2005-09-16

Importance Resampling for Global Illumination

Justin F. Talbot
Brigham Young University - Provo

IMPORTANCE RESAMPLING FOR GLOBAL ILLUMINATION

by

Justin F. Talbot

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computer Science

Brigham Young University

August 2005

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Justin F. Talbot

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Parris K. Egbert, Chair
Date	Bryan S. Morse
Date	Irene Langkilde-Geary

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Justin F. Talbot in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Parris K. Egbert
Chair, Graduate Committee

Accepted for the Department

Parris K. Egbert
Graduate Coordinator

Accepted for the College

G. Rex Bryce,
Associate Dean, College of Physical and Mathematical Sciences

ABSTRACT

IMPORTANCE RESAMPLING FOR GLOBAL ILLUMINATION

Justin F. Talbot

Department of Computer Science

Master of Science

This thesis develops a generalized form of Monte Carlo integration called Resampled Importance Sampling. It is based on the importance resampling sample generation technique. Resampled Importance Sampling can lead to significant variance reduction over standard Monte Carlo integration for common rendering problems. We show how to select the importance resampling parameters for near optimal variance reduction. We also combine RIS with stratification and with Multiple Importance Sampling for further variance reduction. We demonstrate the robustness of this technique on the direct lighting problem and achieve up to a 33% variance reduction over standard techniques. We also suggest using RIS as a default BRDF sampling technique.

ACKNOWLEDGMENTS

Foremost, I would like to thank my wife, Amy Talbot, for her constant support during the writing of this thesis. Her careful editorial work found many mistakes which spared me much embarrassment.

My parents provided the necessary instruction and encouragement to get me into college and started in research.

I would also like to thank my thesis advisor, Parris Egbert, for providing support during both my undergraduate and graduate work. The members of our research group, especially Dave Cline, have provided much needed suggestions and direction on my research and this thesis.

Finally, the Stanford University Computer Graphics Laboratory provided the dragon model, XYZ RGB Inc. made the Thai statue (the elephant) model available, and Paul Debevec created the environment maps used in this thesis.

Contents

List of Figures

Chapter 1

Introduction

Global illumination is the process of turning a virtual scene into a realistic image by simulated light transport. This is a very difficult problem to solve using standard numerical techniques. Instead, current approaches are based on Monte Carlo integration, which allows for very general solutions. In exchange for this generality, the random nature of Monte Carlo integration introduces variance into the solution, which shows up as unacceptable noise in the image. In this thesis we seek robust methods that reduce the variance without compromising the generality of Monte Carlo integration.

We use the word *robust* to refer to a method that works reasonably well on a general class of problems without any information on the specific problem being solved. Robust variance reduction techniques are important in global illumination on two levels. First, the global illumination problem is high-dimensional, and the shape of the integrand is complex and discontinuous. Thus, it is impractical to apply variance reduction techniques that place restrictions on the shape of the integrand or the number of dimensions. Second, we are typically interested in finding algorithms that work across a wide variety of scenes without modification. Without this type of

robustness, significant time and expense must be spent implementing a multitude of special cases, without any guarantee that the resulting "conglomarithm" will work on the *user's* input. Many existing variance reduction techniques do not fit this definition of robust and as a consequence are of limited usefulness in global illumination.

The goal of this thesis is to develop a robust variance reduction technique for global illumination. We achieve our goal by developing a generalization of Monte Carlo integration, called Resampled Importance Sampling (RIS), that results in lower variance estimates for the types of problems that are commonly encountered in global illumination. The new method makes few assumptions about the integrand and can be applied nearly everywhere that standard Monte Carlo integration can be used.

In the following sections we discuss the global illumination problem and the necessity of robust variance reduction methods. Then, in Section 1.2, we summarize the contributions of this thesis. Finally, in Section 1.3 we outline the organization of this thesis.

1.1 Global Illumination

Global illumination is an application of light transport. The goal of light transport is to compute how light propagates from light emitters through a scene consisting of reflectors and absorbers. In global illumination, this information is used to create an image that corresponds to what an observer or camera would see if placed in that scene.

Light transport was first introduced into graphics as a recursive integral by Immel et al. [13] and independently, by Kajiya [15]. Kajiya invented the name, "Rendering Equation," by which the light transport equation is commonly known today. More recently, Veach [27] reformulated this equation as a non-recursive integral over light paths, which is a form more suitable for current global illumination approaches.

The rendering equation (in either form) is noted for its generality. It can eas-

ily incorporate difficult scene geometries, surface reflectance properties, illumination conditions, and camera properties. It has recently been extended to account for participating media [17], subsurface scattering [21, 14], and polarization and fluorescence [30] effects.

Integrating such a general equation is a very difficult problem. Currently, only one integration approach is capable of solving the completely general rendering equation, Monte Carlo integration. Unfortunately, Monte Carlo integration is a probabilistic process and is subject to variance, which appears as noise in the rendered image.

This noise can be reduced in one of two ways. First, the number of samples used for the Monte Carlo estimate can be increased. This reduces the variance, though slowly, and at increased computation cost. Second, a number of different *variance reduction techniques* can be applied. These techniques use statistical "tricks" to achieve faster variance reduction than is achieved by increasing the number of samples. Chapter 3 discusses Monte Carlo integration and some variance reduction techniques in more detail.

We can define our requirement for robust variance reduction in terms of the required *a priori* knowledge. If a variance reduction method requires that certain properties of the integrand be known, this limits its application to only those integrands where the properties are easily available. In general, the more knowledge a method requires, the less robust it is.

1.2 Summary of Original Contributions

Resampled Importance Sampling. We present a novel variance reduction technique for Monte Carlo integration called Resampled Importance Sampling (RIS). It is based on the sampling technique Sampling Importance Resampling (SIR). To estimate the integral we first generate samples from a proposal distribution. These samples are then filtered using a resampling step. The resulting samples are used in

a modified Monte Carlo integration estimator. We prove that the result is unbiased and derive an expression for its variance. We derive expressions for the optimal values of the RIS parameters. When computing the optimal value is impractical, we also find robust approximations. We discuss conditions under which RIS can be expected to give improvement over standard importance sampling.

Resampled Importance Sampling has the following important properties:

1. It is unbiased.
2. It is a true generalization of standard importance sampling and can be used as a direct replacement.
3. It requires no precomputation.
4. It does not rely on any specific properties of the function being integrated.
5. The sampling density does not have to be sampled or normalized.

These properties distiguish RIS from previous attempts to improve importance sampling.

Stratification for RIS. Stratification is an important variance reduction technique that is commonly used with standard importance sampling. Using stratification with the filtering process of Resampled Importance Sampling presents interesting challenges. We show theoretically how to correctly apply stratification in RIS. We propose a practical method that is computationally inexpensive. We also demonstrate that some stratification is necessary for RIS to be a true generalization of Monte Carlo integration.

Multiple Importance Sampling for RIS. The variance of RIS can be further reduced by combination with Multiple Importance Sampling (MIS). Multiple Importance Sampling reduces variance by using multiple proposal distributions to weight

the Monte Carlo samples. We show how to apply MIS both before and after the filtering step of RIS.

Application to Global Illumination problems. Resampled Importance Sampling is a general variance reduction technique. We demonstrate its use on global illumination problems, including BRDF sampling and direct lighting computation. We also show how it can be used to simplify adaptive sampling schemes.

1.3 Thesis Organization

In Chapters 2 and 3 we discuss relevant background material for this thesis. We first summarize methods used for generating samples from distributions. We include a brief summary of basic probability results that we will use in the thesis. We then give an introduction to Monte Carlo integration and common variance reduction techniques. We pay special attention to their application to global illumination.

In Chapter 4 we derive the Resampled Importance Sampling estimator. We show how to apply stratification and Multiple Importance Sampling to improve its variance reduction. We derive heuristics that indicate when RIS will perform better than standard Monte Carlo integration. We also find robust approximations for the RIS parameters and prove their robustness.

In Chapter 5 we apply RIS to problems in global illumination. We show how it can reduce variance in BRDF sampling and direct lighting applications. We demonstrate the effectiveness of the stratification and Multiple Importance Sampling approaches suggested in Chapter 4 and we give numerical results on amount of variance reduction.

Some of the material in this thesis has appeared previously as:

> Justin F. Talbot, David Cline, and Parris K. Egbert. Importance resampling for global illumination. In Kavita Bala and Philip Dutré, editors, *Rendering Techniques 2005 Eurographics Symposium on Rendering*, pages 139–146, Aire-la-Ville, Switzerland, 2005. Eurographics Association.

Chapter 2

Sample Generation Techniques

In this thesis we depend upon the ability to generate samples with a given distribution. This is a basic requirement for Monte Carlo integration and for the variance reduction techniques that will be presented in Chapter 3.

In this chapter, we first summarize some basic probability results and notation in Section 2.1. In Sections 2.2 through 2.4 we describe how to generate samples from some special case distributions. In Section 2.5 we cover a number of techniques for generating samples from general distributions.

2.1 Basic Probability

Probability is the formalized study of uncertainty and randomness. In this section we cover some basic probability definitions and results that we will rely on in this thesis. Monte Carlo integration is based on the common probability results which we describe here. We will also use the results to verify the correctness of our algorithms and to quantify the inevitable error that arises from using a random process.

2.1.1 Random Variables and Samples

Theoretically, random variables are functions that map the outcome of a random process to the n-dimensional real space, $\mathbb{R}^n$. In practice, we can think of them as

variables that do not hold a single value, but a *distribution* of values. The distribution can be expressed as a function relating *possible* values to probabilites. If the function is discrete, it is called a *probability mass function* (pmf). If it is continuous, it is called a *probability density function* (pdf). In the rest of this discussion we will assume that we are using pdfs, although similar results also hold for discrete distributions.

By definition a pdf, p, must comply with two requirements:

$$\forall x : \ p(x) \geq 0$$

and

$$\int_{\Omega} p(x)dx = 1$$

Note that although a pdf refers to a normalized function, in this thesis we will use the term *unnormalized pdf* to refer to a real-valued function that can be tranformed to a pdf by dividing out the appropriate normalizing constant, C. We will use the nonstandard notation $\hat{p}$ to represent an unnormalized pdf ($\hat{p} = Cp$). This will be used to emphasize the fact that the given algorithms do not rely on knowledge of the normalizing constant to work correctly.

A common transformation of the pdf is the cumulative density function (CDF):

$$F(x) = \int_{-\infty}^{x} p(x')dx'$$

This transformation will be used in the inversion method for generating samples from a distribution (see Section 2.5.1). Importantly, there is a one-to-one relationship between pdfs and CDFs.

Unlike random variables, a *sample* is a variable that only holds a single value. However, we can speak of a sample as having a distribution or being drawn from a

distribution if it is generated by a probabilistic process that could have produced a different result. Thus, the distribution of a sample describes what *could have* happened, while the value of the sample tells us what actually did happen.

2.1.2 Expected Value and Variance

Random variables have two important properties that we will use extensively in this thesis. The *expected value* of a random variable is the average value of the variable and is defined for continuous variables as

$$E(X) = \int_{\Omega} x\ p(x)dx \tag{2.1}$$

The expected value is also known as the mean.

The *variance* measures how widely the random variable can deviate from the expected value. It is defined as:

$$V(X) = E\left(X^2\right) - E(X)^2$$

In global illumination applications, variance is unwanted because it appears as noise in the image. We will be interested in choosing random variables that have minimum variance.

2.1.3 Biasedness, Consistency

Estimators are functions of samples drawn from a distribution. As the name implies, these functions are used to estimate certain values using only samples from that distribution. For example, we can estimate the expected value of a random variable, X, using the following estimator:

$$f(x_1, ..., x_N) = \frac{1}{N}\sum_{i=1}^{N} x_i \approx E(X) \tag{2.2}$$

where $\{x_1, ..., x_N\}$ are samples drawn from the distribution of X.

We call an estimator *unbiased* if the expected value of the estimator is equal to the value that is being estimated. Intuitively, an unbiased estimator is correct on average. Otherwise it is called *biased*. An estimator is called *consistent* if its variance and bias approach zero as the number of samples in the sample set approaches infinity.

To be useful in global illumination, the estimator must be at least consistent. Additionally, we prefer estimators that are also unbiased for the reasons given in Veach's thesis [27].

2.2 Uniform Samples

We use the Greek letter ξ ('xi') to represent draws from the uniform distribution between 0 and 1, i.e.:

$$\xi \sim U(0, 1)$$

Uniform samples are commonly used to produce samples from other distributions. Because of this, in the following sections we assume that uniform samples are readily available. Pseudo-random uniform samples are easily generated in most programming languages and in other scientific software. True uniform random variables can be generated using a variety of physical phenomena.

2.3 Sampling Discrete Distributions

Sampling discrete distributions is a common task in global illumination algorithms. In this thesis we will use a common technique which treats the discrete distribution as a piecewise constant continuous function then uses the CDF inversion technique described in Section 2.5.1

A second technique for generating samples from discrete distributions, the *Alias method*, has some advantages over the inversion method if many samples are being

drawn from the same distribution. The Alias method is described in a global illumination context by Burke [4].

2.4 Sampling Mixture Distributions

Mixture distributions have a probability density function that is a weighted sum of other densities:

$$\hat{q}(\theta) = \sum_{i=0}^{N} \alpha_i \hat{p}_i(\theta)$$

Samples can easily be generated from the distribution with pdf $\hat{q}$ in a two-step process. First, we choose i by sampling from a discrete distribution with probability proportional to α_i. We then produce a sample from the distribution with pdf $\hat{p}_i$ using an appropriate sample generation technique.

Evaluating $\hat{q}(\theta)$ requires evaluating all $\hat{p}_i(\theta)$ even if the actual sample was not produced from that distribution.

2.5 Sampling General Distributions

Sampling general distributions is more difficult. The most common techniques, presented below, all use a two-stage sampling method. First, initial samples are generated from a distribution. Second, those samples are modified in some fashion to produce samples that have the desired distribution (or an approximation). The initial samples are usually taken from a uniform distribution, since such samples are readily available.

The first two techniques discussed here produce samples that have the exact desired distribution. This property is nice, since the samples can be used directly in unbiased Monte Carlo integration. However, it comes at the expense of requiring significant knowledge of the shape of the desired distribution.

The final three sampling techniques relax this requirement, but at the cost of exactness. The resulting distributions will only be approximations to the desired

distribution. Although they are approximations, it is still possible to use them in unbiased Monte Carlo integration if proper weighting of the samples is performed.

2.5.1 CDF Inversion

CDF inversion is the most common sampling technique used in global illumination. It permits exact sampling from the desired distribution.

Remember, the CDF is the integral of the pdf:

$$F(x) = \int_{-\infty}^{x} q(x')dx'$$

CDF inversion takes advantage a of special property of the CDF; $F(x)$ is uniformly distributed between 0 and 1. This implies:

$$\xi = F(x)$$

If we solve this equation for x,

$$x = F^{-1}(\xi)$$

we get a simple, efficient method for producing samples, x, from the pdf q. The downside is that it requires a normalized pdf, a closed form integral, and an invertible CDF. These are available in some important, fundamental applications in global illumination. However, these limitations mean that CDF inversion cannot be applied to more difficult, more general global illumination problems.

2.5.2 Rejection Sampling

Rejection Sampling can be used to generate samples from a distribution with density $\hat{q}(\theta)$. To use Rejection Sampling, we must be able to evaluate $\hat{q}(\theta)$ and we must be able to find a pdf $\hat{p}(\theta)$ that bounds $\hat{q}$ (i.e the ratio $\frac{\hat{q}(\theta)}{\hat{p}(\theta)}$ must have a known finite bound, M, for all θ).

We generate a sample in two steps. First, we generate a sample according to $\hat{p}$ using some other sample generation technique. Second, if

$$\xi < \frac{\hat{q}}{\hat{p}M}$$

we accept the sample; otherwise, the sample is rejected. These two steps are repeated until a sample is accepted.

Rejection Sampling can be understood as a dart throwing process. Conceptually, we first generate samples uniformly in the area underneath the scaled distribution $\hat{p}M$. Then, those samples that fall within the area underneath the curve $\hat{q}$ are accepted, while those above are rejected. To be efficient, $\hat{p}M$ should bound $\hat{q}$ as tightly as possible.

The difficulty in using Rejection Sampling arises from the need to find bounding distributions. It typically requires a great deal of a priori knowledge about the shape of $\hat{q}(\theta)$ in order to choose a distribution that will bound it tightly. Also note that finding M requires knowledge of the relative scale between $\hat{q}$ and $\hat{p}$, which is not always available. Although it is simple to extend the Rejection Sampling idea into higher dimensions, finding efficient bounding distributions becomes increasingly difficult as the dimensionality grows.

Finally, in application to Monte Carlo integration, Rejection Sampling is never better than other available sampling techniques [6].

2.5.3 Sampling Importance Resampling

Sampling Importance Resampling (or more simply, importance resampling) is a common method in computational statistics for generating samples from difficult distributions. It is commonly used in sequential importance sampling and particle

filtering [9]. It can also be used to generate samples from Bayesian posterior distributions [10]. Importance resampling was first described by Rubin [23].

Importance resampling can generate samples that approximately have the distribution $\frac{\hat{q}}{C} = q$. To do so, we generate a set of "proposal" samples from a source distribution, p, weight these samples appropriately, then *resample* these samples by drawing samples from them with probability proportional to their weights. This algorithm is given below.

Importance Resampling

1. Generate M proposals ($M \geq 1$) from the source distribution p, $\{x_1, \ldots, x_M\}$.
2. Compute a weight for each proposal, $w(x_j)$.
3. Draw N samples ($N \leq M$), $\{y_1, \ldots, y_N\}$, *with replacement* from the proposals with probability proportional to their weights.

If we choose $w(x_j) = \frac{\hat{q}(x_j)}{p(x_j)}$, then the resulting samples will be approximately distributed according to $\hat{q}$. The effect of the resampling step is to take proposals from the source density, p, and "filter" them, so that the samples have a distribution that approximates q.

We can view M, the number of proposals, as a distribution interpolation variable. When $M = 1$, the samples are marginally distributed according to p. As $M \to \infty$, the distribution of each sample approaches q. As an example, Figure 2.1 shows the distribution of a SIR-generated sample for various values of M when p is uniform and $\hat{q} = \cos(\theta) + \sin^4(6\theta)$.

Importance resampling has been used informally in the global illumination literature. Lafortune et al. [16] used importance resampling to decrease the number of

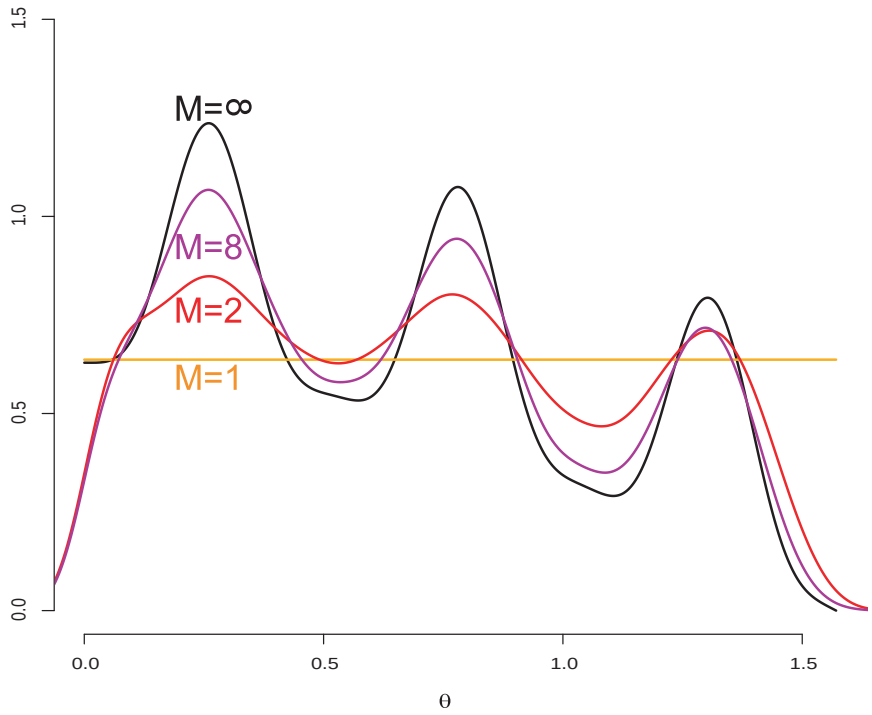

Figure 2.1: Marginal distribution of a sample resulting from importance resampling for different values of M. At $M = 1$, the distribution is p. At $M = \infty$, the distribution is q. For other values of M, the distribution interpolates between p and q, though the exact manner of interpolation is unknown. The low values on the left side for $M = 2$ and $M = 8$ are artifacts of the density estimation method.

visibility tests necessary in bidirectional path tracing. Shirley et al. [24] used resampling to improve direct lighting computations. Burke [4] used importance resampling to sample the distribution of the product of a Phong BRDF model and an illuminating environment map. Chapter 5 formalizes these techniques.

One of the contributions of this thesis is to show how to weight the samples generated by Sampling Importance Resampling to produce an unbiased Monte Carlo integration estimate.

2.5.4 Metropolis Sampling

Metropolis sampling [19] generates a Markov chain with a stationary distribution that is equal to the desired sampling distribution, $\hat{q}$. The advantage of Metropolis

sampling is its generality. It can be used to generate samples from even the most complex distributions. It is very computationally efficient if a large number of samples are desired from the same distribution. Metropolis sampling was first used in global illumination by Veach and Guibas [28].

The Markov chain is formed by proposing a new sample, x', from a distribution, called the transition function, T, which is conditioned on the current sample, x. The proposed sample becomes the current sample if

$$\xi < min\left(1, \frac{\hat{q}(x')T(x|x')}{\hat{q}(x)T(x'|x)}\right)$$

otherwise the old sample is retained as the current sample.

Like the previous technique, Metropolis sampling generates samples that are only approximately distributed according to the desired distribution, $\hat{q}$. As the Markov chain runs, the quality of the approximation improves. Thus, it is common to throw out the initial samples (the "burn-in" period), when the approximation is bad, and only use the subsequent samples, when the approximation is hopefully good. This typically works in practice, although determining an appropriate length of the "burn-in" is difficult. Additionally, the result is still an approximation. Perfect sampling with Metropolis can be achieved using the *Coupling from the Past* method introduced by Propp and Wilson [22].

Fortunately, when Metropolis is applied to Monte Carlo integration, it turns out that we often do not need perfect sampling. Veach and Guibas [28] showed that with appropriate weighting, the integration estimate will be unbiased, even though the samples are only approximately distributed.

2.5.5 Gibbs Sampling

Gibbs sampling [11], also known as successive substitution Markov Chain Monte Carlo (MCMC), is a special case of Metropolis. During each step of the Markov chain, only a single dimension of the parameter vector is updated. Its new value is proposed using the *complete conditional*, the distribution of the free dimension conditioned on keeping all the other dimensions constant. In comparison with standard Metropolis, Gibbs sampling can be more efficient because the proposals come from a more informed distribution, which can lead to a higher acceptance rate.

Gibbs sampling has not yet been applied to Global Illumination. This is largely due to the fact that finding the complete conditionals is too difficult for most global illumination problems. Additionally, Gibbs sampling, due to its single component update strategy, tends to build the Markov chain along axis aligned directions in the primary sampling space. If the integrand is not aligned with an axis, proposing samples only along axis directions leads to a high rejection rate and higher variance. In some cases it can lead to severe bias in the resulting samples if the chain is not (or is nearly not) ergodic.

2.6 Summary

In this chapter we have summarized some basic probability results that will be used in this thesis. We have also discussed different sample generation techniques and their use in global illumination.

Chapter 3

Monte Carlo Integration

Monte Carlo integration is a probabilistic method for integrating difficult functions. It is commonly used in Global Illumination, in preference to more standard integration techniques like numerical quadrature, due to its generality, ease of use, and robustness in high dimensions and to discontinuous functions.

As a probabilistic method, Monte Carlo integration is subject to variance. In Global Illumination, this shows up as undesireable noise in the resulting images (see Figure 3.1). A large number of techniques have been developed to try to reduce the variance of the estimate without increasing the number of samples that have to be taken and thus without increasing the computational cost of the integration.

In this chapter we first discuss Monte Carlo integration in Section 3.1. We then discuss some commonly used variance reduction techniques in Global Illumination. In Section 3.2.1 we describe importance sampling. In Section 3.2.2 we discuss stratified sampling. Finally, in Section 3.2.3 we discuss correlated sampling. A more complete discussion of variance reduction techniques in the context of Global Illumination can be found in Veach's thesis [27].

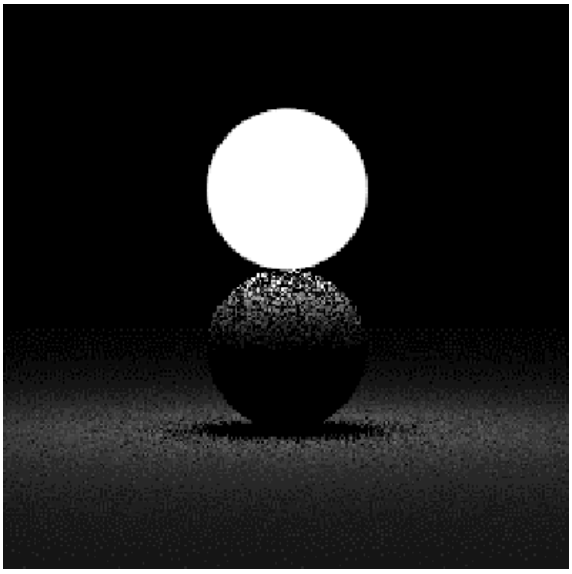

Figure 3.1: This simple image demonstrates the typical speckled noise that is associated with Monte Carlo solutions to the Global Illumination problem.

3.1 Basic Monte Carlo Integration

Monte Carlo integration is based on the fact that the integral of f can be approximated with the following estimator:

$$\int_{\Omega} f(\omega)\, d\omega \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{q(x_i)} \tag{3.1}$$

where the samples $\{x_1, ..., x_N\}$ are drawn from the *sampling distribution* with pdf q. Note that q must be normalized. This expression follows from Equations 2.1 and 2.2.

This estimator provides an unbiased estimate of the integral. Also, the estimator works for general, non-continuous, high dimensional integrals. It only requires that f be evaluated. This generality makes Monte Carlo integration ideal for Global Illumination where the integrands are seldom well-behaved and are almost always high-dimensional.

3.2 Variance Reduction

The variance of the Monte Carlo integration estimator is

$$V\left(\sum_{i=1}^{N}\frac{f\left(x_i\right)}{q\left(x_i\right)}\right)=\frac{1}{N}V\left(\frac{f\left(\omega\right)}{q\left(\omega\right)}\right) \tag{3.2}$$

This implies that as long as $V\left(\frac{f(\omega)}{q(\omega)}\right)$ is finite, then the estimator is also consistent, since the variance will go to zero as $N \to \infty$. This condition is met if $q(\omega) > 0$ whenever $f(\omega) \neq 0$.

This variance expression implies that the error of the estimate only decreases with $O(N^{\frac{1}{2}})$. This convergence is quite slow. To remedy this, a number of variance reduction techniques have been developed to increase the convergence rate.

In the following sections we discuss a few variance reduction techniques that have been applied in Global Illumination.

3.2.1 Importance Sampling

From examination of Equation 3.2, we see that the variance of the Monte Carlo integration estimator is dependent on the variance of the ratio $\frac{f(\omega)}{q(\omega)}$. If the variance of this ratio can be decreased, the overall variance can be reduced without increasing the number of samples.

Importance sampling refers to the technique of choosing the sampling distribution q to minimize the variance of the ratio. Ideally, if $q \propto f$, then the ratio is a constant for all ω. In this case, the variance is zero and there is no error in the estimate. Unfortunately, finding such a q is typically impractical, since it requires integrating f, which is the very problem we are trying to solve. Instead, we try to find a sampling distribution that mimics f. We are greatly limited in our choices of distributions since the distribution must be normalized and it must be easy to generate samples from the distribution. In practice, we generally choose a simple function that has a known

sampling method. If the integrand is a product function, the distribution is typically chosen to match a few of the terms, but seldom all since the complexity of the function grows quickly in the number of terms.

If a specific integrand, f, is going to be integrated repeatedly (possibly with slight variations), it may be worthwhile to dedicate some computation time upfront to find a good q. This approach proceeds in two steps. First, a functional form must be developed that is sufficiently flexible to allow q to closely match f, while still remaining normalized and easy to sample. Second, a method for precomputing the functional parameters must be derived that is fast enough to make the precomputation worthwhile. Some recent examples in global illumination include Structured Importance Sampling by Agarwal et al. [1], a factored BRDF representation by Lawrence et al. [18], and Wavelet Importance Sampling by Clarberg et al. [7].

Although importance sampling can achieve remarkable results, it suffers from a major shortcoming: the sampling density, q, must be chosen specifically for each f. This means that, according to our definition, importance sampling is not robust.

3.2.1.1 Defensive Importance Sampling

Note that although importance sampling will typically decrease the variance, it is also possible to increase the variance if q is chosen poorly. This happens in global illumination since the exact form of f is unknown a priori.

There are two cases that can increase the variance. First, if q is very small where f is relatively large, then the ratio $\frac{f}{q}$ will be extremely large. Second, if q is large where f is relatively small, the ratio will be very small. This wide variation in the ratio leads to increased variance in the Monte Carlo estimator. Typically, the first case is the more problematic of the two. In the second case, the ratio can get no smaller than 0, but in the first case the ratio can approach infinity.

Defensive importance sampling [12] refers to a technique used to reduce the pos-

sibility of producing a very small q where f is large. This is done by replacing q with a mixture distribution that includes a uniform density, $u(\omega)$, over the domain:

$$q'(\omega) = \alpha q(\omega) + (1-\alpha)u(\omega)$$

This approaches guarantees that the new sampling distribution, q', will never be smaller than $(1-\alpha)u(\omega)$. For an integrand, f, with an upper bound U, the following inequality holds:

$$\frac{f(\omega)}{q'(\omega)} \leq \frac{U}{(1-\alpha)u(\omega)}$$

This bound on the ratio places an upper bound on the variance of the estimator and greatly improves its robustness. In general, there is not a clear choice for α, although Veach [27] defends 0.5 as a reasonable default.

3.2.1.2 Multiple Importance Sampling

The defensive importance sampling strategy can be generalized. Veach [27] noted that often we know that f could be mimicked well by any one of a set of sampling distributions $\{q_1, ..., q_K\}$. However, a priori, we don't which one is the best match. In this case we can combine all the distributions into a single density:

$$q'(\omega) = \sum_{i=1}^{K} \alpha_i q_i(\omega)$$

When using this mixture density, the q_i which does match f will bound the ratio $\frac{f}{q'}$ and limit the variance. Although this approach will not always reduce the variance, in general it greatly increases the robustness of Monte Carlo integration.

Veach presents Multiple Importance Sampling in a more general framework where samples are weighted using heuristics that are designed to reduce variance. The mixture distribution approach given here corresponds to the balance heuristic.

Veach also proposed a power and maximum heuristics that do not correspond to a standard Monte Carlo estimate. Instead, they can be represented as a weighted Monte Carlo estimate:

$$\int_{\Omega} f(\omega)\, d\omega \approx \frac{1}{N} \sum_{i=1}^{N} w(x_i) \frac{f(x_i)}{q(x_i)}$$

With some suitable restrictions on w, this estimate is also unbiased. See Chaper 9 of Veach's thesis [27] for more information.

Multiple Importance Sampling has been generalized by Csonka et al. [8] to sequences of integrals. Owen and Zhou [20] demonstrate that defensive and multiple importance sampling can perform poorly under some circumstances and suggest combining multiple importance sampling with control variates to bound the variance. They also show how to use multiple importance sampling with mixed sign functions.

3.2.1.3 Weighted Importance Sampling

Weighted Importance Sampling modifies the Monte Carlo estimate to use a second probability density function, g.

$$\int_{\Omega} f(\omega)\, d\omega \approx \frac{\frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{q(x_i)}}{\frac{1}{N} \sum_{i=1}^{N} \frac{g(x_i)}{q(x_i)}}$$

If g mimics f well, then the numerator and the denominator will fluctuate jointly and the variance of the ratio will be reduced. To see any improvement, g must mimic f better than q. Since g does not have to be sampled we have a wider variety of functions to choose from than when choosing the sampling distribution q. However, g must still be normalized which makes finding an effective density difficult.

Additionally, weighted importance sampling is consistent, but not unbiased. Thus, for small sample sizes the estimate will be noticably incorrect on average.

Weighted Importance Sampling has been used for some global illumination problems. However, the normalization requirement has restricted its use to simple applications where g can be found in closed form. Bekaert et al. [3] use it in a Monte Carlo radiosity application and Balázs et al. [2] describe its use in Monte Carlo radiance shooting algorithms.

3.2.2 Stratified Sampling

The goal of stratified sampling is to improve the uniformity of the sample distribution. It can be shown that if the samples are, in some sense, more evenly spaced in the integrand's domain, the variance will be reduced. Intuitively, there is a higher chance of capturing the important information in the integrand if the samples cover the entire domain than if they are tightly clustered in a small portion of the domain.

Stratified sampling works by dividing the domain, Ω into a set of regions, $\{\Omega_1, ...\Omega_N\}$ and then taking samples from each region. In the special, but common case, where a single sample is taken from each region, the stratified Monte Carlo estimator is

$$\int_\Omega f(w)dw \approx \sum_{i=1}^{N} \frac{f(x_i)}{p_i(x_i)}$$

where p_i is a probability density with support only in region Ω_i.

3.2.3 Correlated Sampling

Correlated Sampling works by finding a second function g, which approximates f, such that $f - g$ is nearly constant. The integral of f is then approximated as:

$$\int_\Omega f(\omega)\, d\omega \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i) - g(x_i)}{q(x_i)} + \int_\Omega g(\omega)d\omega$$

Using Correlated Sampling requires finding g, a closed form approximation of the integrand that can be easily integrated. In general this is difficult to do for global

illumination problems. Szecsi et al. [25] use correlated sampling to improve direct lighting calculations for the limited case of polygonal diffuse lights and phong BRDFs since an approximation can easily be found. A generalization to more complex lights and BRDFs is not readily apparent. Szecsi et al. also demonstrate how to combine correlated sampling and importance sampling in a pseudo-optimal manner.

3.3 Summary

In this chapter we briefly described Monte Carlo integration and its application to Global Illumination. We noted that noisy results, due to variance, are the main problem in Monte Carlo algorithms. We discussed the various variance reduction techniques that have been derived for Monte Carlo integration and their use in Global Illumination.

Chapter 4

Resampled Importance Sampling

In this chapter, we introduce a robust variance reduction technique called *resampled importance sampling* (RIS). This technique results from using Sampling Importance Resampling to generate samples for Monte Carlo integration.

Importantly, resampled importance sampling requires no a priori knowledge of the integrand to achieve substantial variance reduction. This makes it significantly more robust than standard importance sampling, and allows it to be applied to a wider range of problems.

In Section 4.1 we introduce the basic resampled importance sampling estimate. Then, in Section 4.2 we analyze its variance and discuss conditions under which we can expect RIS to give improvements over standard Monte Carlo integration. In Sections 4.3 and 4.4 we show how to choose the RIS parameters in a robust fashion. We apply stratification and Multiple Importance Sampling to RIS in Sections 4.5 and 4.6 respectively. Finally, in Section 4.7 we compare RIS to previous work using resampling in global illumination.

4.1 Resampled Importance Sampling

We want to find the integral I of a function $f(x)$:

$$I = \int_{\Omega} f(x)\, d\mu(x)$$

using Monte Carlo integration:

$$I \approx \hat{I}_{MC} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(y_i)}{\hat{q}(y_i)}$$

This will only work correctly if $\hat{q}$ happens to be normalized and easy to sample perfectly, i.e. using rejection sampling or CDF inversion. To overcome this restriction we must modify the Monte Carlo estimator.

First, we use Sampling Importance Resampling (SIR) to generate samples from $\hat{q}$. Recall that SIR can be used even if $\hat{q}$ is unnormalized or cannot be sampled perfectly.

Second, since the samples generated by SIR are only approximately distributed according to $\hat{q}$, to maintain unbiasedness we must add a weighting term to the standard Monte Carlo estimator. For generality, we allow the weighting term to be a function of both the proposals, $\{x_1, \ldots, x_M\}$, and the resulting samples, $\{y_1, \ldots, y_N\}$:

$$\hat{I}_{ris} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(y_i)}{\hat{q}(y_i)} \cdot w^*(x_1, \ldots, x_M, y_1, \ldots, y_N)$$

The weighting function w^* must be chosen to correct for both the fact that $\hat{q}$ is unnormalized and for the fact that the density of the samples y_i only approximates $\hat{q}$. The appropriate choice of w^* is surprisingly simple. It is the average of the weights

computed in the resampling step:

$$w^*\left(x_1,\ldots,x_M,y_1,\ldots,y_N\right)=\frac{1}{M}\sum_{j=1}^{M}w(x_j)$$

Combining these two equations gives the basic RIS estimator:

$$\hat{I}_{ris}=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{f\left(y_i\right)}{\hat{q}\left(y_i\right)}\right)\cdot\frac{1}{M}\sum_{j=1}^{M}w(x_j) \tag{4.1}$$

For the RIS estimate to be unbiased, two conditions must hold. First, $\hat{q}$, the sampling density, and p, the proposal density, must be greater than zero everywhere that f is non-zero. Second, M and N must be greater than zero. The proof of unbiasedness is given in Appendix A.1.

We can now give the entire Resampled Importance Sampling algorithm.

Resampled Importance Sampling

1. Generate M proposals ($M \geq 1$) from the source distribution p, $\{x_1,\ldots,x_M\}$.
2. Compute a weight for each proposal, $w(x_j)=\frac{\hat{q}(x_j)}{p(x_j)}$.
3. Draw N samples, $\{y_1,\ldots,y_N\}$, *with replacement* from the proposals with probability proportional to the proposal weights.
4. Estimate the integral, I, using Equation (4.1).

The basic RIS estimator has a problem. When performing the Sampling Importance Resampling, there is some chance of choosing proposals multiple times, resulting in repeated samples. This is undesirable in an integration application since repeated samples do not provide any additional information, and are, essentially, wasted. One

major implication of this problem is that when $M = N$, basic RIS does not reduce to standard Monte Carlo integration since the effective sample size is smaller than N.

One possible solution to this problem is to design a resampling procedure that uses sampling without replacement. Unfortunately, it is difficult to derive an unbiased estimator using this approach.

A much easier approach is to stratify the proposals. If each proposal belongs to a single stratum, and a single sample is drawn from each stratum then no sample will be replicated. In Section 4.5 we show how to stratify RIS to avoid the sample replication problem.

4.2 Variance Analysis

Since the RIS estimator is unbiased, the only error in the estimate is due to the variance:

$$V\left(\hat{I}_{ris}\right) = \frac{1}{M} V\left(\frac{f(X)}{p(X)}\right) \quad + \quad \left(1 - \frac{1}{M}\right) \frac{1}{N} V\left(\frac{f(Y)}{q(Y)}\right) \tag{4.2}$$

where X is a random variable with density p and Y is a random variable with density $q = \frac{\hat{q}}{\int \hat{q}}$. The variance term is derived in Appendix A.2.

RIS variance is a combination of two standard Monte Carlo variance terms. The left-hand term is the variance of Monte Carlo integration using p as the sampling density. The right-hand term is the variance resulting from using q as a sampling density, divided by the number of samples. The reciprocal of the number of proposals, M, blends between them. Note that as $M \to \infty$, the contribution of the first term goes to zero. This is due to the fact that as $M \to \infty$, Sampling Importance Resampling produces samples that have the density q.

Compared to standard Monte Carlo integration, the Resampled Importance Sampling estimator adds an additional degree of freedom, the number of proposals, M. This extra flexibility can lead to additional variance reduction when two conditions

are met. First, the following equation must be true:

$$V\left(\frac{f(Y)}{q(Y)}\right) < V\left(\frac{f(X)}{p(X)}\right)$$

This condition holds when q is a better importance sampling density than p, i.e. $\hat{q}$ mimics f better. Otherwise, there will be no advantage to performing the resampling step. Second, computing proposals must be more computationally efficient than evaluating the samples. If not, we would be better off simply computing more samples, rather than wasting time generating proposals.

4.3 Robust Parameter Selection

When using RIS we can freely choose p, $\hat{q}$, M, and N within the unbiasedness constraints given in Section 4.1. Clearly some choices will lead to lower variance than others. In this section we briefly discuss choosing $\hat{q}$ and p. We then formally show how to find efficiency optimal (minimizing variance in a fixed computation time) values of M and N, for given choices of $\hat{q}$ and p.

From the discussion in the previous section we derive three guidelines for choosing $\hat{q}$ and p.

1. We should minimize $V\left(\frac{f(X)}{p(X)}\right)$ and $V\left(\frac{f(Y)}{q(Y)}\right)$. This means choosing p and $\hat{q}$ to be as proportional to f as possible.

2. $V\left(\frac{f(Y)}{q(Y)}\right)$ should be less than $V\left(\frac{f(X)}{p(X)}\right)$. If this is not true, then RIS will not be better than standard Monte Carlo integration.

3. Proposals should be computationally inexpensive. This implies that $\hat{q}$ and p should be cheap to evaluate and p should be easy to sample.

We now derive a heuristic for choosing efficiency optimal values for M and N.

The values are chosen to minimize the overall variance of the RIS estimate under a fixed computation time constraint and given fixed choices of p and $\hat{q}$.

If the total execution time available for computing $\hat{I}_{ris}$ is T_{total}, then we have the following constraint:

$$T_{total} = M\ T_1\ +\ N\ T_2$$

where T_1 is the time necessary to perform steps 1 and 2 of Resampled Importance Sampling and T_2 is the time necessary to perform steps 3 and 4.

We would like to choose values of M and N that minimize Equation (4.2) given the above time constraint. We minimize the variance by substitution to find an optimal value of $K = \frac{M}{N}$:

$$K = \frac{T_2}{T_1}\sqrt{\frac{\frac{T_{total}}{T_2}V\left(\frac{f(X)}{p(X)}\right) - V\left(\frac{f(Y)}{q(Y)}\right)}{\frac{T_{total}}{T_1}V\left(\frac{f(Y)}{q(Y)}\right) - V\left(\frac{f(Y)}{q(Y)}\right)}} \tag{4.3}$$

As should be expected, the optimal ratio of M and N is a function of the variance and the execution time of the proposals and the samples. This equation is derived in Appendix A.3.

The following equations compute the optimal values, M_{opt} and N_{opt} from K. The initial clamping operation ensures that $M_{opt} \geq N_{opt} \geq 1$, which is necessary for the resampling process. The floor operations produce integer values for M_{opt} and N_{opt}. As before, ξ is a uniform random number between 0 and 1.

$$\tilde{K} = \max\left(\min\left(K, \frac{T - T_2}{T_1}\right), 1\right)$$

$$N_{opt} = \left\lfloor \frac{T_{total}}{\tilde{K}T_1 + T_2} + \xi \right\rfloor$$

$$M_{opt} = \left\lfloor N_{opt}\tilde{K} + \xi \right\rfloor$$

4.4 Robust Approximations of M and N

In practice, the true optimal values of M and N cannot be computed since Equation (4.3) relies on unknown parameters–the computation times and the variances. Since computing the variances can be difficult, in this section we introduce a robust approximation of Equation (4.3) that only requires estimates for T_1 and T_2. These times are very simple to estimate in global illumination and other applications.

Using only T_1 and T_2, a simple approximation for K is

$$K^* = \frac{T_2}{T_1} \tag{4.4}$$

Using K^* results in assigning equal amounts of computation time to evaluating proposals and to evaluating samples. Intuitively, we cannot expect to do much better than this since we could at most double the number of samples or double the number of proposals.

Not only is K^* simple and intuitive, it is provably robust. We can show that using K^* instead of the true optimal value, K, will at most double the variance. This is shown rigorously in Appendix A.4.

We have found that this approximation works very well. It is very cheap to compute, and it avoids the necessity of computing Monte Carlo estimates of $V\left(\frac{f(X)}{p(X)}\right)$ and $V\left(\frac{f(Y)}{q(Y)}\right)$. In practice, the slightly larger variance is worth (in an efficiency sense) the reduced computation.

4.5 Stratified Resampled Importance Sampling

In this section, we first briefly discuss stratifying the RIS propsals. We then describe how to stratify the RIS samples. We derive a new stratified RIS estimator and its variance. We use the variance expression to derive heuristics for good stratification.

We describe two simple algorithms for stratifying the samples that attempt to meet some of the heuristic requirements.

4.5.1 Stratifying Proposals

In RIS, the proposals have two purposes: 1) to estimate the normalizing constant of $\hat{q}$ and 2) to provide a discrete approximation of $\hat{q}$ from which the samples will be drawn. For a low-dimensional, smooth $\hat{q}$, stratification of the proposals will improve the estimator's performance greatly. Since the proposals are drawn from p, a normalized continuous pdf, standard stratification techniques can be used.

4.5.2 Stratifying Samples

As noted previously, the basic RIS estimator can result in duplicated samples which leads to effective sample sizes smaller than N. This problem can be avoided by stratifying the proposals, such that a single sample is drawn from each stratum. In this manner, no sample will be repeated.

To effect the stratification, we modify the standard RIS algorithm as follows:

Stratified Resampled Importance Sampling

1. Generate M proposals ($M \geq 1$) from the source distribution p, $\{x_1, \ldots, x_M\}$.
2. Compute a weight for each proposal, $w(x_j) = \frac{\hat{q}(x_j)}{p(x_j)}$.
3. Divide the proposals into N strata of size m_i, where $\sum_{i=1}^{N} m_i = M$.
4. Draw N samples, $\{y_1, \ldots, y_N\}$, one from each stratum, with probability proportional to the proposal weights.
5. Estimate the integral, I, using Equation (4.5).

The stratified RIS estimator is

$$\hat{I}_{ris_2} = \sum_{i=1}^{N} \left(\frac{f\left(y_i\right)}{\hat{q}\left(y_i\right)} \cdot \frac{1}{M} \sum_{j=1}^{m_i} w(x_{ij}) \right) \tag{4.5}$$

where m_i is the number of proposals in stratum i and x_{ij} is the jth proposal in the ith stratum. Importantly, since sample y_i will be selected from just the m_i samples in stratum i, if all the $m_i = 1$, this estimate reduces to standard importance sampling, unlike the basic RIS estimate given in Equation (4.1).

The variance of the stratified estimator is

$$V\left(\hat{I}_{ris_2}\right) = \sum_{i=1}^{N} \left[\frac{1}{m_i} V\left(\frac{f_i(X_i)}{p_i(X_i)} \right) + \left(1 - \frac{1}{m_i}\right) V\left(\frac{f_i(Y_i)}{q_i(Y_i)} \right) \right] \tag{4.6}$$

where X_i is a random variable with density p_i and Y_i is a random variable with density $q_i = \frac{\hat{q}_i}{\int \hat{q}_i}$. This is very similar to the variance equation of standard RIS (Equation (4.2)).

For a given stratification, the variance of stratified RIS is guaranteed to be less than or equal to the variance of stratified importance sampling as long as $V\left(\frac{f(Y)}{q(Y)}\right) < V\left(\frac{f(X)}{p(X)}\right)$. This is a main advantage of stratified RIS over standard RIS.

Stratification is the task of dividing the proposals $\{x_1, ..., x_M\}$ into N strata. The stratification should ensure that each proposal is placed in a single stratum and that no stratum is empty. Within these constraints, there is considerable flexibility on how to choose the stratification. Ideally, we would like to choose strata that minimize Equation (4.6). Finding such optimal strata would require an expensive search through the space of possible stratifications.

Instead, we derive three heuristics from Equation (4.6) that should lead to lower

variance. These heuristics guide the design of the stratification algorithms presented in this section.

Heuristic 1: Make $\sum_{j=1}^{m_i} w_{ij}$ constant for all i. Variation in the total weight in each strata increases the variance.

Heuristic 2: Minimize the areas of the strata w.r.t. the domain of f. We would like to minimize the variance of $\frac{f_i}{q_i}$. However, we have no information on this variance when performing the stratification. Instead, we make the general assumption that f is locally constant. When this assumption is true, minimizing the area of the strata will reduce $V\left(\frac{f_i}{q_i}\right)$.

Heuristic 3: Minimize $V\left(\frac{q_i}{p_i}\right)$. In regions where q_i is a good approximation of f_i, this heuristic will reduce the variance of $\frac{f_i}{p_i}$, thus decreasing the first term of Equation (4.6). This heuristic is met by placing proposals with similar weights ($w(x) = \frac{\hat{q}(x)}{p(x)}$) into the same stratum.

In addition to these heuristics it is also necessary that the stratification process be computationally inexpensive. In the following sections we describe two simple stratification techniques that partially satisfy these heuristics.

4.5.2.1 Equal-proposals Stratification

Equal-proposals stratification is the simplest technique. We simply divide the proposals into N strata with $\frac{M}{N}$ proposals in each. (If $\frac{M}{N}$ is not an integer, the strata are made as equal in size as possible.) If we assume that the order of the proposals indicates some form of spatial locality, then to fulfill heuristic 2 we should not reorder the proposals during stratification. Despite its simplicity, equal-proposals stratification works reasonably well. It has the important property that its variance is

guaranteed to be less than or equal to the variance of standard importance sampling for a fixed number of samples, N.

4.5.2.2 Equal-weights Stratification

Equal-weights stratification is designed to meet the first heuristic. Proposals are divided into strata, such that the sum of the weights in each stratum is nearly equal. Again, for simplicity and to follow the second heuristic, we do not allow rearrangement of the proposals.

To effect this stratification we first compute the weight

$$w_{max} = \frac{\sum_{i=1}^{N} w_i}{N}$$

that should be in each stratum. We then greedily create strata by stepping through the propsals in order. The first proposal is placed in the first stratum. Then, as long as the total weight in the current stratum, w_{total} is less than w_{max}, we continue adding proposals to the current stratum. If adding the next proposal, x, will make the total weight in the stratum larger than w_{max}, then we add the proposal to the current stratum with probability

$$\frac{w_{total} + w(x) - w_{max}}{w(x)}$$

Otherwise, we create a new stratum and add x to the new stratum.

The resulting strata from this approach will, in general, have more equal weight sums than the strata created by uniform stratification. The algorithm also guarantees that each stratum will have at least one proposal. We have found that this stratification leads to effective variance reduction.

4.6 RIS with Multiple Distributions

In this section we describe how to use multiple distributions with Resampled Importance Sampling. It is often the case that we can generate proposals from a number of different distributions. If so, we can increase the robustness of the estimate, and in many cases reduce the variance, by using Multiple Importance Sampling.

If the proposals are all generated in the same domain, or can be easily transformed into a common domain, then we can apply MIS to weight the proposals. This method produces very good results. If the transformation into a common domain is difficult (computationally expensive), then it is not efficient to use MIS on the proposals. However, we can still apply MIS to improve the weighting of the samples.

4.6.1 Multiple Importance Sampling applied to Proposals

Assume we have a set of K proposal densities, $\{p_1, ..., p_K\}$ and a sampling density $\hat{q}$ which is defined over a common domain. Instead of drawing all proposals from a single density, we draw m_i proposals from density p_i. We then use an MIS heuristic to modify the weight, $w(x)$ computed for each of the proposals. For example, with the balance heuristic, the weight for a proposal, x_j, drawn from density, p_i, would be

$$w(x_j) = \frac{m_i p_i(x_j)}{\sum_{k=1}^{K} m_k p_k(x_j)} \frac{\hat{q}(x_j)}{p_i(x_j)}$$

Any MIS heuristic could be used in this context. This new weight is used in all subsequent steps of Resampled Importance Sampling. When performing the resampling step it is not necessary to consider the original proposal densities. Thus, a single sample may be drawn from a set of proposals that initially came from different densities.

4.6.2 Multiple Importance Sampling applied to Samples

If the proposals cannot be easily transformed into the same domain, then we cannot use a common sampling density, $\hat{q}$, and MIS cannot be applied to the proposals. However, we can still apply MIS to the samples.

In this case, we have a set of sampling densities, $\{\hat{q}_1, ...\hat{q}_K\}$, each defined on a different domain. We implement MIS by generating samples from each of the sampling densities, then weighting the samples.

Since the samples are not in the same domain they must be generated from independent sets of proposals. This implies that we must repeat the RIS process K times, once for each density. To do so, we divide the total number of proposals, M, and samples, N, between the K repetitions. Thus, each repetition generates n_i samples from $\hat{q}_i$ using m_i proposals, where $\sum_{i=1}^{K} m_i = M$ and $\sum_{i=1}^{K} n_i = N$. The samples can then be combined using any of the MIS heuristics. Importantly, the MIS weights are computed using the sampling densities, $\{\hat{q}_1, ...\hat{q}_K\}$, not the proposal densities.

4.7 Comparison to Related Work

In this section we compare Resampled Importance Sampling to *Bidirectional Importance Sampling* (BIS) a similar technique developed independently by Burke [4, 5]. Burke suggests two forms of BIS, one based on rejection sampling, the second based on Sampling-Importance Resampling. We only compare RIS to the latter.

The primary differences between our algorithm and Burke's are scope and formality. BIS is seen primarily as a method for improving Monte Carlo integration when applied to product functions. Specifically, it is developed in the context of the direct lighting problem in global illumination. The use of the term "Bidirectional" derives from this application.

We see RIS as a general Monte Carlo variance reduction technique that can be applied to a wide range of problems, not just to global illumination or to product

functions. To emphasize this generality, we derive the RIS estimate and related results independent of any global illumination application. Also, for this reason we consider *Resampled Importance Sampling* to be a more appropriate name, since it does not refer to any specific problem type.

This generality also allows us to be more formal. Most importantly we are able to prove that RIS is unbiased and derive an expression of its variance. Additionally, we show how to robustly choose the RIS parameters with bounds on the resulting variance. We also combine RIS with stratification and Multiple Importance Sampling—other variance reduction techniques.

Finally, RIS is unbiased in general. BIS, as proposed by Burke, is biased if the range of the integrand is multi-dimensional. Specifically, this is a problem when applied to global illumination problems where the range of the integrand is over multi-dimensional colors.

4.8 Summary

In this chapter we presented Resampled Importance Sampling, a novel variance reduction technique that uses Sampling Importance Resampling as the sample generation technique. We showed that it is unbiased and derived a variance expression for it. We also showed how to combine RIS with MIS and stratification to improve the variance reduction.

The variance reduction from RIS results from reducing the number of samples to increase the number of proposals taken. If $\hat{q}$ is a better approximation to f than p, then computing more proposals will reduce the variance. However, it is only efficient to do so if the computation time for proposals is less than the computation time for samples.

Since RIS places no restrictions on $\hat{q}$ it is often quite easy to find a $\hat{q}$ that is much better than the best available p. The requirement that proposals must be much

cheaper to compute than samples is more difficult to fulfill. This serves as an indicator as to which problems will benefit from RIS.

Chapter 5

Results

In this chapter we validate the algorithms derived in the previous chapter by applying them to global illumination problems. We show that RIS can lead to robust variance reduction.

In Sections 5.1 through 5.4 we use RIS to improve the direct lighting calculations in global illumination. We use this application to explore the practical implication of the theories discussed in the previous chapter. Importantly, we demonstrate that the approximate optimal values of M and N derived in Section 4.4 lead to effective variance reduction over standard importance sampling. We also show that the stratification and multiple importance sampling strategies introduced lead to additional variance reduction.

In Section 5.5 we apply RIS to the problem of sampling Bidirectional Reflectance Distribution Functions (BRDFs). In this section we demonstrate that RIS is significantly more robust than standard importance sampling. Although, in this case, RIS cannot outperform specialized implementations of standard importance sampling, we suggest using RIS as a default sampling strategy when a specialized implementation is not yet available.

5.1 Direct Lighting

In general, the direct lighting problem refers to finding the total light that arrives at a point, x' directly from light sources (without bouncing off intermediate surfaces). Often, in global illumination, we're interested in a more specific form of the problem. We want to know how much direct light is reflected from the point, x', in the direction of a second point, x''. Thus, after leaving point x', the light will have bounced exactly one time. We can express this form of the direct lighting problem as an integral over all the points on light sources.

$$L_o(x' \to x'') = \int f_s(x \to x' \to x'')G(x \leftrightarrow x')V(x \leftrightarrow x')L_e(x \to x')dx$$

where f_s is a reflectance term, G is a geometry term, V is a binary visibility term, and L_e is the emitted light. The variable of integration, x, is a point on a light source.

When evaluating this equation in a Monte Carlo path tracer, x' and x'' are already known. Thus the Monte Carlo integration process involves randomly picking sample points, x_i, on light sources and evaluating the Monte Carlo integration estimate

$$L_o(x' \to x'') \approx \frac{1}{N}\sum_{i=1}^{N} \frac{f_s(x_i \to x' \to x'')G(x_i \leftrightarrow x')V(x_i \leftrightarrow x')L_e(x_i \to x')}{p(x_i)}$$

To apply RIS to this problem we need to choose a sampling density $\hat{q}$. As discussed in Section 4.3 the sampling density should be similar to the function that we are integrating and it should be relatively cheap to compute.

The visibility function is usually the most expensive part of the direct lighting equation, so a reasonable choice for $\hat{q}$ is

$$\hat{q} = |f_s(x \to x' \to x'')G(x \leftrightarrow x')L_e(x \to x')|$$

Since the range of the integrand is typically n-component color values, we must use a length function to convert the color to a scalar value. We simply use the luminance of the color. Note that because of the length function, $\hat{q}$ will scale, but not completely cancel out the terms of the integrand.

In the RIS framework we can see that it would be equally valid to approximate the integrand in many other ways. For example, as the computational expense of evaluating f_s or L_e increases, due to more physically realistic surface models or to the calculation of these terms in complex shader programs, it may be more efficient to use

$$\hat{q} = |f'_s(x \to x' \to x'')G(x \leftrightarrow x')V(x \leftrightarrow x')L'_e(x \to x')|$$

where f'_s and L'_e are computationally inexpensive approximations to the true terms.

5.2 Choosing the Number of Samples and Proposals

As we discussed in the previous chapter, the variance reduction from RIS results from reducing the number of samples that are taken in order to increase the number of proposals. If $\hat{q}$ is a good match for f, then computing more proposals can decrease the variance over standard importance sampling.

Figure 5.1 demonstrates the effect of trading of samples for proposals. In the left image, $M = N$, corresponding to standard importance sampling. In the right image, only a single sample is taken, the rest are traded to increase the number of proposals. The right image looks superior to the left in non-shadow regions. This is because $\hat{q} = f$ in these areas. In the shadow regions, however, the right image looks worse. This is because the visibility is only computed once (for the sample), thus its variance is high.

By measuring the variances $V\left(\frac{f(Y)}{q(Y)}\right)$ and $V\left(\frac{f(X)}{p(X)}\right)$, we can apply Equation (4.3) to determine the optimal ratio of M and N for each pixel (see Figure 5.2). However,

Figure 5.1: Dragon sampled with a single primary ray and using RIS with different values of M and N to compute direct lighting (in equal time). On the left, $N = 20$, $M = 20$ and on the right $N = 1$, $M = 60$. The right image has less variance except where visibility is a major component of the variance.

as noted earlier, estimating the variances is very computationally expensive. Instead, we use the approximation given by Equation (4.4). As described in Section 4.4, we must first approximate T_1 and T_2. To do this, we cast a few thousand primary rays. We then track the time necessary to compute the direct lighting at the first hit point. T_1 is the average time necessary to sample the light source and compute $\hat{q}$. T_2 is the average time to check the visibility. The time necessary to estimate these values is negligible.

Across scenes of similar complexity, the values of T_1 and T_2 will probably be quite stable. Thus, these values could be precomputed for a particular renderer implementation. If precomputed, robust RIS would require absolutely no extra computation time over standard importance sampling.

The images in Figure 5.3 show a Thai statue lit by an environment map. The left image uses standard Monte Carlo integration. The right uses the robust values of M

Figure 5.2: The optimal ratio of M and N. Green (lighter) pixels correspond to larger M, red (darker) pixels to larger N.

and N computed from estimated T_1 and T_2 values. In this scene we achieve a 33% reduction in overall variance compared to just using standard Monte Carlo.

5.3 Stratification

This section applies the stratification techniques discussed in Section 4.5 to the direct lighting problem. We show that stratification results in additional variance reduction. For comparison, the top-left image in Figure 5.4 is rendered using RIS without any stratification.

We first stratify the proposals using the standard Latin Hypercube technique. This will directly improve the placement of the proposals in the domain, but will only indirectly improve the placement of the samples. The top-right image in Figure 5.4 shows the results of only stratifying the proposals. The variance in the directly lit areas is reduced significantly. However, the variance in shadowed regions is not largely effected. This is because the evaluation of the proposals does not include the visibility test, so stratifying the proposals does not directly reduce the variance due to visibility.

We can improve the result further by stratifying the samples. This should reduce

Figure 5.3: Direct lighting comparison of RIS and standard Monte Carlo. Both images are rendered using 10 primary rays and approximately equal computation times. The left image uses standard Monte Carlo with 100 shadow rays. The right image uses RIS with the computed robust values from Equation (4.4), $M = 229$, $N = 64$. There is a 33% reduction in variance.

the variance in the shadowed regions. The bottom images in Figure 5.4 were rendered using Latin Hypercube stratification for the proposals with some form of stratification for the samples. The bottom-left image uses the equal-proposals stratification. The bottom-right image uses the equal-weights stratification. Note that both techniques produce additional variance reduction over just stratifying the proposals, especially in the shadow regions.

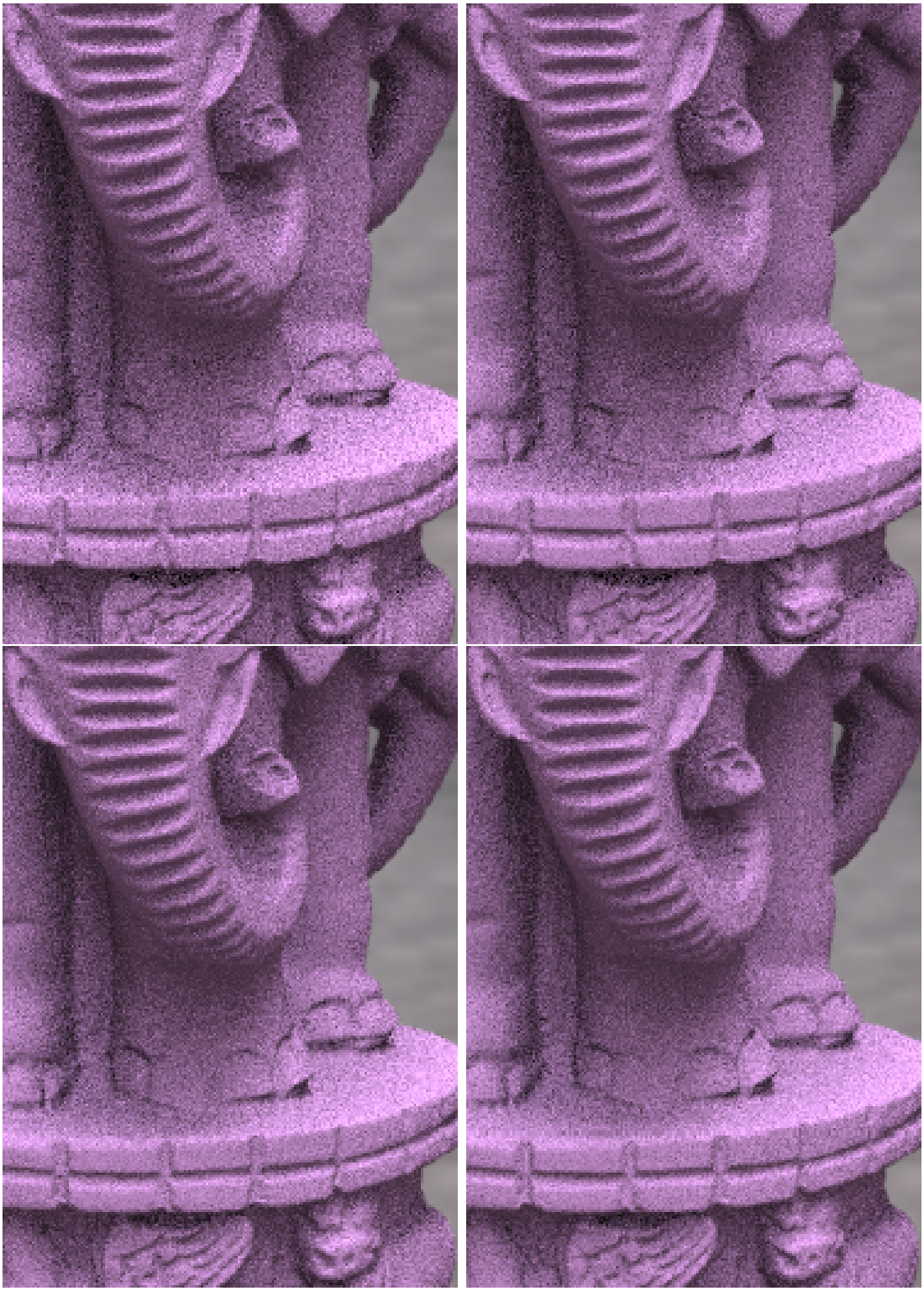

Figure 5.4: Stratification for RIS. Top-left: no stratification. Top-right: proposals stratified using Latin Hypercube (34% variance reduction). Bottom-left image: proposals stratified using Latin Hypercube, samples stratified using equal-proposals stratification (37% variance reduction). Bottom-right: proposals stratified using Latin Hypercube, samples stratified using equal-weights technique (42% variance reduction).

5.4 Multiple Importance Sampling

In this section we apply the Multiple Importance Sampling variance reduction technique to the RIS proposals as described in Section 4.6. The results are similar to those achieved when MIS is applied to standard Monte Carlo integration with some additional variance reduction due to RIS.

Figure 5.5 demonstrates the effectiveness of MIS in a direct lighting application. When combining RIS with MIS (the lower-right image), half of the proposals are generated proportional to the BRDF and half are generated proportional to the environment map. To compute the proposal weights (which included $\hat{q}$), we must place them all in the same domain. To do this we trace a ray in the direction sampled from the BRDF, but only intersect the ray with light sources. Since the number of light sources in a scene will almost always be significantly less than the total number of possible occluders this is reasonably fast. Once we find the nearest light source, we can easily transform the BRDF proposal into the light source domain. This allows us to compute the same $\hat{q}$ for all the proposals.

Samples are then drawn from this set of weighted proposals using the equal-weights stratification method. Each stratum may be made up of proposals from either or both proposal distributions.

In Figure 5.5, notice that the MIS images have lower overall variance than either the BRDF sampling or the light source sampling. The difference between the images with and without RIS is not as drastic. However, RIS reduces the variance in the BRDF sampling case by 60%, in the light source sampling case by 5%, and in the MIS case by 30%.

5.5 Robust Sampling of BRDFs

In the previous sections we showed that RIS can produce better results than standard importance sampling. In this section we demonstrate that RIS can be more

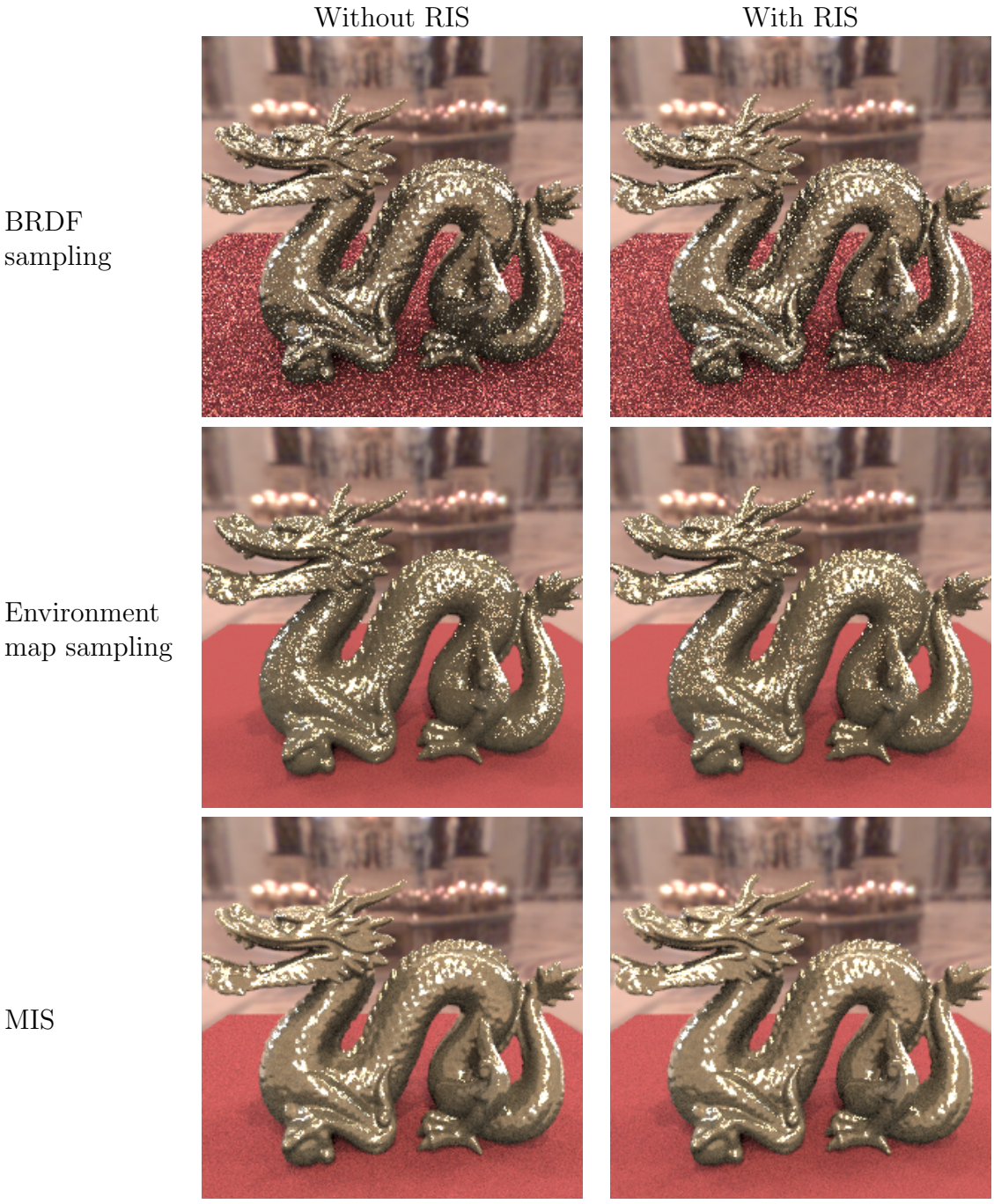

Figure 5.5: Multiple Importance Sampling for RIS proposals. Multiple Importance Sampling, when applied to the RIS proposals, produces similar variance reduction and increased robustness as when applied to standard Monte Carlo integration.

robust than importance sampling. Specifically, we show that RIS can be applied to a wide class of integration problems *without* change. Importance sampling, on the other hand, requires a special case for each specific problem. We use the example of Bidirectional Reflectance Distribution Function (BRDF) sampling to demonstrate.

BRDFs represent the relationship between incident and exitant light at a surface. Traditionally, BRDFs have been sampled with special case distributions developed for each specific BRDF model. Implementing all of these in a global illumination renderer can be very time consuming. This approach can also become unwieldy when the parameters of the BRDF are allowed to vary spatially (as in Bidirectional Texture Distribution Functions). Furthermore, if the BRDF is specified using a shading language, automatically creating a distribution would be difficult.

We would like to find a more robust solution. Ideally, it will improve the sampling of any BRDF model whether or not a good distribution is available for importance sampling. Resampled Importance Sampling can do this.

When sampling BRDFs, we want to sample

$$f = f_s\left(x, \omega_i \rightarrow \omega_o\right) L_i(x, \omega_i) \left|\omega_i \cdot N_x\right|$$

where f_s is the value of the BRDF at point x and L_i is the light arriving at point x from direction ω_i.

To use RIS we need to choose a sampling density $\hat{q}$ that is both closely proportional to f and is inexpensive to compute. Here we will take a common approach and choose

$$\hat{q} = \left| f_s\left(x, \omega_i \rightarrow \omega_o\right) \left|\omega_i \cdot N_x\right| \right|$$

We have dropped the L_i term since it is computationally expensive.

Note that in choosing $\hat{q}$ we have only assumed that the BRDF can be evaluated, which is required for Monte Carlo integration anyway. We did not have to condition the choice on the specific BRDF that is being sampled. Since this choice of $\hat{q}$ works for any BRDF, we can implement RIS once for all BRDF sampling.

In the following example, we recognize that standard Monte Carlo integration with importance sampling could perform much better than the results we show. However, that would require a specialized sampling distribution for each BRDF. Our goal is to show that RIS is more robust since a single implementation can dramatically improve the sampling of very different BRDF models.

Figure 5.6 shows two pairs of spheres each sampled with p equal to a uniform distribution over the hemisphere. The first two spheres have a diffuse BRDF. Our chosen p matches the BRDF exactly, but does not take into account the cosine of the angle with the normal. RIS reduces the variance significantly. The second two spheres have a Cook-Torrance BRDF. In this case, p is a very poor sampling density. Nevertheless, RIS still manages to dramatically improve the sampling quality. We emphasize that no changes are made to the RIS implementation for either BRDF. Achieving similar results with standard importance sampling would require a sampling density selected specifically for each BRDF.

This example suggests another possible use for Resampled Importance Sampling. Since RIS will improve BRDF sampling for any BRDF, it can be used as a default sampling strategy when a specialized BRDF sampling technique is not available. This can greatly improve image quality for a very small implementation cost.

In these examples we used a uniform distribution for p for simplicity. In practice, a better default density would be a cosine-weighted hemisphere with lobes in the reflective and retroreflective directions.

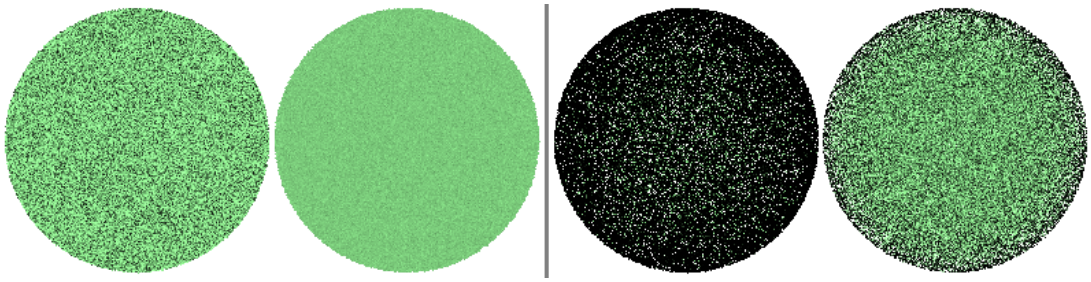

Figure 5.6: Uniformly-lit spheres sampled with a uniform hemispherical distribution. The first two are perfectly diffuse and the second two use a Cook-Torrance BRDF. The first sphere in each pair is rendered without resampling ($N = 1, M = 1$). The second sphere in each pair is rendered with RIS ($N = 1, M = 20$). RIS greatly reduces the variance independent of the BRDF model used.

5.6 Summary

In this chapter we applied Resampled Importance Sampling to two problems in global illumination, direct lighting and BRDF sampling. We showed that, in the case of direct lighting, Resampled Importance Sampling leads to significant variance reduction. We also suggested that, in specific cases where RIS does not do better than standard importance sampling, RIS can serve as an easy to implement default variance reduction technique. This is especially true in cases where a specialized importance sampling density is not generally available.

Chapter 6

Conclusions

We have presented a simple explanation of importance resampling. We have shown how to use importance resampling as the sample generation technique for Monte Carlo integration. The resulting variance reduction technique, Resampled Importance Sampling, is a generalization of standard importance sampling.

We have shown how to compute the optimal resampling parameters, M and N, and how to choose robust approximations of M and N that require significantly less computation time. We have shown how to combine RIS with stratified sampling and with Multiple Importance Sampling.

We have shown that RIS increases the robustness of importance sampling when used in some global illumination problems and we have achieved up to a 33% variance reduction for the direct lighting problem in complex scenes. We have suggested using RIS as a default variance reduction technique when more specialized techniques are not available.

RIS is especially useful when applied to problems that can be factored into two portions that differ greatly in computational cost. The cheaper portion of the problem is evaluated using the proposals. The more expensive portion is evaluated using the

samples. In such problems, computing fewer samples, in order to compute more proposals, makes sense.

Further work needs to be done to find other applications of RIS in Global Illumination. Also, since RIS is a general variance reduction technique, it may find even better uses on problems outside of Global Illumination.

More work needs to be done to find good choices for $\hat{q}$ and p. Unexpectedly, drawing proposals from p accounts for the majority of time to compute the proposals. A lot of work has gone into speeding up visibility computations [29], but not much work has gone into finding faster sample generation techniques. Since RIS uses multiple proposals at a time, techniques that generate proposals in parallel (perhaps using SSE) could be very useful.

Although we have found that our approximate value for the optimal ratio of proposals to samples, K^*, works well, it could be profitable to try to improve the approximation through an adaptive process.

Appendix A

Proofs

A.1 Proof of RIS Unbiasedness

In this section we prove that the RIS estimator, Equation (4.1), is unbiased. Using the identity

$$E_{XY}\left(XY\right) \;=\; E_X\left(E_Y\left(Y|X\right)\cdot X\right) \tag{A.1}$$

we find that

$$E\left(\hat{I}_{ris}\right) \;=\; E\left[E\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(Y_i\right)}{\hat{q}\left(Y_i\right)}\;\middle|\;X_1,..,X_M\right)\cdot\frac{1}{M}\sum_{j=1}^{M}w(X_j)\right]$$

Since the samples Y_i are identically distributed we can pull the sum out of the conditional expectation.

$$E\left(\hat{I}_{ris}\right) \;=\; E\left[E\left(\frac{f\left(Y\right)}{\hat{q}\left(Y\right)}\;\middle|\;X_1,..,X_M\right)\cdot\frac{1}{M}\sum_{j=1}^{M}w(X_j)\right]$$

The sample Y is drawn from a discrete distribution over the proposals, $\{X_1, ..., X_M\}$ so we can expand the conditional expectation as a sum:

$$\begin{aligned} E\left(\hat{I}_{ris}\right) &= E\left[\sum_{k=1}^{M}\left(\frac{f\left(X_k\right)}{\hat{q}\left(X_k\right)}\cdot\frac{w(X_k)}{\sum_{j=1}^{M}w(X_j)}\right)\cdot\frac{1}{M}\sum_{j=1}^{M}w(X_j)\right] \\ &= E\left[\frac{1}{M}\sum_{k=1}^{M}\left(\frac{f\left(X_k\right)}{\hat{q}\left(X_k\right)}\cdot w(X_k)\right)\right] \\ &= E\left(\frac{f\left(X\right)}{\hat{q}\left(X\right)}\cdot w(X)\right) \end{aligned}$$

Using our chosen weighting function $w(X) = \frac{\hat{q}(X)}{p(X)}$ and the fact that the proposals are distributed with density p we can trivially show unbiasedness.

A.2 Proof of RIS Variance

To prove Equation (4.2) we use Equation A.1 and the basic definition of variance:

$$V(XY) = E\left(X^2Y^2\right) - E\left(XY\right)^2$$

to derive

$$\begin{aligned} V(\hat{I}_{ris}) &= V\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(Y_i\right)}{\hat{q}\left(Y_i\right)}\ \cdot\ \frac{1}{M}\sum_{j=1}^{M}w(X_j)\right) \\ &= E\left\{E\left(\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(Y_i\right)}{\hat{q}\left(Y_i\right)}\right)^2\ \middle|\ X_1, ..., X_M\right)\cdot\left(\frac{1}{M}\sum_{j=1}^{M}w(X_j)\right)^2\right\} \\ &\quad - E\left\{E\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(Y_i\right)}{\hat{q}\left(Y_i\right)}\ \middle|\ X_1, ..., X_M\right)\cdot\left(\frac{1}{M}\sum_{j=1}^{M}w(X_j)\right)\right\}^2 \end{aligned}$$

Note that since the Y_i are correlated we cannot directly simplify the variance term. We will reduce the two portions of this equation separately. To reduce the first part

we use the identity

$$E\left(\left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)^2\right)=\frac{1}{N}E\left(X^2\right)+\frac{(N-1)}{N}E\left(X\right)^2$$

to expand the square inside the conditional expectation. We can then replace the resulting expectations with summations.

$$\begin{aligned}
&E\left\{E\left(\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(Y_i\right)}{\hat{q}\left(Y_i\right)}\right)^2\;\middle|\;X_1,...,X_M\right)\cdot\left(\frac{1}{M}\sum_{j=1}^{M}w(X_j)\right)^2\right\}\\
&=E\left\{\left(\frac{1}{N}E\left(\frac{f\left(Y\right)^2}{\hat{q}\left(Y\right)^2}\;\middle|\;X_1,...,X_M\right)+\frac{(N-1)}{N}E\left(\frac{f\left(Y\right)}{\hat{q}\left(Y\right)}\;\middle|\;X_1,...,X_M\right)^2\right)\right.\\
&\quad\left.\cdot\left(\frac{1}{M}\sum_{j=1}^{M}w(X_j)\right)^2\right\}\\
&=E\left\{\left(\frac{1}{N}\sum_{k=1}^{M}\left(\frac{f\left(X_k\right)^2}{\hat{q}\left(X_k\right)^2}\frac{w(X_k)}{\sum_{j=1}^{M}w(X_j)}\right)+\frac{(N-1)}{N}\left(\sum_{k=1}^{M}\frac{f\left(X_k\right)}{\hat{q}\left(X_k\right)}\frac{w(X_k)}{\sum_{j=1}^{M}w(X_j)}\right)^2\right)\right.\\
&\quad\left.\cdot\left(\frac{1}{M}\sum_{j=1}^{M}w(X_j)\right)^2\right\}\\
&=\frac{1}{M^2N}E\left(\sum_{k=1}^{M}\frac{f\left(X_k\right)^2}{\hat{q}\left(X_k\right)^2}w(X_k)\cdot\sum_{j=1}^{M}w(X_j)\right)+\frac{(N-1)}{M^2N}E\left(\left(\sum_{k=1}^{M}\frac{f\left(X_k\right)}{\hat{q}\left(X_k\right)}w(X_k)\right)^2\right)\\
&=\frac{1}{M^2N}\left(ME\left(\frac{f(X)^2}{\hat{q}(X)^2}w(X)^2\right)+M(M-1)E\left(\frac{f(X)^2}{\hat{q}(X)^2}w(X)\right)E\left(w(X)\right)\right)\\
&\quad+\frac{(N-1)}{M^2N}\left(ME\left(\frac{f(X)^2}{\hat{q}(X)^2}w(X)^2\right)+M(M-1)E\left(\frac{f(X)}{\hat{q}(X)}w(X)\right)^2\right)\\
&=\frac{1}{MN}\left[E\left(\frac{f(X)^2}{p(X)^2}\right)+(M-1)E\left(\frac{f(X)^2}{\hat{q}(X)p(X)}\right)E\left(\frac{\hat{q}(X)}{p(X)}\right)\right.\\
&\quad\left.+(N-1)E\left(\frac{f(X)^2}{p(X)^2}\right)+(M-1)(N-1)E\left(\frac{f(X)}{p(X)}\right)^2\right]
\end{aligned}$$

The final step results from substituting for the weight function, $w(x)=\frac{\hat{q}(x)}{p(x)}$.

The second part of the variance equation can be reduced easily using the result of Appendix A.1:

$$E\left\{E\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(Y_i\right)}{\hat{q}\left(Y_i\right)}\,\middle|\,X_1,...,X_M\right)\cdot\left(\frac{1}{M}\sum_{j=1}^{M}w(X_j)\right)\right\}^2 = E\left(\frac{f(X)}{p(X)}\right)^2$$

Combining the two parts of the equation gives us

$$V(\hat{I}_{ris}) = \frac{1}{MN}\left[E\left(\frac{f(X)^2}{p(X)^2}\right) + (M-1)E\left(\frac{f(X)^2}{\hat{q}(X)p(X)}\right)E\left(\frac{\hat{q}(X)}{p(X)}\right) + (N-1)E\left(\frac{f(X)^2}{p(X)^2}\right)\right.$$
$$\left. +(M-1)(N-1)E\left(\frac{f(X)}{p(X)}\right)^2 - MNE\left(\frac{f(X)}{p(X)}\right)^2\right]$$

By rearranging and combining terms we can simplify this.

$$V(\hat{I}_{ris}) = \frac{1}{MN}\left[NV\left(\frac{f(X)}{p(X)}\right) + (M-1)\left(E\left(\frac{f(X)^2}{\hat{q}(X)^2p(X)^2}\right)E\left(\frac{\hat{q}(X)}{p(X)}\right) - E\left(\frac{f(X)}{p(X)}\right)^2\right)\right]$$

To reduce this further it is necessary to recognize that

$$E\left(\frac{\hat{q}(X)}{p(X)}\right) = \int \hat{q}$$

and

$$E\left(\frac{f\left(X\right)^2}{\hat{q}\left(X\right)p\left(X\right)}\right) = \int \frac{f^2}{\hat{q}}$$

Thus,

$$E\left(\frac{f\left(X\right)^2}{\hat{q}\left(X\right)p\left(X\right)}\right)E\left(\frac{\hat{q}(X)}{p(X)}\right) = \int \frac{f^2}{\frac{\hat{q}}{\int\hat{q}}}$$
$$= E\left(\frac{f(Y)^2}{q(Y)^2}\right)$$

and,

$$E\left(\frac{f(X)}{p(X)}\right) = \int f$$
$$= E\left(\frac{f(Y)}{q(Y)}\right)$$

where Y is a random variable with density $q = \frac{\hat{q}}{\int \hat{q}}$. Therefore we can express the variance as a function of two random variables, X and Y with densities p and q respectively.

$$\frac{1}{MN}\left[NV\left(\frac{f(X)}{p(X)}\right) + (M-1)V\left(\frac{f(Y)}{q(Y)}\right)\right]$$

A little more rearranging produces the form given in Equation (4.2).

A.3 Proof of Optimal K

In this section we prove Equation (4.3) by minimizing the RIS variance (Equation 4.2)

$$V\left(\hat{I}_{ris}\right) = \frac{1}{M}V\left(\frac{f(X)}{p(X)}\right) \; + \; \left(1-\frac{1}{M}\right)\frac{1}{N}V\left(\frac{f(Y)}{q(Y)}\right)$$

under the constraint that

$$T_{total} = M\ T_1 \; + \; N\ T_2$$

To simplify this process we make the substitution

$$K = \frac{M}{N}$$

This leads to

$$V\left(\hat{I}_{ris}\right) = \frac{1}{M}V\left(\frac{f(X)}{p(X)}\right) \; + \; \left(1-\frac{1}{M}\right)\frac{K}{M}V\left(\frac{f(Y)}{q(Y)}\right)$$

with the modified constraint that

$$T_{total} = M\ T_1\ +\ \frac{M}{K}\ T_2$$

We solve for M in the constraint and directly substitute it into the variance expression

$$V\left(\hat{I}_{ris}\right) = \frac{T_1 + \frac{T_2}{K}}{T} V\left(\frac{f(X)}{p(X)}\right)\ +\ \left(1 - \frac{T_1 + \frac{T_2}{K}}{T}\right) K \frac{T_1 + \frac{T_2}{K}}{T} V\left(\frac{f(Y)}{q(Y)}\right)$$

We take the derivative and set it equal to zero to find the minimum.

$$\begin{aligned}
& -\frac{T_2}{K^2 T} V\left(\frac{f(X)}{p(X)}\right) + \left(\frac{T_1}{T} - \frac{{T_1}^2}{T^2} + \frac{{T_2}^2}{K^2 T^2}\right) V\left(\frac{f(Y)}{q(Y)}\right) \stackrel{\text{set}}{=} 0 \\
\iff \qquad & K^2 \left(\frac{T_1}{T} - \frac{{T_1}^2}{T^2}\right) V\left(\frac{f(Y)}{q(Y)}\right) = \frac{T_2}{T} V\left(\frac{f(X)}{p(X)}\right) - \frac{{T_2}^2}{T^2} V\left(\frac{f(Y)}{q(Y)}\right) \\
\iff \qquad & K^2 = \frac{T_2 \left(V\left(\frac{f(X)}{p(X)}\right) - \frac{T_2}{T} V\left(\frac{f(Y)}{q(Y)}\right)\right)}{T_1 \left(V\left(\frac{f(Y)}{q(Y)}\right) - \frac{T_1}{T} V\left(\frac{f(Y)}{q(Y)}\right)\right)}
\end{aligned}$$

This leads directly to Equation (4.3).

A.4 Proof of Variance Bound on K^*

In Section 4.4 we claimed that using K^* instead of K would result in no more than twice the variance. In this section we prove that result.

Using

$$K^* = \frac{T_2}{T_1}$$

results in evenly splitting the time between proposals and samples. Thus,

$$M^* = \frac{\frac{T}{2}}{T_1}$$

and

$$N^* = \frac{\frac{T}{2}}{T_2}$$

Since decreasing M or N will not decrease the variance, we only consider the case of increasing M or N. Within the overall time constraint the maximum possible values for M and N are

$$M_{max} = 2M^*$$

and

$$N_{max} = 2N^*$$

This follows directly from the fact that M^* and N^* use half of the total time.

We will split the proof into two cases. First, we prove that using M^* instead of M_{max} can no more than double the variance:

$$\begin{aligned}
&\frac{1}{M^*}V\left(\frac{f(X)}{p(X)}\right) + \left(1 - \frac{1}{M^*}\right)\frac{1}{N}V\left(\frac{f(Y)}{q(Y)}\right) \\
&\qquad = \frac{2}{M_{max}}V\left(\frac{f(X)}{p(X)}\right) + \left(1 - \frac{2}{M_{max}}\right)\frac{1}{N}V\left(\frac{f(Y)}{q(Y)}\right) \\
&\qquad = 2\left(\frac{1}{M_{max}}V\left(\frac{f(X)}{p(X)}\right) + \left(\frac{1}{2} - \frac{1}{M_{max}}\right)\frac{1}{N}V\left(\frac{f(Y)}{q(Y)}\right)\right) \\
&\qquad \leq 2\left(\frac{1}{M_{max}}V\left(\frac{f(X)}{p(X)}\right) + \left(1 - \frac{1}{M_{max}}\right)\frac{1}{N}V\left(\frac{f(Y)}{q(Y)}\right)\right)
\end{aligned}$$

We follow a similar approach to show that using N^* instead of N_{max} will no more than double the variance:

$$\begin{aligned}
&\frac{1}{M}V\left(\frac{f(X)}{p(X)}\right)+\left(1-\frac{1}{M}\right)\frac{1}{N^*}V\left(\frac{f(Y)}{q(Y)}\right)\\
&\qquad=\frac{1}{M}V\left(\frac{f(X)}{p(X)}\right)+\left(1-\frac{1}{M}\right)\frac{2}{N_{max}}V\left(\frac{f(Y)}{q(Y)}\right)\\
&\qquad=2\left(\frac{1}{2M}V\left(\frac{f(X)}{p(X)}\right)+\left(1-\frac{1}{M}\right)\frac{1}{N_{max}}V\left(\frac{f(Y)}{q(Y)}\right)\right)\\
&\qquad\leq 2\left(\frac{1}{M}V\left(\frac{f(X)}{p(X)}\right)+\left(1-\frac{1}{M}\right)\frac{1}{N_{max}}V\left(\frac{f(Y)}{q(Y)}\right)\right)
\end{aligned}$$

This concludes our proof.

Bibliography

[1] Sameer Agarwal, Ravi Ramamoorthi, Serge Belongie, and Henrik Wann Jensen. Structured importance sampling of environment maps. *ACM Trans. Graph.*, 22(3):605–612, 2003.

[2] Benedek Balázs, László Szirmay-Kalos, and Antal György. Weighted importance sampling in shooting algorithms. In *SCCG '03: Proceedings of the 19th Spring Conference on Computer Graphics*, pages 177–184, New York, NY, USA, 2003. ACM Press.

[3] Philippe Bekaert, Mateu Sbert, and Yves D. Willems. Weighted importance sampling techniques for Monte Carlo radiosity. In *Proceedings of the 11th Eurographics Workshop on Rendering*, pages 35–46. Eurographics, 2000.

[4] David Burke. Bidirectional importance sampling for illumination from environment maps. Master's thesis, University of British Columbia, 2004.

[5] David Burke, Abhijeet Ghosh, and Wolfgang Heidrich. Bidirectional importance sampling for direct illumination. In Kavita Bala and Philip Dutré, editors, *Rendering Techniques 2005 Eurographics Symposium on Rendering*, pages 147–156, Aire-la-Ville, Switzerland, 2005. Eurographics Association.

[6] Yuguo Chen. Another look at rejection sampling through importance sampling. *Statistics and Probability Letters*, 72:277–283, 2005.

[7] Petrik Clarberg, Wojciech Jarosz, Tomas Akenine-Möller, and Henrik Wann Jensen. Wavelet importance sampling: Efficiently evaluating products of complex functions. In *Proceedings of ACM SIGGRAPH 2005*. ACM Press, July 2005.

[8] Ferenc Csonka, László Szirmay-Kalos, and György Antal. Generalized multiple importance sampling for Monte-Carlo global illumination. *Machine Graphics & Vision International Journal*, 11(1):3–20, 2002.

[9] Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2000.

[10] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, London, second edition, 2004.

[11] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In *Readings in Uncertain Reasoning*, pages 452–472. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990.

[12] Tim Hesterberg. Weighted average importance sampling and defensive mixture distributions. *Technometrics*, 37(2):185–194, May 1995.

[13] David S. Immel, Michael F. Cohen, and Donald P. Greenberg. A radiosity method for non-diffuse environments. In David C. Evans and Russell J. Athay, editors, *SIGGRAPH '86: Proceedings of the 13th Annual Conference on Computer*

Graphics and Interactive Techniques, pages 133–142, New York, NY, USA, 1986. ACM Press.

[14] Henrik Wann Jensen, Stephen R. Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *SIGGRAPH '01: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pages 511–518, New York, NY, USA, 2001. ACM Press.

[15] James T. Kajiya. The rendering equation. In David C. Evans and Russell J. Athay, editors, *SIGGRAPH '86: Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques*, pages 143–150, New York, NY, USA, 1986. ACM Press.

[16] Eric P. Lafortune and Yves D. Willems. Reducing the number of shadow rays in bidirectional path tracing. In V. Skala, editor, *Proceedings of the Winter School of Computer Graphics and CAD Systems '95*, pages 384–392, Plzen, Czech Republic, 1995. University of West Bohemia.

[17] Eric P. Lafortune and Yves D. Willems. Rendering participating media with bidirectional path tracing. In *Proceedings of the Eurographics Workshop on Rendering Techniques '96*, pages 91–100, London, UK, 1996. Springer-Verlag.

[18] Jason Lawrence, Szymon Rusinkiewicz, and Ravi Ramamoorthi. Efficient BRDF importance sampling using a factored representation. *ACM Trans. Graph.*, 23(3):496–505, 2004.

[19] Nicholas Metropolis and Stanislaw Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.

[20] Art Owen and Yi Zhou. Safe and effective importance sampling. *Journal of the American Statistical Association*, 95(449):135–143, March 2000.

[21] Matt Pharr and Pat Hanrahan. Monte Carlo evaluation of non-linear scattering equations for subsurface reflection. In *SIGGRAPH '00: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pages 75–84, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[22] James Gary Propp and David Bruce Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms*, pages 223–252, New York, NY, USA, 1996. John Wiley & Sons, Inc.

[23] Donald B. Rubin. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. Discussion of Tanner and Wong. *Journal of the American Statistical Association*, 82:543–546, 1987.

[24] Peter Shirley, Changyaw Wang, and Kurt Zimmerman. Monte Carlo techniques for direct lighting calculations. *ACM Trans. Graph.*, 15(1):1–36, 1996.

[25] László Szécsi, Mateu Sbert, and László Szirmay-Kalos. Combined correlated and importance sampling in direct light source computation and environment mapping. *Comput. Graph. Forum*, 23(3):585–594, 2004.

[26] Justin F. Talbot, David Cline, and Parris K. Egbert. Importance resampling for global illumination. In Kavita Bala and Philip Dutré, editors, *Rendering Techniques 2005 Eurographics Symposium on Rendering*, pages 139–146, Aire-la-Ville, Switzerland, 2005. Eurographics Association.

[27] Eric Veach. *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, 1997.

[28] Eric Veach and Leonidas J. Guibas. Metropolis light transport. In *SIGGRAPH '97: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, pages 65–76, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[29] Ingo Wald. *Realtime Ray Tracing and Interactive Global Illumination*. PhD thesis, Computer Graphics Group, Saarland University, 2004. Available at http://www.mpi-sb.mpg.de/~wald/PhD/.

[30] Alexander Wilkie, Robert F. Tobler, and Werner Purgathofer. Combined rendering of polarization and fluorescence effects. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*, pages 197–204, London, UK, 2001. Springer-Verlag.