Notes on Monte Carlo

The following are definitions and derivations that are explicitly written down with an extent of rigor that I personally find useful as it regards the implementation.

1 Monte Carlo Estimator

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).

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). And it is clear that the power heuristic meets both weight function criteria.

2.3 Russian Roulette

Russian roulette offers a way to terminate paths while maintaining an unbiased estimate. After picking an arbitrary termination probability $q \in [0, 1]$ (usually increasing as the integrand becomes smaller), we define a new discrete estimator $\boldsymbol{R} \in \{\frac{1}{1-q}\boldsymbol{M}_n^*, \vec{0}\}$ s.t. $P(\boldsymbol{R} = \frac{1}{1-q}\boldsymbol{M}_n^*) = 1-q$ and $P(\boldsymbol{R} = \vec{0}) = q$. Then

$$\mathrm{E}[\boldsymbol{R}] = (1-q) \cdot \frac{1}{1-q}\mathrm{E}[\boldsymbol{M}_n^*] + q \cdot \vec{0} = \mathrm{E}[\boldsymbol{M}_n^*]$$

3 Light Transport

3.1 Rendering Equation

Radiance is flux per unit projected area per unit solid angle (watts/(steradian$\cdot$m^2)), which is what we seek to measure. The rendering equation describes outgoing radiance from a point $\boldsymbol{x}$ in a direction $\boldsymbol{\Theta}$:

$$L(\boldsymbol{x} \to \boldsymbol{\Theta}) = L_e(\boldsymbol{x} \to \boldsymbol{\Theta}) + \int_{\Omega_{\boldsymbol{x}}} f_r(\boldsymbol{x}, \boldsymbol{\Psi} \to \boldsymbol{\Theta}) L(\boldsymbol{x} \leftarrow \boldsymbol{\Psi}) |\boldsymbol{N}_{\boldsymbol{x}} \cdot \boldsymbol{\Psi}|\, d\boldsymbol{\omega}_{\boldsymbol{\Psi}}$$

Given incoming direction(s) $\boldsymbol{\Psi}$, BRDF f_r, incoming radiance L, emitted radiance L_e, and surface normal $\boldsymbol{N}_{\boldsymbol{x}}$. Alternatively, the area formulation of the rendering equation states that

$$L(\boldsymbol{x} \to \boldsymbol{\Theta}) = L_e(\boldsymbol{x} \to \boldsymbol{\Theta}) + \int_A f_r(\boldsymbol{x}, \boldsymbol{\Psi} \to \boldsymbol{\Theta}) L(\boldsymbol{y} \to -\boldsymbol{\Psi}) V(\boldsymbol{x}, \boldsymbol{y}) \frac{|\boldsymbol{N}_{\boldsymbol{x}} \cdot \boldsymbol{\Psi}||\boldsymbol{N}_{\boldsymbol{y}} \cdot -\boldsymbol{\Psi}|}{r_{\boldsymbol{xy}}^2}\, dA_{\boldsymbol{y}}$$

because incoming radiance is equivalent to ougoing radiance from every other point $\boldsymbol{y}$ in the scene, with the visibility term V to account for obstructions. The area formulation enables importance sampling of light sources in a scene, so it is useful to use it for "direct" illumination and the preceding hemispherical formulation for "indirect" illumination:

$$L_r(\boldsymbol{x} \to \boldsymbol{\Theta}) = \int_A f_r(\boldsymbol{x}, \boldsymbol{\Theta} \to \boldsymbol{\Psi}) L_e(y \to -\boldsymbol{\Psi}) V(\boldsymbol{x}, \boldsymbol{y}) G(\boldsymbol{x}, \boldsymbol{y})\, dA_{\boldsymbol{y}} +$$

$$\int_{\omega_{\boldsymbol{x}}} f_r(\boldsymbol{x}, \boldsymbol{\Theta} \to \boldsymbol{\Psi}) L_i(x \leftarrow \boldsymbol{\Psi}) |\boldsymbol{N}_{\boldsymbol{x}} \cdot \boldsymbol{\Psi}| \, d\omega_{\boldsymbol{\Psi}}$$

where L_i is reflected radiance from the incoming direction $\boldsymbol{\Psi}$. The light transport problem is suitable for Monte Carlo integration.