Volume xx (200y), Number z, pp. 1–4

Supplementary material A: Derivations of results

EGSR 2014 Submission ID 1034

In this document, we present detailed derivations for the results presented in the paper (boxed equations here).

1. Analysis of importance sampling

1.1. Exact variance of importance sampling

The primary importance sampling estimator is $\tilde{\mu}_{\text{is},1} = f(\mathbf{X})/g(\mathbf{X}), \quad \mathbf{X} \sim g(x)$ and the corresponding secondary estimator is $\tilde{\mu}_{\text{is,N}} = \frac{1}{N}\sum f(\mathbf{X}_i)/g(\mathbf{X}_i), \quad \mathbf{X}_i \sim g(x)$. These estimators are unbiased and the extent of variance reduction depends critically on judicious choice of the importance function. Writing $\alpha(x) = 1/g(x)$, the definition of variance of the primary importance sampling estimator is

$$\boxed{\mathrm{V}\left(\tilde{\mu}_{\text{is},1}\right) \quad = \quad \int\limits_{\mathcal{D}} f^2(x)\alpha(x) \ \mathrm{d}x \ - \ I^2} \tag{1}$$

The Fourier transform preserves inner products, and since the Fourier transform of the rhs of eq. 1 could be written in two ways:

$$\mathrm{V}\left(\tilde{\mu}_{\text{is},1}\right) \ = \ \int \left(\hat{f}_\Pi \otimes \hat{f}_\Pi\right)(\omega) \ \ \hat{\alpha}(\omega) \, \mathrm{d}\omega \ - \ \left(\hat{f}_\Pi(0)\right)^2 \tag{2}$$

$$\mathrm{V}\left(\tilde{\mu}_{\text{is},1}\right) \ = \ \int \left(\hat{f}_\Pi \otimes \hat{\alpha}\right)(\omega) \ \ \hat{f}_\Pi(\omega) \, \mathrm{d}\omega \ - \ \left(\hat{f}_\Pi(0)\right)^2 \tag{3}$$

respectively. In either case, choosing $\hat{\alpha}(\omega) = \delta(\omega)$ results in a variance proportional to that of the integrand. These equations provide different insight into the choice of importance function for variance reduction. Eq. 2 suggests that ideally, $g(x) = 1/\alpha(x)$ should be chosen so that $\hat{\alpha}(\omega)$ contains all its energy at frequencies where the square of the integrand has no energy. Eq. 3, on the other hand implies that the ideal situation is when $(\hat{f}_\Pi \otimes \hat{\alpha})(\omega) = \delta(\omega)$, or that the weighting function warps the integrand into a constant.

Discussion: The main shortcoming of this analysis is its dependence on the phase of the integrand for exact prediction of variance. This is not surprising, since the variance due to IS is not shift-invariant. In fact, the benefit of IS hinges on deliberate correlation between the importance function and integrand (which translates to phase in the Fourier analysis). Although this equation is not of practical value unless some information is available about the phase of the integrand, we derive an upper bound that is easy to evaluate.

1.2. Multiple importance sampling

A common strategy in rendering, when the integrand is a product of two or more terms, is multiple importance sampling (MIS). Rather than sampling the product, MIS samples from the average of the distributions. We show that, while this is mathematically equivalent to sampling from the average, the two strategies are statistically different. estimator to that of an importance sampling estimator which uses the averaged importance functions. First we will compare the variance of the MIS estimator with that of the average distribution IS estimator, using just two importance functions. Later, we provide the general result.

The integral I could be expressed mathematically as either

$$\int\limits_{\mathcal{D}} f(x) \ \mathrm{d}x \ = \ \int\limits_{\mathcal{D}} \frac{f(x)}{\bar{g}(x)}\bar{g}(x) \ \mathrm{d}x \quad \text{or as} \tag{4}$$

$$\int\limits_{\mathcal{D}} f(x) \ \mathrm{d}x \ = \ \int\limits_{\mathcal{D}} \frac{f(x)}{2\bar{g}(x)}g_1(x) \ \mathrm{d}x + \int\limits_{\mathcal{D}} \frac{f(x)}{2\bar{g}(x)}g_2(x) \ \mathrm{d}x.$$

where $\bar{g}(x) = (g_1(x) + g_2(x))/2$. While both expressions are mathematically equal, they lead to unbiased estimators that have different statistical behaviour (variances). The first, $\tilde{\mu}_{\text{avg,is}}$, simply importance samples the averaged pdf $\bar{g}(x)$. The second, $\tilde{\mu}_{\text{mis}}$, samples independently from $g_1(x)$ and $g_2(x)$ and combines them, as though they were drawn from the average distribution.

Estimating each of the two integrals on the rhs using $N/2$ samples yields the N-sample MIS estimator (with balance heuristic)

$$\tilde{\mu}_{\mathrm{mis,N}} = \frac{2}{N}\left(\sum_{i=1}^{N/2}\frac{f(\mathbf{X}_{1i})}{2\bar{g}(\mathbf{X}_{1i})} + \sum_{i=1}^{N/2}\frac{f(\mathbf{X}_{2i})}{2\bar{g}(\mathbf{X}_{2i})}\right), \quad (5)$$

where $\mathbf{X}_{1i} \sim g_1(x)$ and $\mathbf{X}_{2i} \sim g_2(x)$. The main distinction of this estimator from the two-sample importance sampling estimator (from the average distribution $\bar{g}(x)$) is that in the latter $\mathbf{X}_1$ and $\mathbf{X}_2$ are both drawn according to $\bar{g}(x)$. The variance of the MIS estimator, from first principles, is

$$\begin{aligned}\mathrm{V}\left(\tilde{\mu}_{\mathrm{mis,N}}\right) &= \frac{1}{2N}\left(2\left\langle\frac{f^2(x)}{\bar{g}(x)^2}\right\rangle_{\bar{g}} - 4I^2 + 2I_{\bar{g}1}I_{\bar{g}2}\right) \\ &= \mathrm{V}\left(\tilde{\mu}_{\mathrm{avg,isN}}\right) + 2\frac{I_{\bar{g}1}I_{\bar{g}2} - I^2}{2N} \quad (6)\end{aligned}$$

where $I_{\bar{g}i} = \left\langle\frac{f(x)}{\bar{g}(x)}\right\rangle_{g_i}$. Assume that $I_{\bar{g}1}I_{\bar{g}2} - I^2 > 0$. Since $2I = I_{\bar{g}1} + I_{\bar{g}2}$ (by definition), we arrive at $(I_{\bar{g}1} - I_{\bar{g}2})^2 < 0$, proving (by contradiction) that $\mathrm{V}\left(\tilde{\mu}_{\mathrm{mis},2}\right) \leq \mathrm{V}\left(\tilde{\mu}_{\mathrm{avg,is2}}\right)$. The equality holds when $g_1(x) = g_2(x)$. In this case, of two sampling distributions, the MIS estimator (with balance heuristic) has lower variance than the averaged pdf estimator.

In the general case, with N_g different importance functions ($N_g = 2$ for the above case), the variance of the MIS estimator is

$$\boxed{\mathrm{V}\left(\tilde{\mu}_{\mathrm{mis,N}}\right) = \mathrm{V}\left(\tilde{\mu}_{\mathrm{avg,isN}}\right) + \frac{\kappa}{NN_g}} \quad (7)$$

where $\kappa = \sum_{j=1}^{N_g}\sum_{j=1}^{N_g} I_{\bar{g}i}I_{\bar{g}j} - N_g(N_g-1)I^2$.

1.3. Summary of variance reduction and convergence

Our derivations for variance were for the respective primary estimators. eg. $\tilde{\mu}_{\mathrm{is},1}$. If the samples are independent, the variance of the corresponding secondary estimators is obtained as

$$\mathrm{V}\left(\tilde{\mu}_{\mathrm{is,N}}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{V}\left(\tilde{\mu}_{\mathrm{is},1}\right) = \frac{\mathrm{V}\left(\tilde{\mu}_{\mathrm{is},1}\right)}{N}. \quad (8)$$

In summary, while importance sampling can provide significant benefit, this reduction is a constant factor and the convergence of the error of the estimator remains $O(N^{-0.5})$.

2. Combining with jittered sampling

We formally analyse the variance of jittered importance sampling. We discuss the benefit of jittered antithetic importance sampling. Recall that, in jittered sampling, the domain $\mathcal{D}$ is partititioned into N strata so that $\mathcal{D} = \bigcup \mathcal{D}_i,\ 1 < i < N$ with $\mathcal{D}_i \cap \mathcal{D}_j = \emptyset, \forall i \neq j$ with proportional (equal) allocation.

2.1. Jittered importance sampling

Consider a single sample from each stratum, $\mathbf{X}_i \sim g(x),\ x \in \mathcal{D}_i$, where $g(x)$ is a pdf defined on the entire domain $\mathcal{D}$. Since the integral of the importance function is potentially different over (unequal) strata, to avoid bias,

$$\tilde{\mu}_{\mathrm{jis,N}} = \sum_{i=1}^{N}\beta i\,\frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)}, \quad \text{where} \quad \beta i = \int_{\mathcal{D}_i} g(x)\ \mathrm{d}x. \quad (9)$$

In image synthesis, it is common to pass multi-dimensional jittered samples through the inverse cdfs of the importance function [PH10, Jak10] then $\beta i = 1/N$. Then the variance of $\tilde{\mu}_{\mathrm{jis,N}}$ is

$$\begin{aligned}\mathrm{V}\left(\tilde{\mu}_{\mathrm{jis,N}}\right) &= \frac{1}{N^2}\sum_{i=1}^{N}\left(\int_{\mathcal{D}_i}\frac{f^2(x)}{g(x)}\ \mathrm{d}x - I_i^2\right) \\ &= \frac{\mathrm{V}\left(\tilde{\mu}_{\mathrm{is,N}}\right)}{N} + \frac{\sum_{i=1}^{N}\sum_{j\neq i}^{N} I_iI_j}{N^2} \quad (10)\end{aligned}$$

where I_i is the integral of $f(x)$ within the domain $\mathcal{D}_i$.

2.2. Jittered multiple-importance sampling

First consider stratification of an N-sample MIS estimator with two importance functions. The complication introduced by stratification of the MIS estimator is that the strata are different for each importance function. Let $\mathcal{D}_i^1$ and $\mathcal{D}_i^2$ denote the strata induced by the importance functions $g_1(x)$ and $g_2(x)$ respectively. The formulation that leads to the N-sample jittered MIS estimator is the decomposition of the integral into integrals over the strata:

$$I = \sum_{i=1}^{N/2}\left(\int_{\mathcal{D}_i^1}\frac{f(x)}{2\bar{g}(x)}g_1(x)\ \mathrm{d}x + \int_{\mathcal{D}_i^2}\frac{f(x)}{2\bar{g}(x)}g_2(x)\ \mathrm{d}x\right).$$

Note the different strata in the rhs. Also, since the importance functions integrate to $2/N$ within each stratum, there is a multiplication by this factor. The jittered MIS estimator (with balance heuristic) is the sum of the primary estimators in each of the strata

$$\tilde{\mu}_{\mathrm{jmis,N}} = \frac{2}{N}\left(\sum_{i=1}^{N/2}\frac{f(\mathbf{X}_{1i})}{2\bar{g}(\mathbf{X}_{1i})} + \sum_{i=1}^{N/2}\frac{f(\mathbf{X}_{2i})}{2\bar{g}(\mathbf{X}_{2i})}\right), \quad (11)$$

where $\mathbf{X}_{ki} \sim g_k(x), x \in \mathcal{D}_i^k,\ k = 1, 2$. Its variance is

$$\frac{4}{N^2}\left(\sum_{i=1}^{N/2}\mathrm{V}\left(\frac{f(\mathbf{X}_{1i})}{2\bar{g}(\mathbf{X}_{1i})}\right)_{g_1} + \sum_{i=1}^{N/2}\mathrm{V}\left(\frac{f(\mathbf{X}_{2i})}{2\bar{g}(\mathbf{X}_{2i})}\right)_{g_2}\right) \quad (12)$$

The first variance term in this expression can be written as

$$\int_{\mathcal{D}} \breve{f}^2(x) g_1(x) \ \mathrm{d}x - \left(\int_{\mathcal{D}} \breve{f}(x) g_1(x) \ \mathrm{d}x \right)^2 + \psi_1 (13)$$

where $\breve{f}(x) = f(x)/(2\bar{g}(x))$ and the remainder term is

$$\psi_k = \sum_{i=1}^{i<N/2} \sum_{j \neq i}^{j<N/2} \int_{\mathcal{D}_i^k} \breve{f}(x) g_k(x) \ \mathrm{d}x \int_{\mathcal{D}_j^k} \breve{f}(x) g_k(x) \ \mathrm{d}x.$$

Substituting this (similarly the second term), in eq. 12 yields

$$\mathrm{V}\left(\tilde{\mu}_{\mathrm{jmis,N}}\right) = \frac{2}{N} \mathrm{V}\left(\tilde{\mu}_{\mathrm{avg,isN}}\right) + \frac{4}{N^2} \left(\eta_{12} + \psi_1 + \psi_2 - I^2/2 \right)$$

where $\eta_{ij} = \left\langle \breve{f}(\mathbf{X}) \right\rangle_{g_i} \left\langle \breve{f}(\mathbf{X}) \right\rangle_{g_j}$. Generalizing to N_g distributions, the variance of the jittered MIS estimator is

$$\boxed{\mathrm{V}\left(\tilde{\mu}_{\mathrm{jmis,N}}\right) = \frac{N_g}{N} \mathrm{V}\left(\tilde{\mu}_{\mathrm{avg,isN}}\right) + \frac{N_g^2}{N^2} \psi} \qquad (14)$$

where $\psi = \left(\sum_{i=1}^{i<\frac{N}{N_g}} \sum_{j \neq i}^{j<\frac{N}{N_g}} \eta_{ij} + \sum_{i=1}^{i<N_g} \psi_i - \frac{N_g - 1}{N_g} I^2 \right)$ is the error term. Note that ψ_i is now defined so its summations run up to N/N_g instead of $N/2$ as defined earlier. As with importance sampling, we have proved that jittered sampling improves the convergence of the MIS estimator from $O(N^{-1/2})$ to $O(N^{-1})$, at the cost of an additional error term.

3. A note on variance in the Fourier domain

The error of a single estimate, expressed in the Fourier domain, is

$$\Delta = \hat{f_\Pi}(0) - \int \hat{\mathbf{S}}(\omega) \ \hat{f_\Pi}(-\omega) \, \mathrm{d}\omega. \qquad (15)$$

Note that this error is a random variable since it is a function of $\mathbf{S}(x)$. Bias and variance were derived [SK13] as the expectation and variance of this error.

3.1. Upper bound for the variance

The equation for the variance is incorrect in previous work [SK13]. We show that their result is an upper-bound. Applying the variance operator to the error in eq. 15), , substituting $\hat{f_\Pi}(\omega) = |\hat{f}(\omega)| \, \mathrm{e}^{-\imath \Phi_f}$ and simplifying,

$$\begin{aligned} \mathrm{V}\left(\Delta\right) &= |\hat{f}(\omega)|^2 \int \mathrm{V}\left(\hat{\mathbf{S}}_{re}(\omega)\right) \cos^2 \Phi_f \ \mathrm{d}\omega \\ &+ \ |\hat{f}(\omega)|^2 \int \mathrm{V}\left(\hat{\mathbf{S}}_{re}(\omega)\right) \sin^2 \Phi_f \ \mathrm{d}\omega \\ &\leq \int |\hat{f_\Pi}(-\omega)|^2 \ \mathrm{V}\left(\hat{\mathbf{S}}(\omega)\right) \ \mathrm{d}\omega. \end{aligned} \qquad (16)$$

where we use the shorthand notation $\hat{\mathbf{S}}_{re}$ and $\hat{\mathbf{S}}_{im}$ for the real and imaginary components of $\hat{\mathbf{S}}(\omega)$ respectively.

The sampling function for the primary importance sampling estimator is $\mathbf{S}(x) = \alpha(\mathbf{X}) \delta(x - \mathbf{X})$, $\mathbf{X} \sim g(x)$, where $g(x)$ is a pdf and $\alpha(x) = 1/g(x)$. Therefore, the Fourier transformed sampling function is $\hat{\mathbf{S}}(\omega) = \alpha(\mathbf{X}_i) \, \mathrm{e}^{-\imath 2\pi\omega \mathbf{X}_i}$. Adding the variances of the real and imaginary components of $\hat{\mathbf{S}}(\omega)$, simplifying and substituting in eq. 16, we obtain the upper bound for the primary importance sampling estimator as

$$\mathrm{V}\left(\tilde{\mu}_{\mathrm{is,1}}\right) \leq \int |\hat{f_\Pi}(-\omega)|^2 \ \mathrm{d}\omega \int_{\mathcal{D}} \frac{1}{g(x)} \ \mathrm{d}x \qquad (17)$$

3.2. Biased importance sampling

When samples for numerical integration are distributed non-uniformly in the domain, the estimates need to be weighted by the reciprocal of the normalized probability density to yield unbiased results. This is often expensive and normalization is sometimes a hurdle. Here, we analyze the estimator that results from simply taking the mean of the unweighted primary estimates.

Consider a single sample $\mathbf{S}(\mathbf{X}) = \delta(x - \mathbf{X})$ where $\mathbf{X} \sim g(x)$. Then, $\hat{\mathbf{S}}(\omega) = \mathrm{e}^{-2\pi\imath\omega\mathbf{X}}$, and the expected spectrum is

$$\left\langle \hat{\mathbf{S}}(\omega) \right\rangle_g = \int \mathrm{e}^{-2\pi\imath\omega x} \ g(x) \, \mathrm{d}x = \hat{g}(\omega).$$

This expectation remains the same for N iid samples. So, the *expected Fourier spectrum of a sampling function is the Fourier spectrum of the probability distribution function of its samples.*

A biased, but convenient, primary importance sampling estimator $\tilde{\mu}_{\mathrm{bis}}$ is simply $\tilde{\mu}_{\mathrm{bis,1}} = f(\mathbf{X})$, $\mathbf{X} \sim g(x)$. The bias of this estimator is obtained by applying the expectation operator on either side of eq. 15 and substituting $\left\langle \hat{\mathbf{S}}(\omega) \right\rangle = \hat{g}(\omega)$:

$$\left\langle \Delta(\tilde{\mu}_{\mathrm{bis}}) \right\rangle = \hat{f_\Pi}(0) - \int \hat{g}(\omega) \ \hat{f_\Pi}(-\omega) \, \mathrm{d}\omega. \qquad (18)$$

That is, to keep bias low, the Fourier spectrum of the pdf must be complementary to the integrand. An upper bound for the bias is simply the inner product of the amplitude spectrum of the sampling distribution and the amplitude spectrum of the integrand.

The exact variance of this biased estimator is

$$\mathrm{V}\left(\tilde{\mu}_{\mathrm{bis},1}\right) = \int\limits_{\mathcal{D}} f^2(x)g(x)\ \mathrm{d}x - \left(\int\limits_{\mathcal{D}} f(x)g(x)\ \mathrm{d}x\right)^2. \tag{19}$$

An upper bound can be derived from eq. 16, for which we need to know $\mathrm{V}\left(\hat{\mathbf{S}}(\omega)\right)_g$. But,

$$\mathrm{V}\left(\mathrm{e}^{-2\pi\imath\omega\mathbf{X}}\right)_g = \mathrm{V}\left(\cos(2\pi\imath\omega\mathbf{X})\right)_g + \mathrm{V}\left(\sin(2\pi\imath\omega\mathbf{X})\right)_g \leq 1.$$

Substituting this in eq. 16, we obtain an upper bound for the variance of this biased estimator as

$$\mathrm{V}\left(\tilde{\mu}_{\mathrm{bis},1}\right) \leq \int \left|\hat{f}_{\Pi}(-\omega)\right|^2 \mathrm{d}\omega\ . \tag{20}$$

Comparing this to the standard importance sampling estimator, we see that this upper bound is lower if $\int_{\mathcal{D}} \frac{1}{g(x)}\ \mathrm{d}x > 1$.

References

[Jak10] JAKOB W.: Mitsuba Renderer, 2010. http://www.mitsuba-renderer.org.

[PH10] PHARR M., HUMPHREYS G.: *Physically Based Rendering, Second Edition: From Theory To Implementation*, 2nd ed. Morgan Kaufmann Publishers Inc., 2010.

[SK13] SUBR K., KAUTZ J.: Fourier Analysis of Stochastic Sampling Strategies for Assessing Bias and Variance in Integration. *ACM Trans. Graph. 32*, 4 (2013), 128:1–128:12.