#5

Monte Carlo

L. Olson

September 10, 2015

Department of Computer Science
University of Illinois at Urbana-Champaign

why

- Central to the PageRank (and many many other applications in fincance, science, informatics, etc) is that we randomly process something
- what we want to know is “on average” what is likely to happen
- what would happen if we have an infinite number of samples?
- let’s take a look at integral (a discrete limit in a sense)

integration

- integral of a function over a domain

$$\int_{x \in D} f(x)\, dA_x$$

- the size of a domain

$$A_D = \int_{x \in D} dA_x$$

- average of a function over some domain

$$\frac{\int_{x \in D} f(x) dA_x}{A_D}$$

integral example

The average "daily" snowfall in Champaign last year

- domain: year (1d time interval)
- integration variable: day
- function: snowfall depending on day

$$average = \frac{\int_{day \in year} s(day) d_{day}}{lengthofyear}$$

integral example

The average snowfall in Illinois

- domain: Illinois (2d surface)
- integration variable: (x, y) location
- function: snowfall depending on location

$$average = \frac{\int_{location \in Illinois} s(location) d_{location}}{areaofillinois}$$

integral example

The average snowfall in Illinois today

- domain: Illinois $\times$ year (3d space-time)
- integration variable: location and day
- function: snowfall depending on location and day

$$average = \frac{\int_{day \in year} \int_{location \in Illinois} s(location, day) d_{location,day}}{areaofillinois \cdot lengthofyear}$$

discrete random variables

- random variable x
- values: $x_0, x_1, \ldots, x_n$
- probabilities $p_0, p_1, \ldots, p_n$ with $\sum_{i=0}^{n} p_i = 1$

throwing a die (1-based index)

- values: $x_1 = 1, x_2 = 2, \ldots, x_6 = 6$
- probabilities $p_i = 1/6$

expected value and variance

- expected value: average value of the variable

$$E[x] = \sum_{j=1}^{n} x_j p_j$$

- variance: variation from the average

$$\sigma^2[x] = E[(x - E[x])^2] = E[x^2] - E[x]^2$$

throwing a die

- expected value: $E[x] = (1 + 2 + \cdots + 6)/6 = 3.5$
- variance: $\sigma^2[x] = 2.916$

estimated $e[x]$

- to estimate the expected value, choose a set of random values based on the probability and average the results

$$E[x] = \frac{1}{N}\sum_{j=1}^{N} x_i$$

- bigger N gives better estimates

throwing a die

- 3 rolls: $3, 1, 6 \rightarrow E[x] \approx (3 + 1 + 6)/3 = 3.33$
- 9 rolls: $3, 1, 6, 2, 5, 3, 4, 6, 2 \rightarrow E[x] \approx (3 + 1 + 6 + 2 + 5 + 3 + 4 + 6 + 2)/9 = 3.51$

law of large numbers

- by taking N to ∞, the error between the estimate an the expected value is statistically zero. That is, the estimate will converge to the correct value

$$P\left(E[x] = lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} x_i\right) = 1$$

continuous random variable

- random variable: x
- values: $x \in [a, b]$
- probability: density function $\rho(x)$ with $\int_a^b \rho(x)\, dx = 1$
- probability that the variable is value x: $\rho(x)$

uniformly distributed random variable

- $\rho(x)$ is constant
- $\int_a^b \rho(x)\,dx = 1$ means $\rho(x) = 1/(b-a)$

continuous extensions

- expected value

$$E[x] = \int_a^b x\rho(x)\,dx$$

$$E[g(x)] = \int_a^b g(x)\rho(x)\,dx$$

- variance

$$\sigma^2[x] = \int_a^b (x - E[x])^2\rho(x)\,dx$$

$$\sigma^2[g(x)] = \int_a^b (g(x) - E[g(x)])^2 p(x)\,dx$$

- estimating the expected value

$$E[g(x)] \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i)$$

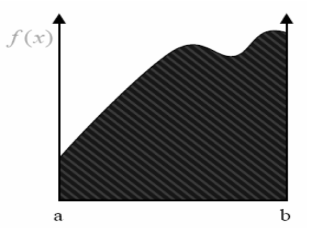
multidementional extensions

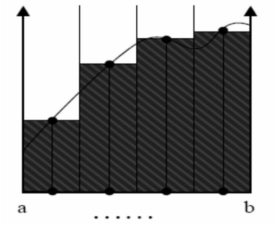
- difficult domains (complex geometries)
- expected value

$$E[g(x)] = \int_{x \in D} g(x)\rho(x)\, dA_x$$

(deterministic) numerical integration

- split domain into set of fixed segments
- sum function values with size of segments (Riemann!)





algorithm

We have for a random sequence $x_1, \ldots, x_n$

$$\int_0^1 f(x)\,dx \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

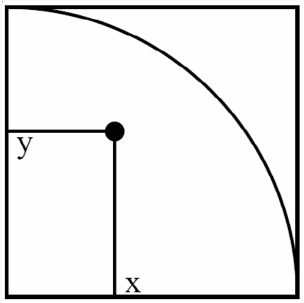
```
n=100
x=np.random.rand(n)
a=f(x)
s=a.sum()/n
```

2d: example computing π

Use the unit square $[0,1]^2$ with a quarter-circle

$$f(x,y) = \begin{cases} 1 & (x,y) \in \textit{circle} \\ 0 & \textit{else} \end{cases}$$

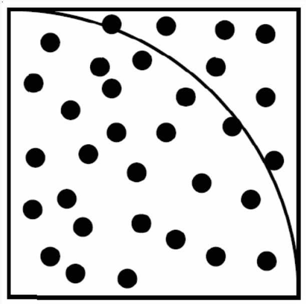
$$A_{quarter-circle} = \int_0^1 \int_0^1 f(x,y)\, dxdy = \frac{pi}{4}$$



2d: example computing π

Estimate the area of the circle by randomly evaluating $f(x, y)$

$$A_{quarter-circle} \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i, y_i)$$



2d: example computing π

By definition

$$A_{quarter-circle} = \pi/4$$

so

$$\pi \approx \frac{4}{N} \sum_{i=1}^{N} f(x_i, y_i)$$

2d: example computing π, algorithm

```
input N
call rand in 2d
for i=1:N
    sum = sum + f(x_i, y_i)
end
sum = 4 * sum/N
```

2d: example computing π, algorithm

The expected value of the *error* is $\mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$

- convergence does not depend on dimension
- deterministic integration is very hard in higher dimensions
- deterministic integration is very hard for complicated domains

clt: central limit theorem

The distribution of an average is close to being normal, even when the distribution from which the average is computed is not normal.

What?

- Let $x_1, \ldots, x_n$ be some independent random variables from any PDF
- Consider the sum $S_n = x_1 + \cdots + x_n$
- The expected value is $n\mu$ and the standard deviation is $\sigma\sqrt{n}$
- That is, $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ approaches the normal distribution
- What? The sample mean has an error of $\sigma/\sqrt{n}$

now what?

- How does one minimize the noise in this random process?
- pick better samples!
- Use more samples where the impact is high: where f is large

$$I \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f(x)}{p(x)}$$

- So pick a distribution similar to f

$$p_{optimal} \propto f(x)$$