Five Common Misconceptions about Bias in Light Transport Simulation

Toshiya Hachisuka

Aarhus University

Abstract

Monte Carlo integration has been a common approach for light transport simulation over many years. This widespread usage, however, seems to come with certain misconceptions regarding the behavior of error in Monte Carlo algorithms. In particular, unbiased Monte Carlo integration has been blindly (and often implicitly) considered as a holy grail of accurate light transport simulation by some people. The word "unbiased" has been used even as a marketing buzzword for some commercial rendering systems - treating "unbiasedness" as something absolutely good. Such misconceptions now seem to start creeping into the academic research in light transport simulation. This manuscript lists five such misconceptions. We provide theoretical reasoning and some examples in order to clarify each misconception why it is wrong. We believe that the list can be used as teaching notes on the topic of light transport simulation as well as a sanity check for claims of research in light transport simulation.

1. Introduction

Since the formulation the rendering equation [Kaj86], Monte Carlo integration has been the basis of light transport simulation over many years. For example, path tracing [Kaj86] solves the rendering equation by random sampling of light transport paths, where the rendering equation is expanded into an infinite sum of integrals using the Neumann expansion. The path integral formulation [Vea98] provides a mathematically rigorous connection between light transport simulation and Monte Carlo integration.

Due to the use of random numbers, a solution given by Monte Carlo integration is naturally a random variable. The most common approach to quantify the characteristics of a given random variable is to consider its expected value. If the expected value of an estimator is exactly equal to the desired solution of the integral, the estimator is *unbiased*. Otherwise, the estimator is *biased*, where the bias is defined as the difference between the expected value and the correct solution. The computational error of an unbiased estimator is thus solely characterized by the randomness of the solution (i.e., noise), whereas the computational error of a biased estimator is also affected by bias in addition to noise.

One common argument regarding bias in light transport simulation is that unbiased methods are "more accurate" than biased methods. This argument seems logical at a first glance, since error of biased methods contains bias in addition to noise as defined above. This difference in the definitions apparently gives some people a false impression that unbiased estimators should have smaller error. Given this seemingly logical reasoning, the unbiased approach has been sometimes considered as "a holy grail" of accurate light transport simulation, both in some academic research and industrial applications. This argument, however, is at least not exact, given that some biased methods do provide far more accurate solutions than unbiased methods under the same computation time. A prominent example would be irradiance caching [WRC88].

This manuscripts is the author's attempt to resolve the following five common misconceptions related to bias in light transport simulation algorithms:

1. Unbiased methods are more accurate than biased methods;
2. Unbiased methods always converge to the correct solution;
3. Unbiased methods are parameter-free;
4. Quantifying error is possible only with unbiased methods;
5. Markov chain algorithms are unbiased and consistent.

Naturally, the above list is not meant to be complete. We however believe it covers some critical misconceptions that are self-contradictory in the academic community and the industry. It is the author's experience that these points are often misunderstood even among experts of light transport simulation. The list may be useful for teaching the topic of light transport simulation, since these misconceptions are far easier to occur when people are learning this topic. We also believe that the list can serve as a sanity check for claims in new research results in light transport simulation.

)

2. Background

Before diving into the details of each misconception, we briefly review some basic definitions regarding Monte Carlo integration.

2.1. Monte Carlo Integration

The problem we want to solve is the following definite integral over some domain Ω:

$$I = \int_{\Omega} f(\mathbf{x}) d\mathbf{x}, \tag{1}$$

where $\mathbf{x} \in \Omega$ and $f : \Omega \to \mathbb{R}$ is a scalar function. In the context of light transport simulation, Ω is the space of light transport paths, I is the pixel intensity to be computed, and $f(\mathbf{x})$ is the energy transported by the path defined by $\mathbf{x}$. For the brevity of equations, we consider the Lebesgue measure, but the following discussion holds for the path measure in path integral formulation [Vea98].

Monte Carlo integration [MU49] is a numerical integration method that provides an approximate solution of such a definite integral. The method is based on the definition of the expected value of a random variable $f(\mathbf{x})/p(\mathbf{x})$ with probability density function $p(\mathbf{x})$ and its estimation using a finite number of independent samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in \mathbf{X}$:

$$E[f(\mathbf{X})] = \int_{\Omega} f(\mathbf{x}) p(\mathbf{x})\, d\mathbf{x}. \tag{2}$$

Using the definition of expected values, we can formulate the definite integral as the expected value of a random variable $f(\mathbf{x})/p(\mathbf{x})$ with the probability density function $p(\mathbf{x})$;

$$\mathrm{E}\left[\frac{f(\mathbf{X})}{p(\mathbf{X})}\right] = \int_{\Omega} \frac{f(\mathbf{x})}{p(\mathbf{x})} p(\mathbf{x})\, d\mathbf{x} = \int_{\Omega} f(\mathbf{x})\, d\mathbf{x} = I. \tag{3}$$

Note that we need to use $f(\mathbf{x})/p(\mathbf{x})$ as the random variable rather than $f(\mathbf{x})$. Monte Carlo integration estimates this expected value with N samples as

$$I = \mathrm{E}\left[\frac{f(\mathbf{X})}{p(\mathbf{X})}\right] \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f(\mathbf{x}_i)}{p(\mathbf{x}_i)}. \tag{4}$$

2.2. Consistency and Unbiasedness

A Monte Carlo estimator with N samples, $f_N(\mathbf{X})$, is consistent if the estimate converges almost surely to the correct solution with an infinite number of samples. In the context of Monte Carlo integration, a *consistent* estimator satisfies

$$\lim_{N \to \infty} \mathrm{P}\left[\int_{\Omega} f(\mathbf{x})\, d\mathbf{x} - f_N(\mathbf{X}) = 0\right] = 1, \tag{5}$$

where $\mathrm{P}[A]$ returns the probability that the boolean expression A is true. An estimator that does not satisfy the above condition is called an *inconsistent* estimator.

Another important concept is unbiasedness. A Monte Carlo estimator $f_N(\mathbf{X})$ is *unbiased* if the expected value of the estimate is equal to the correct solution. To be concrete, an unbiased estimator satisfy

$$\mathrm{E}[f_N(\mathbf{X})] - I = 0. \tag{6}$$

It is important to note that an unbiased estimator is unbiased *regardless* of the number of samples. A *biased* estimator simply does not satisfy the above condition and the bias $\mathrm{B}[f_N(\mathbf{X})]$ is defined as

$$\mathrm{B}[f_N(\mathbf{X})] = \mathrm{E}[f_N(\mathbf{X})] - I. \tag{7}$$

2.3. Error Estimation

Since every Monte Carlo estimate is a random variable, error is also a random variable. In order to quantify the error of an estimate, we consider the expected squared error $\mathrm{E}[(f_N(\mathbf{X}) - I)^2]$. We can expand this expression by assuming that samples are independent and identically distributed according to $p(\mathbf{X})$,

$$\begin{aligned}
\mathrm{E}\left[(f_N(\mathbf{X}) - I)^2\right] &= \mathrm{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f(\mathbf{x}_i)}{p(\mathbf{x}_i)} + \mathrm{B}[f_N(\mathbf{X})] - I\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f(\mathbf{x}_i)}{p(\mathbf{x}_i)} + \mathrm{B}[f_N(\mathbf{X})]\right)^2\right] - 2I\left(I + \mathrm{B}[f_N(\mathbf{X})]\right) + I^2 \\
&= \frac{1}{N}\mathrm{V}\left[\frac{f(\mathbf{X})}{p(\mathbf{X})}\right] + \left(\mathrm{B}[f_N(\mathbf{X})]\right)^2,
\end{aligned} \tag{8}$$

where $\mathrm{V}[\mathbf{X}]$ returns the variance of $\mathbf{X}$. This result shows that the expected squared error of an estimator vanishes as $O(N^{-1})$ (thus the error vanishes as $O(N^{-0.5})$) if it is unbiased ($\mathrm{B}[f_N(\mathbf{X})] = 0$). This result also shows that bias appears as an additional error in the form of $(\mathrm{B}[f_N(\mathbf{X})])^2$.

There are two remarks on this commonly known "error estimation of MC integration" that are noteworthy. First, the equation is true only if the samples are independent and identically distributed. In many light transport algorithms, such as photon mapping and even regular path tracing, samples are not independent due to path reusing. It is thus generally wrong to use the above equation directly to discuss the average error over the rendered image.

Second, the derivation assumes that $\mathrm{V}[f(\mathbf{X})/p(\mathbf{X})]$ is finite and exists. Again, this is hardly true in light transport simulation, where the function $f(\mathbf{X})/p(\mathbf{X})$ can have discontinuities and singularities, depending on the scene configuration. If the variance is infinite, we can only rely on the the strong law of large numbers that tells us

$$\lim_{N \to \infty} \mathrm{P}\left[E\left[\frac{f(\mathbf{X})}{p(\mathbf{X})}\right] - \frac{1}{N}\sum_{i=1}^{N}\frac{f(\mathbf{x}_i)}{p(\mathbf{x}_i)} = 0\right] = 1. \tag{9}$$

It also means that even the well-known convergence rate $O(N^{-1})$ is *not necessarily applicable* if $\mathrm{V}[f(\mathbf{X})/p(\mathbf{X})]$ is not finite. Such a situation can easily occur, for example, if the geometry term [Vea98] is not importance sampled.

3. Misconceptions

In the following, we describe each misconception and show why they are indeed “misconceptions”. Note that it is not the author’s intention to invalidate scientific contributions of papers which may be making some claims based on the following misconceptions.

3.1. Unbiased methods are more accurate than biased methods

Misconception: The argument behind this claim is that, according to Equation 8, biased methods have additional systematic error as bias to its expected value, whereas unbiased methods has no such additional error.

Explanation: This claim is not accurate for two reasons. Firstly, Equation 8 only discusses *expected* squared error which is not equal to the error of a rendered image for a specific run. An actual rendered image obtained from an unbiased estimator is not the expected value itself, but a random estimate. Secondly, and perhaps more importantly, biased methods could have significantly lower variance than unbiased methods. The claim ignores possible reduction of variance in biased methods. It is very well possible that a biased estimator gives us a solution with a lower actual error with the same number of samples.

To be more precise, consider an unbiased estimator $u_N(\mathbf{X})$ where $\mathrm{B}[u(\mathbf{X})] = 0$ and a biased estimator $b_N(\mathbf{X})$ where $\mathrm{B}[b(\mathbf{X})] \neq 0$. By definition, we have

$$\mathrm{E}\left[(u_N(\mathbf{X}) - I)^2\right] = \frac{1}{N}\mathrm{V}[u(\mathbf{X})], \tag{10}$$

and

$$\mathrm{E}\left[(b_N(\mathbf{X}) - I)^2\right] = \frac{1}{N}\mathrm{V}[b(\mathbf{X})] + (\mathrm{B}[b(\mathbf{X})])^2. \tag{11}$$

However, neither equation supports the following claim of improved accuracy of $u_N(\mathbf{X})$ over $b_N(\mathbf{X})$;

$$\mathrm{P}\left[(u_N(\mathbf{X}) - I)^2 < (b_N(\mathbf{X}) - I)^2\right] > 0.5, \tag{12}$$

which says that the biased estimator $b_N(\mathbf{X})$ can be closer to the correct solution more often than the competing unbiased estimator $u_N(\mathbf{X})$. Neither Equation 10 nor Equation 11 even implies this expected behavior

$$\mathrm{E}\left[(u_N(\mathbf{X}) - I)^2\right] < \mathrm{E}\left[(b_N(\mathbf{X}) - I)^2\right] \tag{13}$$

due to potential difference in variances. If the two estimators have the same variance, then this claim is true. The claim, however, is often used for comparing unbiased methods and biased methods of different characteristics, which is not appropriate. One cannot claim that a newly developed unbiased method is “better” than exiting biased methods just because it is unbiased. For example, progressive photon mapping [HOJ08] is indeed a biased method that was demonstrated to often outperform unbiased methods.

3.2. Unbiased methods always converge to the correct solution

Misconception: Unbiasedness is often considered to ensure convergence to the correct solution. In other words, unbiasedness is tend to be tied with consistency. This argument effectively claims that Equation 6 implies Equation 5.

Explanation: Perhaps somewhat counterintuitive, but an unbiased estimator does not always satisfy Equation 5. Consistency and unbiasedness are in fact independent properties of an estimator. A method can be unbiased and inconsistent at the same time.

To illustrate the independence of consistency and unbiasedness, one can think of a simple experiment of computing the expected value of dice throwing. We can consider the following four estimators for all combinations of consistency and unbiasedness:

- **Consistent and unbiased estimator**:
  The sum of numbers obtained so far divided by the number of throws.
- **Inconsistent and unbiased estimator**:
  The number obtained from the last throw regardless of the preceding throws.
- **Consistent and biased estimator**:
  The sum of numbers obtained so far divided by the number of throws + 1.
- **Inconsistent and biased estimator**:
  The number 4, regardless of the outcomes of the throws.

Notice that there exists an *inconsistent* and *unbiased* estimator even in this simple example. This estimator is unbiased since the expected value of the number obtained from the last throw is equal to the correct value of 3.5. This estimator however is inconsistent because the estimate never converges to this expected value, even with an infinite number of throws.

In the simple example above, it is easy to see that the estimator is not inconsistent. In practice, it is less trivial to check if a new light transport simulation method is consistent or not, and the details will depend on the specific algorithm. The take-home message here is that unbiasedness does not always imply consistency, thus one needs to be careful when making such a claim. We will explain an even more subtle example with Markov chain Monte Carlo sampling for light transport simulation as another misconception.

3.3. Unbiased methods are parameter-free

Misconception: Based on the definition of error in unbiased Monte Carlo integral estimators, it is often claimed that the only parameter of such unbiased estimators is the number of samples, hence unbiasedness is an attractive property for making a rendering algorithm easy to use. This claim often is used as reason for introducing unbiased methods into a movie production, where biased methods with lots of parameters make the whole pipeline difficult to use without understanding each parameter. Likewise, this claim is used as a reasoning to prefer unbiased Monte Carlo methods over biased ones.

Explanation: Even if an estimator is unbiased and consistent, it might have parameters that greatly affects its variance (hence average error). For example, one could think of a two-stage importance sampling algorithm, where the first stage constructs a PDF that is used the second stage to render the actual image. In this case, even though the entire algorithm can be unbiased (and potentially consistent), we need to decide how many samples to draw in each stage and how to construct the PDF. This will result in multiple user-defined parameters that will affect performance.

Whether a particular algorithm has many parameters or not is entirely independent from its unbiasedness. In fact, if estimators are consistent, regardless of their unbiasedness, the result will converge to the same solution simply by increasing the number of samples.

3.4. Quantifying error is possible only with unbiased methods

Misconception: One technical reasoning to prefer unbiased methods over biased methods is that error in unbiased methods is solely characterized by noise, and thus it is easy to quantify. The claim often goes on to say that bias is impossible to quantify, thus one cannot use biased methods for applications of light transport simulation where estimating the accuracy is important.

Explanation: This claim contains a certain truth: error estimation of unbiased methods is readily available under certain conditions. For example, we can derive a probabilistic error bound using the t-distribution. What is often forgotten are the assumptions for such derivation. For the example of the t-distribution, we need to assume that samples are drawn exactly from the normal distribution and variance is finite and exists. These two assumptions are violated in light transport simulation (e.g., due to the geometry term of shadow ray connections). Claiming that error estimation is possible with unbiased estimators is at least misleading since the estimated error can be arbitrary wrong if these assumptions do not hold.

There also exists work that provides quantification of error for biased methods. In particular, the analysis and estimation of error of bias in photon mapping has been explored [HJJ10]. It is therefore possible to quantify error in some biased light transport simulation methods. One might still claim that estimation of bias needs some assumptions, thus this is not a misconception. However, we should keep it in mind that assumptions are made even for error estimation with unbiased estimators, as discussed above. Ultimately, we should always perform numerical experiments to see if error estimations are in fact useful.

3.5. Markov chain algorithms are unbiased and consistent

Misconception: Throughout the literature, it is well recognized that the original Markov chain Monte Carlo method is biased and consistent. The reason is that the distribution of samples converges to the target distribution for infinitely long Markov chains by definition. The difference between the initial distribution and the target distribution is called start-up bias. Veach proposed to eliminate start-up bias in order to make Metropolis light transport (MLT) unbiased. The misconception is that this technique makes MLT unbiased and consistent.

Explanation: The start-up bias elimination in fact makes the algorithm inconsistent, at least, with a single run with one Markov chain. The inconsistency is due to the necessary weighting of the Markov chain in the start-up bias elimination. This misconception has nothing to with the convergence of the distribution, which happens regardless of the start-up bias elimination. The main issue here is the introduced weighting.

Without the start-up bias elimination, the biased and consistent MLT converges to the correct solution using a single chain. When the start-up bias is used, the estimate obtained from one chain is scaled by a random scaling factor, making this algorithm inconsistent for an infinitely long run. To correctly understand this concept, consider a set of initial samples from many different runs. The average over many independent chains of the same length still converges to the correct solution, which means that the algorithm is unbiased. This is hinted by Equation 11.7 in Veach's thesis, where N is the chain length. This is why MLT with the start-up bias elimination is inconsistent.

Acknowledgements The author would like to thank Iliyan Georgiev for his constructive feedback on the manuscript.

References

[HJJ10] HACHISUKA T., JAROSZ W., JENSEN H.: A progressive error estimation framework for photon density estimation. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2010) 29* (December 2010), 144:1–144:12. 4

[HOJ08] HACHISUKA T., OGAKI S., JENSEN H. W.: Progressive photon mapping. In *ACM SIGGRAPH Asia 2008 Papers* (2008), pp. 130:1–130:8. 3

[Kaj86] KAJIYA J. T.: The rendering equation. In *ACM SIGGRAPH 1986 Papers* (1986), vol. 20, pp. 143–150. 1

[MU49] METROPOLIS N., ULAM S.: The monte carlo method. *Journal of the American Statistical Association 44*, 247 (1949), 335–341. 2

[Vea98] VEACH E.: *Robust Monte Carlo methods for light transport simulation*. PhD thesis, Stanford University, Stanford, CA, USA, 1998. AAI9837162. 1, 2

[WRC88] WARD G. J., RUBINSTEIN F. M., CLEAR R. D.: A ray tracing solution for diffuse interreflection. In *ACM SIGGRAPH Computer Graphics* (1988), vol. 22, ACM, pp. 85–92. 1