal value of parameter γ is close to 2 since this is the point where Tsallis divergence becomes the variance $V[\mathcal{F}]$.

In the image synthesis example, the integrand is not a weighted sum but a product where the sampling pdfs approximately mimic the factors. In this case, no zero variance MIS is possible. Unlike in the numerical examples, we took small number of samples for each pixel according to practical scenarios. The small sample number made the optimization target equation unreliable unless parameter γ is reduced from 2 that corresponds to the variance. Looking at Figure 10, we can observe that when γ becomes smaller, the computed weight maps become slightly less accurate in average but also less noisy, which eventually reduces the computation error.

We have also seen in the examples that the differences between the variances corresponding to the optimal values for the different parameters γ of the Tsallis divergence are

small. This is in accordance with any φ -divergence being a second order approximation of χ^2 divergence [17], thus obtaining any minimum within a wide range of γ parameters would be a good solution. Aside from $\gamma = 2$, which corresponds to $V[\mathcal{F}]$, the most interesting Tsallis divergence from a practical perspective is the Kullback-Leibler divergence, corresponding to $\gamma = 1$.

10. Conclusions and Future Work

We have shown in this paper that finding the optimal weights for MIS can be presented as finding the minimum φ -divergence between the normalized integrand and the importance pdf. Being convex, a φ -divergence allows convex optimization. We have singled out the Tsallis divergences with parameter γ as they have the further advantage that the equations for the optimal value consist of making the γ -moments equal. We have tested our results with numerical examples and solving an illumination problem.

In our future work we will investigate the optimal efficiencies, i.e., taking into account the cost of sampling. This has been done for the variance $V[\mathcal{F}]$ (i.e., $\gamma = 2$) in [4,6] where the implicit equation for the optimal solution was given. However, the efficiency equation does not allow convex optimization. Using the γ -moments, or a different family of φ -divergences, could help in obtaining a robust solution for the efficiency.

Author Contributions: Conceptualization, M.S.; Investigation, M.S. and L.S.-K.; Software, M.S. and L.S.-K.; Validation, M.S. and L.S.-K.; Visualization, M.S. and L.S.-K.; Writing—original draft, M.S.; Writing—review & editing, M.S. and L.S.-K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by OTKA K-124124 and by Grant PID2019-106426RB-C31 funded by MCIN/AEI/10.13039/501100011033.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 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, Los Angeles, CA, USA, 6–11 August 1995; pp. 419–428.
2. Veach, E. Robust Monte Carlo Methods for Light Transport Simulation. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 1997.
3. Sbert, M.; Havran, V.; Szirmay-Kalos, L. Variance analysis of multi-sample and one-sample multiple importance sampling. *Comput. Graph. Forum* **2016**, *35*, 451–460.
4. Sbert, M.; Havran, V. Adaptive multiple importance sampling for general functions. *Vis. Comput.* **2017**, *33*, 845–855.
5. Sbert, M.; Havran, V.; Szirmay-Kalos, L. Multiple importance sampling revisited: Breaking the bounds. *EURASIP J. Adv. Signal Process* **2018**, *2018*, 15.
6. 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, VRCAI '14, Shenzhen, China, 30 November–2 December 2014; ACM: New York, NY, USA, 2014; pp. 141–150.
7. Lu, H.; Pacanowski, R.; Granier, X. Second-Order Approximation for Variance Reduction in Multiple Importance Sampling. *Comput. Graph. Forum* **2013**, *32*, 131–136.
8. Elvira, V.; Martino, L.; Luengo, D.; Bugallo, M.F. Efficient multiple importance sampling estimators. *IEEE Signal Process. Lett.* **2015**, *10*, 1757–1761.
9. Elvira, V.; Martino, L.; Luengo, D.; Bugallo, M.F. Generalized Multiple Importance Sampling. *Stat. Sci.* **2019**, *34*, 129–155.
10. Sbert, M.; Havran, V.; Szirmay-Kalos L.; Elvira V. Multiple importance sampling characterization by weighted mean invariance. *Vis. Comput.* **2018**, *34*, 843–852.
11. Sbert, M.; Havran, V.; Szirmay-Kalos L. Optimal Deterministic Mixture Sampling. In Proceedings of the Short Papers. Eurographics 2019, Genova, Italy, 6–10 May 2019; Eurographics Association: Aire-La-Ville, Switzerland, 2019; pp. 73–76.
12. Vorba, J.; Hanika, J.; Herholz, S.; Mueller, T.; Krivanek, J.; Keller, A. Path guiding in production. In Proceedings of the ACM SIGGRAPH '19: Special Interest Group on Computer Graphics and Interactive Techniques Conference, Los Angeles, CA, USA, 28 July 2019.
13. Cornebise, J.; Moulines, E.; Olsson, J. Adaptive methods for sequential importance sampling with application to state space models. *Stat. Comput.* **2008**, *18*, 461–480.

14. Míguez, J. On the performance of nonlinear importance samplers and population Monte Carlo schemes. In Proceedings of the 22nd International Conference on Digital Signal Processing (DSP), London, UK, 23–25 August 2017; pp. 1–5.
15. Müller, T.; McWilliams, B.; Rousselle, F.; Gross, M.; Novák, J. Neural importance sampling. *ACM Trans. Graph.* **2019**, *38*, 1–19.
16. Sbert, M.; Elvira, V. Generalizing the balance heuristic estimator in multiple importance sampling. *Entropy* **2022**, *24*, 191.
17. Nielsen, F.; Nock, R. On the chi square and higher-order chi distances for approximating f-divergences. *IEEE Signal Process. Lett.* **2014**, *21*, 10–13.
18. Murray, D.; Benzait, S.; Pacanowski, R.; Granier, X. On Learning the Best Local Balancing Strategy. In Proceedings of the Eurographics 2020—Short Papers, Norrköping, Sweden, 25–29 May 2020; Eurographics Association: Aire-La-Ville, Switzerland, 2020.
19. Kondapaneni, I.; Vevoda, P.; Grittmann, P.; Skřivan, T.; Slusallek, P.; Krivánek, J. Optimal multiple importance sampling. *ACM Trans. Graph.* **2019**, *38*, 1–14.
20. Csiszár, I. Information-type measures of difference of probability distributions and indirect observation. *Stud. Sci. Math. Hung.* **1967**, *2*, 229–318.
21. Wikipedia, F-Divergence—Wikipedia, The Free Encyclopedia, Wikipedia, The Free Encyclopedia. Available online: <http://en.wikipedia.org/w/index.php?title=F-divergence&oldid=1068442350> (accessed on 5 April 2022).
22. Igal, S. On f-Divergences: Integral Representations, Local Behavior, and Inequalities. *Entropy* **2018**, *20*, 383.
23. Szirmay-Kalos, L.; Sbert, M. Robust sample budget allocation for MIS. In Proceedings of the Eurographics 2022—Short Papers, Reims, France, 25–29 April 2022; Eurographics Association: Aire-La-Ville, Switzerland, 2022.
24. Nielsen, F.; Nock, R. A closed-form expression for the Sharma–Mittal entropy of exponential families. *J. Phys. Math. Theor.* **2011**, *45*, 032003.
25. Wikipedia, Tsallis Entropy—Wikipedia, The Free Encyclopedia, Wikipedia, The Free Encyclopedia. Available online: <http://en.wikipedia.org/w/index.php?title=Tsallis%20entropy&oldid=1072173071> (accessed on 5 April 2022).
26. Vajda, I. On Metric Divergences of Probability Measures. *Kybernetika* **2009**, *45*, 885–900.
27. Neumann, L.; Neumann, A.; Szirmay-Kalos, L. Compact metallic reflectance models. *Comput. Graph. Forum* **1999**, *18*, 161–172.
28. Rigau, J.; Feixas, M.; Sbert, M. Refinement Criteria Based on f-Divergences. In Proceedings of the EGRW '03: Proceedings of the 14th Eurographics workshop on Rendering Leuven Belgium, 25–27 June 2003; Dutre, P., Suykens, F., Christensen, P.H., Cohen-Or, D., Eds.; Eurographics Association: Aire-La-Ville, Switzerland, 2003.