Probability and Monte Carlo Estimators

6

CS184: Computer Graphics and Imaging

March 5, 2019

1 Quick Terminology

Expectation: probability-weighted average of all possible values. In the discrete case, given by

$$E[X] = \sum_i x_i p_i$$

In the continuous case, given by

$$E[X] = \int xp(x)dx$$

.

Variance: the expected value of the squared deviation from the mean, or how spread apart values are from the mean. Given by

$$\begin{aligned} Var(x) &= E[(X - E[X])^2] &(1)\\ &= E[X^2] - E[X]^2 &(2) \end{aligned}$$

Cumulative Distribution Function (CDF): probability that a sample from distribution X will take a value less than or equal to x.

Lagrange Multipliers: a method to find the maxima or minima of a function $f(x)$ with constraints $g(x) = 0$. We create a function

$$L(x, \lambda) = f(x) + \lambda g(x)$$

and look for critical points where the gradient of L is 0. The critical points of L are the maxima/minima of f.

2 Inversion Method

Recall the inversion method from class. Given a uniform random variable U in the interval $[0, 1]$, we can generate a random variable from any other one dimensional distribution if we have access to the inverse of its cumulative distribution function, $F^{-1}(x)$. We simply have to return $X = F^{-1}(U)$. This is how we choose sample points when running a ray tracing algorithm.

1. What function of U will return a sample from the exponential distribution (with parameter λ)? This distribution has density

$$p_\lambda(x) = \lambda e^{-\lambda x}$$

and is defined for $x \geq 0$.

Solution: First, we need to calculate the CDF:

$$F_\lambda(x) = \int_0^x p_\lambda(x)dx = \int_0^x \lambda e^{-\lambda x}dx = -e^{-\lambda x}\Big|_0^x = 1 - e^{-\lambda x}$$

Its inverse is

$$F_\lambda^{-1}(x) = -\frac{\log(1-x)}{\lambda}$$

Thus we can return

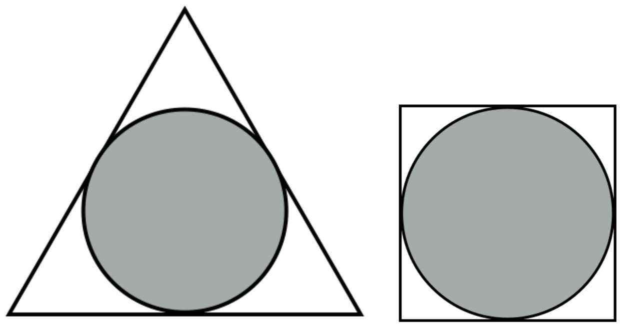
$$-\frac{\log(1-U)}{\lambda}$$

to sample from the exponential distribution.

3 Rejection Sampling

Recall that rejection sampling is one way of using the Monte-Carlo method to sample from a probability distribution. We repeatedly sample values from a proposed distribution, then accept or reject that sample based on whether or not it falls within a probability density function that we know how to sample from. If the sample falls outside of the PDF, then we reject that sample. The remaining samples should be uniformly distributed within our target probability function.

The figure below shows two methods for estimating π using rejection sampling. Method (a) generates random points uniformly in an equilateral triangle. Method (b) generates random points uniformly in a square. Both methods estimate π using the ratio of the number of points that lie within the shape's tangent circle to the total number of points sampled.



(a) Triangle (b) Square

1. In method (a), suppose there are k points out of all n points sampled that lie within the triangle's tangent circle. Then, a formula for the estimated value of π is:

   Solution: We can relate the ratio between k and n to the ratio between the area of the circle and the area of the triangle. Note that the area of an equilateral triangle with an inscribed circle of radius r is $3\sqrt{3}r^2$. Set up the ratio and then solve for π.

$$\frac{k}{n} = \frac{\pi r^2}{3\sqrt{3}r^2}$$
$$\pi = \frac{3\sqrt{3}k}{n}$$

2. In method (b), suppose there are k points out of all n points sampled that lie within the square's tangent circle. Then, a formula for the estimated value of π is:

   Solution: As above, we relate the ratio between k and n to the ratio between the area of the circle and the area of the square. The square with an inscribed circle of radius r has sides of length $2r$.

$$\frac{k}{n} = \frac{\pi r^2}{4r^2}$$
$$\pi = \frac{4k}{n}$$

3. Which method ((a) or (b)) has lower variance with the same number of samples?

   Solution: (b) has lower variance, because a larger percentage of the points will land in the shaded area.

4. What are some downsides to using uniform random sampling?

Solution: It may take a lot of samples to get accurate representation of desired distribution, especially if the sample points lie on one side of the distribution. To fix this problem, we want to sample more intelligently - see importance sampling.

Optional: there's also a notion of discrepancy - from Wikipedia, the discrepancy of a sequence is low if the proportion of points in the sequence falling into an arbitrary set B is close to proportional to the measure of B. With that in mind, in many cases people use low-discrepancy sequences to sample. When these sequences are used to numerically approximate an integral, the method is called quasi-Monte Carlo integration.

5. What are some downsides to using rejection sampling for high-dimensional spaces?

Solution: We run into the problem commonly known as the "curse of dimensionality," which means that many more samples will be rejected than will be accepted. This means that the efficiency of rejection sampling is dramatically reduced in higher-dimensional spaces.

4 Importance Sampling

Suppose we are trying to integrate a function f on some domain (In the case of ray tracing, the sum of the total lighting incident on a surface). The true value of the integral is

$$I = \int f(x)dx.$$

A typical Monte Carlo estimator for this value is

$$\frac{f(X)}{p(X)}$$

where X is a random variable sampled from the probability density p. The expectation of this estimator is

$$E\left[\frac{f(X)}{p(X)}\right] = \int \frac{f(x)}{p(x)}p(x)dx = \int f(x)dx = I.$$

This is the most basic form of *importance sampling*. This estimator has variance equal to

$$E\left[\left(\frac{f(X)}{p(X)}\right)^2\right] - I^2 = \int \frac{f(x)^2}{p(x)}dx - I^2.$$

As the previous question demonstrated, estimators with a lower variance are desirable because we don't have to take as many samples to get an accurate guess, and therefore speed up our renderer! Let's try and find an estimator that minimizes the variance.

1. In order to solve for the minimum variance sampling distribution p we would have to use the calculus of variations. However, we can solve the discrete case involving probability mass functions by using only Lagrange multipliers. In this case, we have a discrete "function" f_i, $i = 1, \ldots, n$, and a probability mass function p_i. What is the discrete distribution p_i that solves

$$\min_{p_i} \sum_{i=1}^{n} \frac{f_i^2}{p_i}, \quad \text{subject to } \sum_{i=1}^{n} p_i = 1, p_i \geq 0?$$

Solution: The Lagrangian is

$$L = \sum_i \frac{f_i^2}{p_i} + \lambda \left(\sum_i p_i - 1 \right)$$

which has partial derivatives

$$\frac{\partial L}{\partial p_i} = -\frac{f_i^2}{p_i^2} + \lambda.$$

Setting these equal to zero yields that we have $p_i = |f_i|/\sqrt{\lambda}$ for all i. In order to satisfy the constraint, we must thus have

$$1 = \frac{1}{\sqrt{\lambda}} \sum_i |f_i| \quad \Rightarrow \quad \sqrt{\lambda} = \sum_i |f_i| \quad \Rightarrow \quad p_i = \frac{|f_i|}{\sum_j |f_j|}$$

Thus in the case where the f_i are nonnegative, we see that the distribution p is exactly proportional to f. However, we are trying to estimate $\sum_j f_j$ in the discrete case, so we cannot know this p without already knowing the quantity we are trying to compute! The same reasoning carries over to the continuous case, where the minimum variance distribution is

$$p(x) = \frac{f(x)}{\int f(x')dx'} = \frac{f(x)}{I}$$

5 Multiple Importance Sampling

In question 4, we showed that the optimal estimator is the distribution $f(x)$ itself! However, in many cases, like rendering, we don't know what that distribution is. One way to remedy that problem is using multiple different sampling distributions that each guess at the true distribution.

With that in mind, suppose we have two different sampling methods p_1 and p_2 for I. Samples from either of them can provide an estimate of I, equal to

$$\frac{f(X_i)}{p_i(X_i)}$$

when $X_i \sim p_i$. However, we might not be sure which of p_1 or p_2 provides a lower variance estimate for f, and thus may be interested in creating a weighted combination of these estimators.

1. What is an unbiased estimator?

> **Solution:** The expectation of an unbiased estimator is equal to the quantity being estimated.

2. What constraint on the two constants w_1 and w_2 will make the following estimator unbiased?

$$w_1\frac{f(X_1)}{p_1(X_1)} + w_2\frac{f(X_2)}{p_2(X_2)}$$

> **Solution:**
> $$E\left[w_1\frac{f(X_1)}{p_1(X_1)} + w_2\frac{f(X_2)}{p_2(X_2)}\right] = w_1 I + w_2 I = (w_1 + w_2)I$$
> For this estimator to be unbiased, we need that $w_1 + w_2 = 1$.

3. What is a sufficient constraint on the two functions $w_1(x)$ and $w_2(x)$ for the following estimator to be unbiased?

$$w_1(X_1)\frac{f(X_1)}{p_1(X_1)} + w_2(X_2)\frac{f(X_2)}{p_2(X_2)}$$

> **Solution:**
> $$\begin{aligned} E\left[w_1(X_1)\frac{f(X_1)}{p_1(X_1)} + w_2(X_2)\frac{f(X_2)}{p_2(X_2)}\right] &= \int w_1(x)\frac{f(x)}{p_1(x)}p_1(x)dx + \int w_2(x)\frac{f(x)}{p_2(x)}p_2(x)dx \\ &= \int \left(w_1(x)f(x) + w_2(x)f(x)\right)dx \\ &= \int f(x)(w_1(x) + w_2(x))dx. \end{aligned}$$

This will be equal to I if we apply the constraint that $w_1(x) + w_2(x) = 1$ for all x.