Monte Carlo Simulation to Estimate Value of Pi

The Problem

Rabbit and Tortoise are at it again. They've been tasked with approximating the value of π using Monte Carlo Simulations. Rabbit and Tortoise have come up with different approaches as stated below. Your task is to follow both of these approaches and report your findings.

Monte Carlo Simulations

In statistics, we often run into integrals of the form

$$I = \int_a^b h(x)g(x)dx$$

These are hard to compute using traditional methods. In such cases, we can use Monte Carlo Simulations to obtain an estimate for the value of the integral by transforming the integration problem into a procedure of sampling values from a probability distribution and calculating a statistic over those samples.

In this case, if the function $g(x)$ is positive on the interval (a, b) and the integral of the function is finite with $C = \int_b^a g(x)$, we can define a new probability distribution on the interval (a, b):

$$p(x) = \frac{g(x)}{C}$$

We can now rewrite the original integration as

$$I = C\int_a^b h(x)p(x)dx = CE_{p(x)}[h(x)]$$

The integral is same as a constant times the expected value of the integrand function. The expected value is approximated by sampling N points and computing the sample mean

$$E_{p(x)}[h(x)] \approx \frac{1}{N} \sum_{i}^{N} h(x_i)$$

Part A: Circle Approximation

As we know, area of a circle can be written as

$$I = \int_{-r}^{r} \int_{-r}^{r} \mathbb{1}(x^2 + y^2 \leq r^2) dx dy$$

where $\mathbb{1}$ represents the Indicator Random Variable, which takes the value 1 when the condition evaluates to true and 0 when it evaluates to false.

If we obtain this integral, we can calculate π by using $\pi = \frac{I}{r^2}$. To evaluate the integral, we will apply the method mentioned above.

Since, x and y follow uniform distributions in the range $[-r, r]$, we have $p(x) = 1/2r$ and $p(y) = 1/2r$. Hence the integral becomes

$$I = (2r)(2r) \int \int \mathbb{1}(x^2 + y^2 \leq r^2) p(x) p(y) dx dy$$

$$= 4r^2 \int \int \mathbb{1}(x^2 + y^2 \leq r^2) p(x) p(y) dx dy$$

which can be written as

$$I \approx 4r^2 \frac{1}{S} \sum_{s=1}^{S} \mathbb{1}(x_s^2 + y_s^2 \leq r^2)$$

where S is the number of points sampled.

Thus, sampling points (x, y) with each following the distribution mentioned above and estimating the value of the indicator random variable for each point, we can get an estimate for the integral. Hence pi can be estimated.

Run the program multiple times. Starting off with 10 samples. For each successive run, increase the number of samples to 10 times the previous. Go on till 10^7 points.

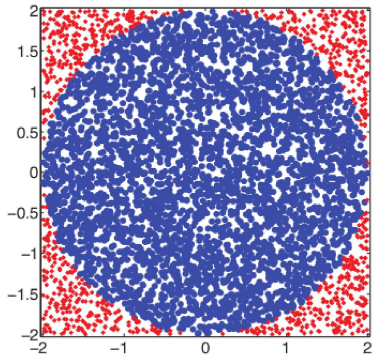

Fig 2.19 from Kevin Murphy - Machine Learning, A Probabilistic Perspective

For each run, make plots like the one given above. The colour of the points represents the value of the indicator random variable.

Part B: Sphere Approximation

This is Tortoise's view of the problem. Using your understanding of the previous section, extrapolate for the case of a sphere (3 Dimensions).