ISyE 6416: Computational Statistics
Spring 2017

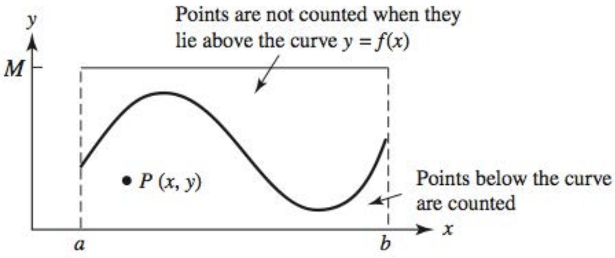
Lecture 13: Monte Carlo Methods

Prof. Yao Xie

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Determine area under a curve

- determine area under a curve $y = h(x)$, continuous function with $0 \leq h(x) \leq M$ over the closed interval $a \leq x \leq b$
- select points at random from within the rectangular region
- area under the curve / area of rectangle $\approx$ number of points counted below curve / total number of random points



Monte Carlo integration

- To compute $\mathbb{E}[h(X)] = \int h(x)f(x)dx$
  $h(x)$: integrand, $f(x)$: probability density function
- Monte Carlo approach to approximate the integration:
  Take i.i.d. samples $X_1, \ldots, X_n$ from pdf $f(x)$.
  Then take sample average

$$\hat{I}_1 = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$$

- Convergence: law of large numbers, as $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \to \mathbb{E}[h(X)]$$

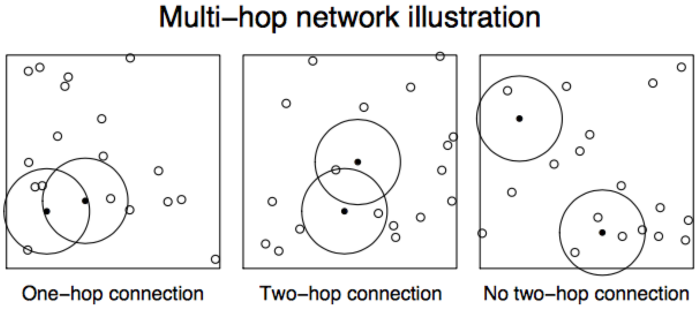
Monte Carlo methods

- Monte Carlo methods (or Monte Carlo experiments) are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results.
- most useful when it is difficult or impossible to use other approaches.
- Monte Carlo methods are mainly used in three distinct problem classes: optimization, numerical integration, and generating draws from a probability distribution.

Estimating probabilities

- Evaluating probability can be rewritten using indicator function as
$$\mathbb{P}(X \in \mathcal{A}) = \mathbb{I}_{\mathcal{A}}(x)$$
- Widely used in engineering: example, wireless multi-hop networks
- There are $m \geq 3$ nodes randomly distributed in the unit square
- Each node can communicate directly with any other node that is within a distance r of it
- A two-hop connection arises between nodes 1 and 2 when nodes 1 and 2 are farther than r but are both within distance r of node j for one or more $j \in \{3, \ldots, m\}$

Source: A. Owen, Monte Carlo theory, methods and examples.



$m = 40$

- Question: What is the probability of forming a two-way connection?
- This probability is an integral over $2m = 40$ dimensional space
- Run 10,000 independent replications of this problem
- Among them, 650 cases have two-hop connection
- The probability is estimated to be

$$\hat{p} = 650/10000 = 0.065.$$

- How good is the estimation?
  CLT-based 99% confidence interval for p

$$\hat{p} \pm \underbrace{z_{0.005}}_{2.58} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

  leading to

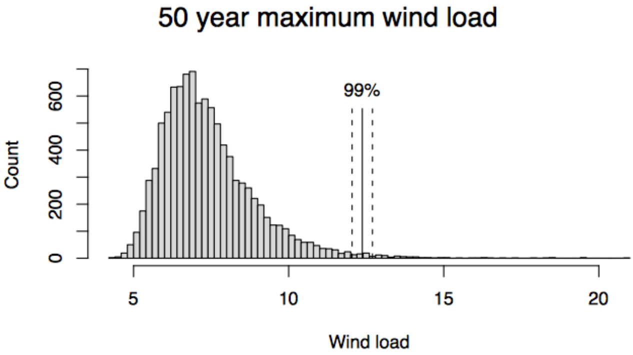
$$0.065 \pm 2.58\sqrt{\frac{0.065 \times 0.945}{10000}} = 0.065 \pm 0.0064$$

	Lower	Upper
CLT	0.05865	0.07135
Exact	0.05881	0.07161

Estimating tail probability

- How strong a wind should we construct a building to be resistant to?

$$\mathbb{P}(W \leq w) = \exp(-\exp((w-52)/4))$$



Source: A. Owen, Monte Carlo theory, methods and examples.

Estimating tail probability

- Suppose the random variable X is binomially distributed with m trials and success probability p. We want to evaluate the right-tail probability $\alpha = \mathbb{P}\{X \geq z\}$
- evaluating tail probability is important in hypothesis testing: finding the significance level of the test, we need to evaluate the tail probability of the test statistics under the null distribution
- tail probability also used for risk management in insurance, and portfolio investment
- For z much larger than mp, α is very small, and estimating this small probability accurately is not easy

Convergence property

- If X is square integrable, we have

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \Rightarrow \mathcal{N}\left(\mathbb{E}[h(X)], \frac{1}{n}\mathsf{Var}[h(X)]\right)$$

- Estimate the order of the rate of convergence by

$$\sqrt{v/n}, \quad v = \frac{1}{n-1}\sum_{i=1}^{n}[h(X_i) - \frac{1}{n}\sum_{j=1}^{n} h(X_j)]^2$$

- disadvantage: slow convergence rate $n^{-1/2}$
- can we reduce v by generating samples smartly?

Important Sampling

- variance reduction method in Monte Carlo integration
- main idea: using samples generated from a different distribution rather than the distribution given
- suppose $g(x)$ is another pdf

$$\int h(x) f(x) dx = \int h(x) \underbrace{\frac{f(x)}{g(x)}}_{\text{likelihood ratio}} g(x) dx$$

change the function to be integrated from

$$h(x) \to h(x) \frac{f(x)}{g(x)}$$

change the pdf from $f(x)$ to $g(x)$

- based on this, now we can sample $Y_1, Y_2, \ldots, Y_n$ i.i.d. from $g(x)$
- estimate the integration as

$$\hat{I}_2 = \frac{1}{n} \sum_{i=1}^{n} h(Y_i) \underbrace{\frac{f(Y_i)}{g(Y_i)}}_{\text{weighted by likelihood ratio}}$$

- we can choose $g(x)$ smartly so this estimator $\hat{I}_2$ has smaller variance than the one based on direct sampling $\hat{I}_1$
- choose a distribution to encourage "important" values

How to choose a pdf to sample from

- we want to show that var$\{\hat{I}_2\} \leq$ var$\{\hat{I}_1\}$, since both methods are unbiased: $\mathbb{E}\{\hat{I}_1\} = \mathbb{E}\{\hat{I}_2\}$, or equivalently $\mathbb{E}\{\hat{I}_2^2\} \leq \mathbb{E}\{\hat{I}_1^2\}$:

$$\int [h(x)\frac{f(x)}{g(x)}]^2 g(x)dx \leq \int h(x)^2 f(x)dx, (*)$$

- choose $g(x) = \frac{|h(x)|f(x)}{\int |h(v)|f(v)dv}$, we can show (*) is true
- main idea: choose $g(x)$ to resemble $|h(x)|f(x)$, the standard error tends to be reduced
- difficulty: if $|h(x)|f(x)$ is unknown, we have to approximate it

Proof.

$$\int [\frac{h(x)f(x)}{g(x)}]^2 g(x)dx$$

$$= \int [h(x)f(x)]^2/g(x)dx$$

$$= \int [h(x)f(x)]^2/(|h(x)|f(x))dx \cdot [\int |h(v)|f(v)dv]$$

$$= (\int |h(x)|f(x)dx)^2 \text{ Cauchy Schwartz}$$

$$\leq \int |h(x)|^2 f(x)dx \underbrace{\int f(x)dx}_{=1}$$

$$= \int h^2(x)f(x)dx$$

□