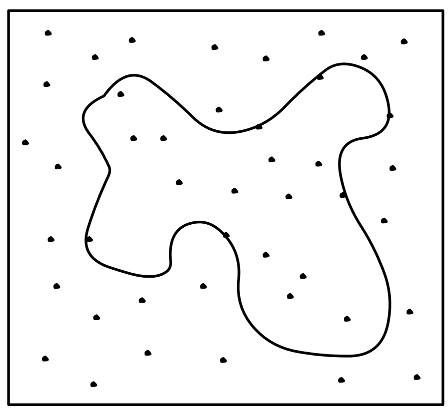
CS667 **Lecture 6: Monte Carlo Integration** 02/10/05

Venkat Krishnaraj Lecturer: Steve Marschner

1 Idea

The main idea of Monte Carlo Integration is that we can estimate the value of an integral by looking at a large number of random points in the domain of that integral and adding up the values. In general, we find that as we increase the number of random points, the addition of those values converge towards the integral.

To illustrate the above idea, let us use a simple example.



As shown above, suppose we have a region in the square and we only know how to find if a particular point is in it or not. If we drop one million random points on the square, and count how many fall inside that, we will get an approximate measure of the area (area fraction in reality).

Similar example
Telephone poll asking a yes/no question to estimate number of people who own cars. For this poll to work, we need to know size of population correctly, and more importantly, ensure that there is a uniform chance at selecting any given person (so that our final results reflect the overall population).

The above examples are just counting or computing area (in other words, integrating a function that is one or zero). But we can extend this principle to integrate complex integrals whose values cannot be analytically easily determined.

For example, we try and compute the average height of a given terrain. To perform this experiment, we drop 1 million sticky leaflets from a plane and we measure the altitude where each of those sticky leaflets land (or of each of those leaflets itself). We can see that the average height of the terrain will be average height over all those leaflets (however, we need to ensure that the leaflets are thrown in such a way that they are uniformly distributed and that they are sticky so that they dont collect at low lying regions or valleys to give incorrect results).

Another example, a Telephone survey to compute average income in the United States. To perform this experiment, we could go about asking every person in US their income and averaging their responses (which is quite a tedious task, and probably not possible). However, instead, we could pick a subset, like asking a thousand people and averaging their responses to estimate the average income. However, we need to be careful of two things here. Again, as we have seen in the previous case, we have to ensure unform distribution (ie the subset is a good indicator of the US population in general) Also, we need to take into account that with a phone survey, you will only get average over those people who own phones (weighted by how many phone numbers they have.)

Another issue, if you survey 100 people and happen to include one CEO in that survey, then the results could be quite misleading. Say we get 99 people with average income at 50k, and one CEO with a salary of 5 million (100 times the average), the mean will be about 100k, which is inaccurate. With a more typical person instead of the CEO, we would have got a mean of 50K. This is sometimes the reason as to why we may compute medians instead of means when possible.

In relation to graphics, this means that we need to be careful while integrating a distribution with outliers (for example, a scene with many normal light sources and a few very bright light source.) If we integrate in the presence of outliers, we will tend to get incorrect results.

2 Quick review of probability

2.1 Discrete Random Variable

Definition:
A variable that takes on one of a finite number of different values for every trial of an experiment is called a Discrete Random Variable

Eg
Throw a die, define random variable X as a number that comes up.
X has equal probability of taking on the values $1, 2...6$ (the 6 outcomes happen equally frequently).
The probability space R is a set of values X can take on, and the distribution of X assigns a frequency or probability to each. I'll call these elements $\omega_1 ... \omega_6$

Example for the die
The die can take any discrete value between 1 to 6. We shall call this set R. Also, each of those face values are equally likely.
Thus, mathematically,
$R = \{1, 2, ...6\}$
and $p(i) = \frac{1}{6}$ for each i which is an element of $\{1, 2...6\}$ Here $\omega_i = i$

We write $X \sim p(i)$, where p is the probability distribution for X

2.1.1 A few mathematical definitions

Expected Value or Mean

$E\{x\} = \sum_{i=1}^{n} p(\omega_i)\omega_i$

This is a number you will get if you take many samples of x and average their values.

Eg for the die $E\{x\} = \frac{1}{6} \cdot 1 + ... \frac{1}{6} \cdot 6 = \frac{21}{6} = 3.5$

If we take a bunch of samples and average them, then we get the sample mean which is an estimate of the expected value

If $Xi \sim p$, for $i = 1$ to N (they are samples of X),

then $\frac{1}{n}\sum_{i=1}^{n} x_i$ is an estimate of $E\{X\}$

Linearity of expectation $E\{X + Y\} = E\{X\} + E\{Y\}$

Variance

$\sigma^2\{X\} = E\{X - E\{X\}\}^2 = \sum_{i=1}^{n} p(\omega_i)(\omega_i - E\{X\})^2$

The variance tells us something about how spread out the distribution is

A handy formula : $\sigma^2\{x\} = E\{X^2\} - E\{X\}^2$

Proof :

$$\begin{aligned} \sigma^2\{X\} &= E\{X - E\{X\}\}^2 \\ &= E\{X^2 - 2XE\{X\} + E\{X\}^2\} \\ &= E\{X^2\} - 2E\{XE\{X\}\} + E\{E\{X\}^2\} \\ &= E\{X^2\} - E\{X\}^2 \end{aligned}$$

since $E\{XE\{X\}\} = E\{X\}E\{X\}$ and $E\{E\{X\}^2\} = E\{X\}^2$

Eg Variance of die

$E\{x^2\} = \frac{1}{6} \cdot 1^2 + \frac{1}{6} \cdot 2^2 + \frac{1}{6} \cdot 3^2 + ... + \frac{1}{6} \cdot 6^2 = \frac{91}{6}$

Then, $\sigma^2\{X\} = E\{X^2\} - E\{X\}^2 = \frac{91}{6} - (\frac{21}{6})^2 = 2\frac{11}{12}$

Alternative way of getting the result :

$E\{(X - E\{X\})^2\} = \frac{1}{6}(1 - 3\frac{1}{2})^2 + \frac{1}{6}(2 - 3\frac{1}{2})^2 + ... + \frac{1}{6}(6 - 3\frac{1}{2})^2 = 2\frac{11}{12}$

Definition $\sigma\{X\} = \sqrt{\sigma^2\{X\}}$ Relation between standard deviation $\sigma\{X\}$ and variance

Another simple example

$\Omega = \{0, 1\} p(1) = p, p(0) = 1 - p$

$E\{X\} = 1 * p + 0 * (1 - p) = p$

$E\{X^2\} = 1.p + 0.(1 - p) = p$

$E\{X\}^2 = p^2$

$\sigma^2\{X\} = p - p^2 = p(1 - p)$

Note that for $p = 0$ or $p = 1$, variance is zero

2.2 Continuous Case

In this case, Ω is an infinite set (eg the real interval [0,1]) and then the random variable X is called a continuous random variable.

In the continuous case, the probability is a measure: it assigns a finite probability for any finite(suitable) subset of R.

But for well behaved distributions, we can talk about a probability density function, or pdf, $p(x)$, where $p(x)$ is the probability of landing near x

$p(x)dx = P_r\{a < x < a + dx\}$ (or from $a - \frac{dx}{2}$ to $a + \frac{dx}{2}$)

$P_r\{X \in S\} = \int_S dp$ where $\int_\omega dp = 1$

Here, expected value is an integral instead of a sum:

$E\{X\} = \int_\Omega xp(x)dx$ or $\int_\Omega xdp(x)$

The relations for the discrete case hold true, viz $E\{X + Y\} = E\{X\} + E\{Y\}$

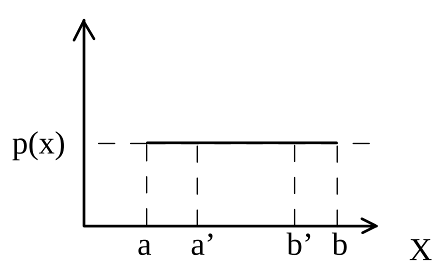
$\sigma^2\{X\} = \int_\Omega (x - \overline{x})dp(x)$

Same manipulation as before shows

$\sigma^2\{X\} = E\{X^2\} - E\{X\}^2$

Simple example for continuous probability :

Uniform distribution from a to b $p(x) = \frac{1}{(b-a)}$



$P_r(x \in [a', b']) = \int_{[a',b']} dp = \int_{a'}^{b'} \frac{1}{b-a} dx = \frac{(b'-a')}{(b-a)}$

$$\begin{aligned} E\{X\} &= \int_a^b x dp \\ &= \int_a^b x p(x) dx \\ &= \int_a^b \frac{x}{b-a} dx \\ &= \frac{1}{2} \frac{b^2 - a^2}{b-a} \\ &= \frac{b+a}{2} \end{aligned}$$

Now, we will compute the variance

$$\begin{aligned} \sigma^2\{X\} &= E\{X^2\} - E\{X\}^2 \\ &= \int_a^b x^2 dp - (\frac{b+a}{2})^2 \\ &= \int_a^b x^2 p(x) dx - (\frac{b+a}{2})^2 \\ &= \int_a^b \frac{x^2}{b-a} dx - (\frac{b+a}{2})^2 \\ &= \frac{1}{3} \frac{b^3 - a^3}{b-a} - (\frac{b+a}{2})^2 \\ &= \frac{b^2 + ab + a^2}{3} - \frac{b^2 + a^2 + 2ab}{4} \\ &= \frac{4b^2 + 4ab + 4a^2 - 3b^2 - 3a^2 - 6ab}{12} \\ &= \frac{b^2 - 2ab + a^2}{12} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

3 Monte Carlo Integration

Based on the fact that the expected value of an random variable is an integral, we can compute this integral experimentally.

The basic form, we want to compute

$I = \int_\Omega f(x) dx$

We want to do it using random samples generated according to a distribution p. This means we will take samples $x_i \sim p$, evaluate some function $g(x_i)$ and average :

$\overline{I} = \frac{1}{N}\sum_i g(x_i)$

We saw earlier that $E\{\overline{I}\} = E\{g(x)\} = \int_\Omega g(x)p(x)dx$

So, what do you want for g?

Clearly $g(x) = \frac{f(x)}{p(x)}$

$E\{\frac{f(x)}{p(x)}\} = \int_\Omega \frac{f(x)}{p(x)}.p(x)dx = \int_\Omega f(x)dx$

This is the core idea

While coding, we need to thus take care of 2 important aspects :

- What kind of probability distribution we are using
- What is the function that we are integrating

As we can see, our estimated value may not be the exact value of the integral, however, before trying to express variance quantitatively, there are two methods to lower it

- Importance Sampling
- Stratification

In the next class, we shall see the above two techniques and also derive a mathematical expression for the variance of $\overline{I}$, and how it is related to the variance of g.