Monte Carlo Notes

The following are definitions and derivations that are explicitly written down with the hope that they will help guide the ray tracer implementation.

1 Monte Carlo Estimator and Radiance

We intend to probabilistically approximate the integral $\int_{D^*} f$ for some integrable function $f : D \to R$, where $D^* \subseteq D$. The Monte Carlo estimator will suffice: given n iid samples $\boldsymbol{X}_i \in D^*$ s.t. $\boldsymbol{X}_i \sim p \implies p(D \setminus D^*) = 0$ (with the usual restriction that $\int_D p = 1$), we define our estimator

$$\boldsymbol{M}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{f(\boldsymbol{X}_i)}{p(\boldsymbol{X}_i)}$$

$$\implies \begin{cases} \mathrm{E}[\boldsymbol{M}_n] & = \int_D p(\boldsymbol{x}) \cdot \frac{1}{n} \sum_{i=1}^{n} \frac{f(\boldsymbol{x})}{p(\boldsymbol{x})} \, d\boldsymbol{x} \\ & = \frac{1}{n} \sum_{i=1}^{n} \int_{D^*} \frac{f(\boldsymbol{x})}{p(\boldsymbol{x})} p(\boldsymbol{x}) \, d\boldsymbol{x} \\ & = \int_{D^*} f = \mu \\ \mathrm{Var}[\boldsymbol{M}_n] & = \mathrm{Var}[\frac{1}{n} \sum_{i=1}^{n} \frac{f(\boldsymbol{X}_i)}{p(\boldsymbol{X}_i)}] \\ & = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}[f(\boldsymbol{X}_i)/p(\boldsymbol{X}_i)] \\ & = \frac{1}{n} \mathrm{Var}[f(\boldsymbol{X}_i)/p(\boldsymbol{X}_i)] \end{cases}$$

Thus $\lim_{n\to\infty} \mathrm{Var}[\boldsymbol{M}_n] = 0$. The definition of variance suggests that increasing the number of samples reduces squared error: $\mathrm{Var}[\boldsymbol{M}_n] = \mathrm{E}[(\boldsymbol{M}_n - \mu)^2]$, which in turn suggests that the estimator converges to the desired integral (which could perhaps be rationalized as a consequence of the law of large numbers).

In a ray tracing application, we seek to approximate the integral term (reflected radiance L_r) of the rendering equation:

$$L_r(\boldsymbol{x}, \omega_o) = \int_\Omega f(\boldsymbol{x}, \omega_i, \omega_o) L_i(\boldsymbol{x}, \omega_i) \cos\theta \, d\omega_i$$

with BSDF f, incoming radiance L_i, incoming direction ω_i, and outgoing direction ω_o at point $\boldsymbol{x}$.

2 Improving Estimator Efficiency

2.1 Importance Sampling

Suppose we pick p s.t. $p = kf$, where f is the estimated function from before. Then $\int_D p = 1 \implies k = 1/\int_{D^*} f$, in which case the estimator term $\frac{f(\boldsymbol{X}_i)}{p(\boldsymbol{X}_i)} = \int_{D^*} f = \mu$ already. Then $\mathrm{Var}[\boldsymbol{M}_n] = 0$ immediately. While this ideal p defeats the purpose of the Monte Carlo estimator, it intuitively follows that picking p that roughly conforms to the "shape" of f will decrease estimator variance. In practice, this means making p large when the contribution from f is large and vice-versa for when the contribution from f is relatively small.

2.2 Multiple Importance Sampling (MIS)

It may be desireable to utilize multiple densities p_i when estimating the rendering equation. Veach et al (1997) offers the multi-sample Monte Carlo estimator:

$$\boldsymbol{M}_n^* = \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n} w_i(\boldsymbol{X}_{i,\,j}) \frac{f(\boldsymbol{X}_{i,\,j})}{p_i(\boldsymbol{X}_{i,\,j})}$$

given a set of densities $\{p_1, \ldots, p_n\}$ and n_i samples drawn for each p_i and $\boldsymbol{X}_{i,\,j} \sim p_i$.

We expect that the bias, $\beta(\boldsymbol{M}) = \mathrm{E}[\boldsymbol{M}] - \int_{D^*} f$ is still zero so long as we impose the conditions that **(W1)** $\sum_{i=1}^{n} w_i(\boldsymbol{x}) = 1$ when $f(\boldsymbol{x}) \neq 0$ and **(W2)** $w_i(\boldsymbol{x}) = 0$ when $p_i(\boldsymbol{x}) = 0$:

The multi-sample estimator is unbiased: $\beta(\boldsymbol{M}_n^*) = 0$. Each random sample $\boldsymbol{X}_{i,\,j}$ is not necessarily identically-distributed but they are nevertheless independent, so we can manipulate the expectation accordingly, assuming each $n_i \geq 1$:

$$\begin{aligned}
\mathrm{E}[\boldsymbol{M}_n^*] &= \int_D \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n} w_i(\boldsymbol{x}) \frac{f(\boldsymbol{x})}{p_i(\boldsymbol{x})} \cdot p_i(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \int_{D^*} \sum_{i=1}^{n} w_i(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \int_{D^*} f, \text{ by } \textbf{(W1)}
\end{aligned}$$

QED

Veach et al offers the power heuristic as a "good" weighting function: $w_i(\boldsymbol{x}) = \frac{[n_i p_i(\boldsymbol{x})]^\gamma}{\sum_k [n_k p_k(\boldsymbol{x})]^\gamma}$, where $\gamma = 1$ produces the simpler balance heuristic ($\gamma = 2$ is often sufficient).