Seoul
National
University

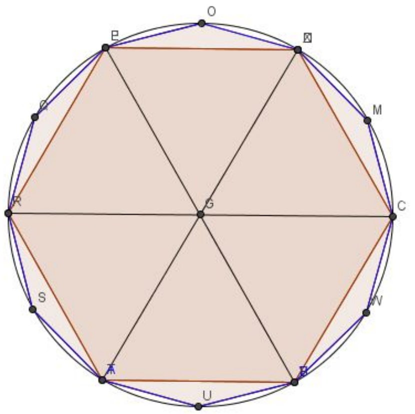
Monte Carlo Method

Name: Chong-kwon Kim

SCONE
Lab.

Computation of the Constant π

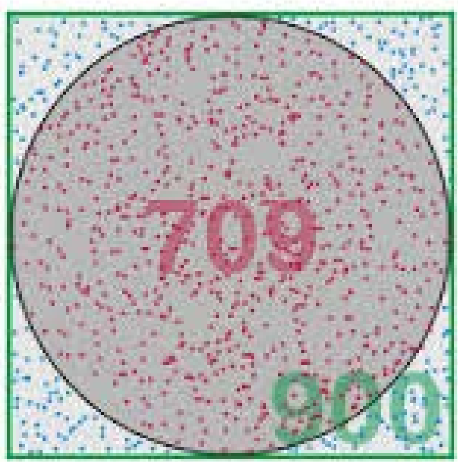
- One of the most famous & oldest problems in mathematics
  - The Bible says that π=3



$\pi = \frac{C}{D}$

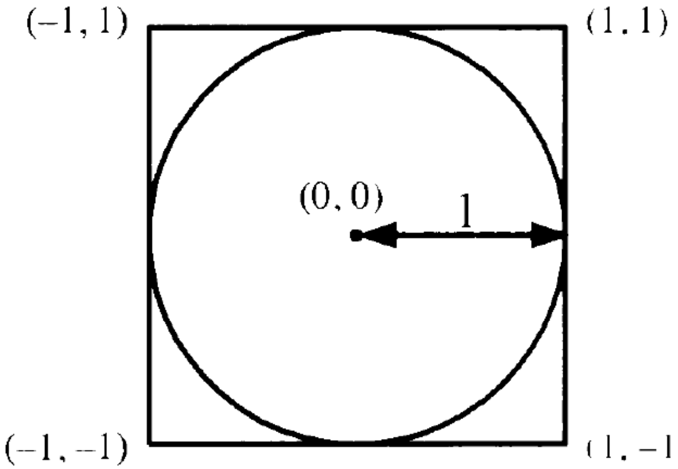
- The old wisdoms found out that
  - π can be bounded between inscribed and circumscribed polygons

- Monte Carlo method (simulation) is another technique to estimate π
  - Count the numbers of randomly selected points inside and outside of the circle, respectively



π = 709/900 * 4 = 3.1511..

Monte Carlo Method



- Estimate the constant π
  - Pick randomly a point (x, y), x, y∈(0, 1) and check if the point is in the circle
  - Let Z = 1, if $\sqrt{x^2 + y^2 \leq 1}$

    0, ow

    Pr(Z=1) = $\pi/4$
  - Repeat the experiment (Simulation) many times (m) and let Z_i be the result of i-th run
  - Let W = $\sum_{i=1}^{m} Z_i$ ➔ E[W]=$\mathrm{m} \cdot (\frac{\pi}{4})$
  - Let $W' = \left(\frac{4}{m}\right) W$, then by Chernoff inequality

$$\Pr(|W' - \pi| \geq \varepsilon\pi) = \Pr\left(\left|W - \frac{m\pi}{4}\right| \geq \frac{\varepsilon m\pi}{4}\right)$$
$$= \Pr(|W - E[W]| \geq \varepsilon E[W])$$
$$\leq 2e^{-m\pi\varepsilon^2/12}$$

W~B(m, $\frac{\pi}{4}$)

For 0 < δ ≤ 1,

Pr(X ≥ $(1 + \delta)\mu$) ≤ $e^{-\frac{\mu\delta^2}{3}}$

(ε, δ)–Approximation

- Definition
  - A simulation is (ε, δ)–approximation for V if the output X of the simulation satisfies

    $\Pr(|X - V| \leq \epsilon V) \geq 1 - \delta$

- To make the constant π estimation be (ε, δ)–approximation
  - From $\Pr(|W - E[W]| \geq \varepsilon E[W]) \leq 2e^{-\frac{m\pi\varepsilon^2}{12}}$

    $\Pr(|W - E[W]| < \varepsilon E[W]) \geq 1 - 2e^{-m\pi\varepsilon^2/12}$

  - From $\delta \geq 2e^{-m\pi\varepsilon^2/12}$, $m \geq \frac{12\ln(\frac{2}{\delta})}{\pi\varepsilon^2}$

Repeat the same experiment many times

(ε, δ)−Approximation

- More generally, Claim
  - Let X_i, i=1,2,⋯,m be i.i.d. indicator random variables with E[X_i]=μ

    If $m \geq 3\ln\left(\frac{2}{\delta}\right)/\varepsilon^2\mu$

    ➔ Then the experiment $\{X_i\}$ is an (ε, δ)−approximation for μ

    ➔ $\Pr(|\frac{1}{m}\sum_{i=1}^{m} X_i - \mu| \geq \varepsilon\mu) \leq \delta$
- Proof is basically the same as the constant π estimation

Exercise 10.1

- Definition: FPRAS(Fully Polynomial Randomized Approximation Scheme)
  - Given an input x and parameters ε, δ with $\varepsilon > 0, \delta < 1$, an FPRAS algorithm outputs an (ε, δ)−Approximation to V(x) in time that is polynomial in 1/ε, ln(1/δ) and the size of input x

Application: DNF

- Consider the complement of CNF
- By the de Morgan's rule

  $\overline{CNF}$ ➔ DNF (Disjunctive Normal Form)

  $(\overline{x_1} + x_2 + \overline{x_3}) \cdot (\overline{x_2} + \overline{x_4}) \cdot (x_1 + \overline{x_3} + \overline{x_4})$

  ➔ $(x_1 \cdot \overline{x_2} \cdot x_3) + (x_2 \cdot x_4) + (\overline{x_1} \cdot x_3 \cdot x_4)$

CNF: Satisfiability
Is there a solution?
Most of random assignments make the formula FALSE

DNF: No solution
Existence of a FALSE assignment
Most of random assignments make the formula TRUE

Count # satisfying random assignments & check if # $\equiv 2^n$

K-SAT: Cascaded modification
➔ MC

Random assignments
Monte Carlo

Simple Monte Carlo for DNF

- Let c(F) be # satisfying assignments of a DNF formula F
- A naïve approach to estimate c(F)

DNF Counting Algorithm 1
1. $\mathrm{X} \leftarrow 0$
2. For $k = 1, \ldots, m$ do
   a) Generate random assignment of n variables
   b) If the random assignment satisfies F, $\mathrm{X} \leftarrow X + 1$
3. Return $\mathrm{Y} \leftarrow (X/m)2^n$

- X_k: Indicator random variable

  $X_k = 1$, if k−th random assignment is a satisfying one
  0, ow
- $\Pr(X_k = 1) = \frac{c(F)}{2^n}$
- $\mathrm{E}[\mathrm{X}] = \mathrm{E}[\sum_{k=1}^{m} X_k] = \mathrm{m} \cdot \frac{c(F)}{2^n}$
- $\mathrm{E}[\mathrm{Y}] = \mathrm{c}(\mathrm{F})$

Simple Monte Carlo for DNF

- How many iterations (m) are required to make X/m be an (ε, δ)−approximation for $c(F)/2^n$?
  - From $m \geq 3\ln\left(\frac{2}{\delta}\right)/\varepsilon^2\mu$ ➔ $m \geq 3 \cdot 2^n \ln\left(\frac{2}{\delta}\right)/\varepsilon^2 c(F)$

- What is the condition that make the algorithm FPRAS?
  - c(F) = $2^n/\alpha(n)$

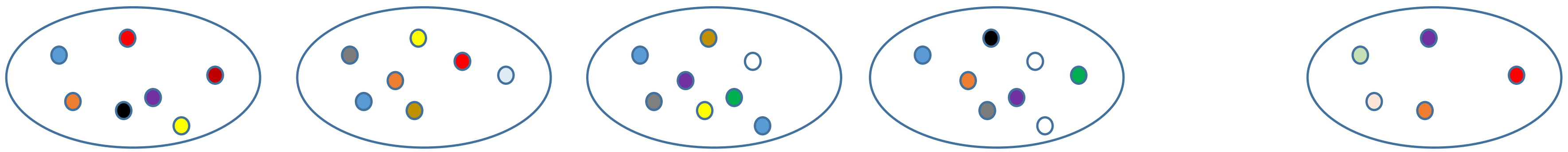
- If c(F) is polynomial, we need to perform O(2^n) iterations to find a satisfying assignment

➔ Require better sampling techniques that find a few satisfying assignments

FPRAS for DNF

- How to *efficiently* estimate the c(F)?
- Consider a DNF, F=$C_1 + C_2 + \cdots + C_t$
  - If any of clause is satisfied, then F is satisfied
  - Assume $C_i = x_1 \cdot \bar{x}_2 \cdot x_3$ ➔ $x_1 = T,\ x_2 = F,\ x_3 = T$
    ➔ Other literals such as $x_4,\ x_5,..$ can be either T/F
  - If there are n literals, then there are 2^{n-3} satisfying assignments
  - Let SC_i be a set of satisfying assignments of C_i that consists of l_i literals
  - ➔ $|SC_i|$=2^{n-l_i}
- Let
  - U= {(i, a) | $1 \le i \le t\ and\ a \in SC_i$}
  - Let S be the set of distinctive assignments that satisfy F
    - $S = \cup_{i=1}^{t} SC_i$
    - C(F) = $|\cup_{i=1}^{t} SC_i| \le |U|$

Note: A same assignments may occur many times in U



SC_1 SC_2 SC_3 SC_4 SC_t

FPRAS for DNF

- How to estimate c(F) (= $|S| = |\bigcup_{i=1}^{t} SC_i|$)?
  - We know the size of U= {(i, a) | $1 \le i \le t$ and $a \in SC_i$}
    - |U| = $\sum_{i=1}^{t} |SC_i|$
    - It is easy to find SC_i(and $|SC_i|$), but the same satisfying assignment can appear in many SC_i
  - How many times a same satisfying assignment occur in different clauses?
  - Estimate |U|/|S|
- Sketch of a Monte Carlo simulation scheme
  - Select an assignment in SC_i, and check if it appear in other SC_j, then systematically remove it from the set
  - ➔ Count only the first appearance
  - S={(i, a) | $1 \le i \le t,\ \ a \in SC_i,\ \ a \notin SC_j, for\ j < i$}

Uniform sample over SC_i
➔ $|SC_i| / \sum_i |SC_i|$

- Sampling method
  - Selection of (i, a) pairs
  - First sample i and then sample a in SC_i
  - Then examine if it satisfies $SC_j, for\ j < i$

FPRAS for DNF

DNF Counting Algorithm 2
1. $X \leftarrow 0$
2. For $k = 1, \dots, m$ do
   a) With probability $|SC_i|/\sum_i |SC_i|$, choose $a \in SC_i$
   b) If $a \notin SC_k$ for all $k < i$, $X \leftarrow X + 1$
3. Return $Y \leftarrow (X/m)\cdot \sum_i |SC_i|$

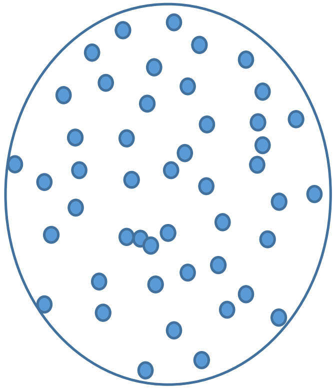
- Theorem:
  - The above algorithm is FPRAS for the DNF counting problem when $m = (3t/\varepsilon^2)\ln(2/\delta)$
- Proof
  - First show that sampling based on $|SC_i|/\sum_i |SC_i|$ is uniform sampling over |U|
    - Pr((i,a) is sampled) = Pr(i is sampled)·Pr(a is selected | i sampled)
      = ($|SC_i|$/|U|) · (1/$|SC_i|$) = 1/|U|
  - Prob. that a random sample passes the test 2 b)) $\geq 1/t$

  ➔ $\mu = E[X_i] \geq 1/t$

Note: $m \geq 3\ln\left(\frac{2}{\delta}\right)/\varepsilon^2\mu$

Sampling Method

- Probe the sample space uniformly
  - The DNF example showed that sampling method itself is as important as the main problem

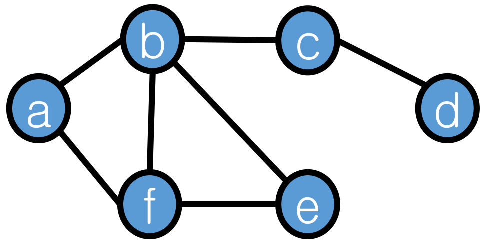


Sample space: Ω

- Definition: ε−Uniform sample of Ω
  - ω: Sampling instance
  - For any $S \subseteq \Omega$, $\left|\Pr(\omega \in S) - \frac{|S|}{|\Omega|}\right| \leq \varepsilon$
- Definition: **FPAUS**(Fully Polynomial Almost Uniform Sampler)
  - A sampling algorithm is FPAUS if, given an input x and parameter ε, it generates an ε−uniform sample of Ω(x) and running time is polynomial of $ln\varepsilon^{-1}$ and the size of the input x

Example: Independent Set

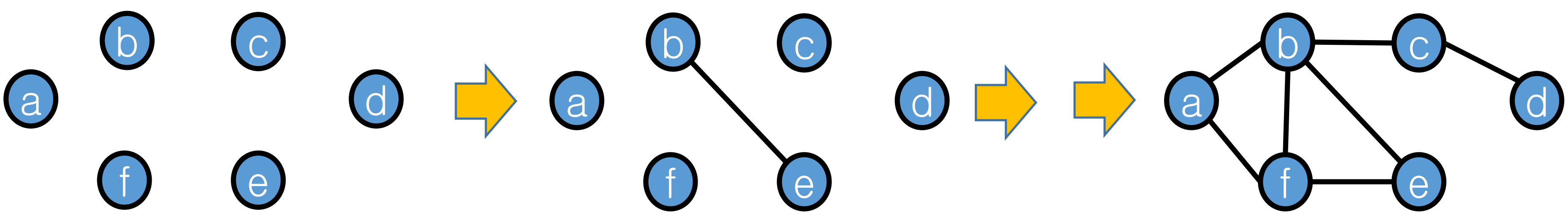
- Recall the **independent sets** of a Graph (Chapter 6)
  - A subset of nodes that are not directly connected



{a}, {a,c}, {b,d}, {a,c,e} are example of independent sets {{a,c,d} is not an indep. set

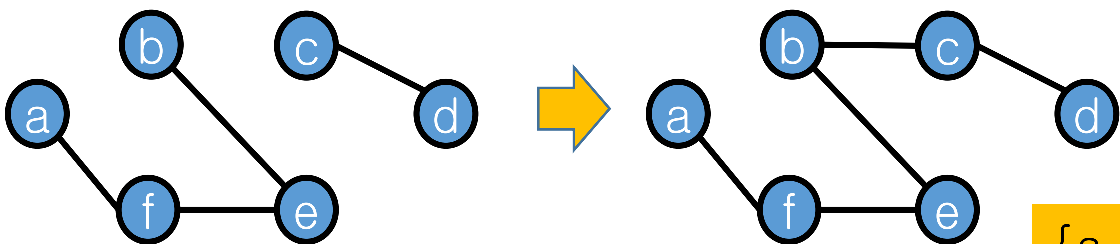
- **Estimating # independent sets** in a graph G=(V, E)

- How?
  - Start from a primitive case and proceed to the original graph



Example: Independent Set

- Suppose m=|E|, and randomly order the edges
- Define $G_i = (V, E_i)$ where E_i has the first i random edges
  - G_0: Graph with no edges
  - $G_m \equiv G$
- Let $\Omega(G_i)$ be the set of independent sets in G_i
- $|\Omega(G_0)| = ??$
  - Every subset of V is an independent set of G_o ➔ 2^n, where n = |V|



{a, b. c} is an indep. set of G_4, but not of G_5

- Note that G_i is derived from G_{i-1} by adding one randomly selected edge
  - Some of subsets ∈ $\Omega(G_{i-1})$ is no longer independent in G_i

Example: Independent Set

- $|\Omega(G_m)| = \frac{|\Omega(G_m)|}{|\Omega(G_{m-1})|} \cdot \frac{|\Omega(G_{m-1})|}{|\Omega(G_{m-2})|} \cdots \cdots \frac{|\Omega(G_2)|}{|\Omega(G_1)|} \cdot \frac{|\Omega(G_1)|}{|\Omega(G_0)|} \cdot |\Omega(G_0)|$

- Let $r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$

  ➔ $|\Omega(G_m)| = 2^n \cdot \prod_{i=1}^{m} r_i$

- Develop estimates $\tilde{r}_i$ for r_i such that the compound error R=$\prod_{i=1}^{m} \frac{\tilde{r}_i}{r_i}$ is bounded

  ➔ $\Pr(|R-1| \leq \epsilon) \geq 1-\delta$

(ε, δ)−approximation

Example: Independent Set

- Claim:

  If $\tilde{r}_i$ is an $(\varepsilon/2m, \delta/m)$–approximation for r_i (for i=1,2,⋯,m)

  ➔ Then $\Pr(|R-1| \leq \epsilon) \geq 1-\delta$

- Proof:

  – For each i, $\Pr\left(|\tilde{r}_i - r_i| \leq \frac{\varepsilon}{2m} r_i\right) \geq 1 - \frac{\delta}{m}$

  ➔ $\Pr\left(|\tilde{r}_i - r_i| > \frac{\varepsilon}{2m} r_i\right) < \frac{\delta}{m}$

  – $\Pr\left(\cup_{i=1}^{m}(|\tilde{r}_i - r_i| > \frac{\varepsilon}{2m} r_i)\right) \leq \sum_{i=1}^{K} \Pr\left(|\tilde{r}_i - r_i| > \frac{\varepsilon}{2m} r_i\right) < \delta$

  ➔ $\Pr\left(\cap_{i=1}^{m}(|\tilde{r}_i - r_i| \leq \frac{\varepsilon}{2m} r_i)\right) \geq 1 - \delta$

  ➔ $\Pr\left((1-\frac{\varepsilon}{2m})^m \leq \prod_{i=1}^{m} \frac{\tilde{r}_i}{r_i} \leq (1+\frac{\varepsilon}{2m})^m\right) \geq 1-\delta$

  – The lemma holds because $(1-\frac{\varepsilon}{2m})^m \geq 1-\varepsilon$, $\left(1+\frac{\varepsilon}{2m}\right)^m \leq 1+\varepsilon$

Example: Independent Set

- Estimation of r_i
  - Sample independent sets in $\Omega(G_{i-1})$ and compute # sets also belong to $\Omega(G_i)$

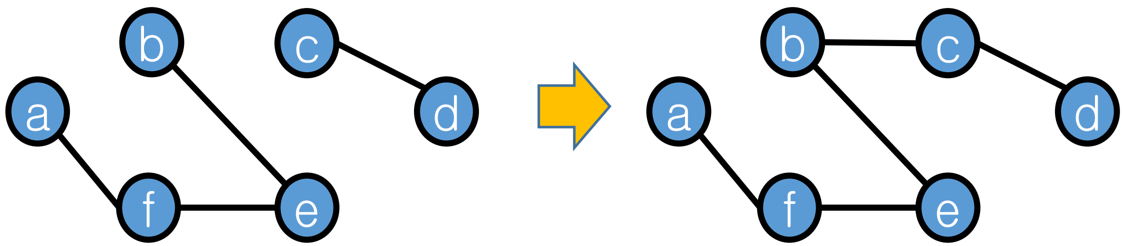
> Given G_i and G_{i-1}
> 1. $\mathrm{X} \leftarrow 0$
> 2. Repeat for M$(= 1296 \cdot m^2\varepsilon^{-2}\ln(2m/\delta))$ independent trials
>   a) Generate an $(\varepsilon/6m)$ uniform sample from $\Omega(G_{i-1})$
>   b) If the sample is independent set of G_i, $\mathrm{X} \leftarrow X + 1$
> 3. Return $\widetilde{r_i} \leftarrow X/M$

- Claim:
  - The procedure to estimate r_i is an $(\varepsilon/2m,\ \delta/m)-$approximation for r_i

First, prove the claim
Then, How to generate $\varepsilon-$Uniform sample?
➔ Markov Chain Monte Carlo Method

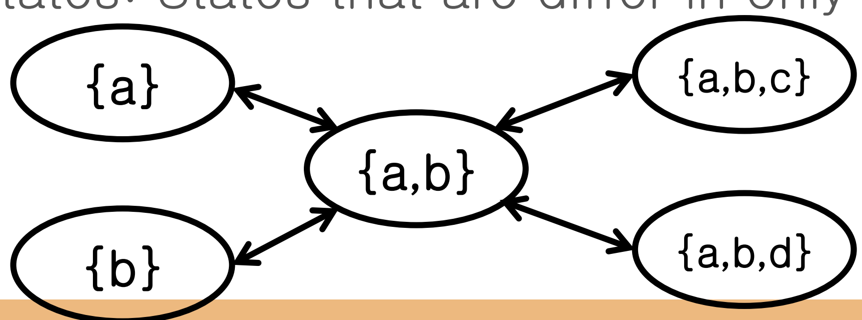
MC^2

- MCMC, MC^2: Markov Chain Monte Carlo
- Use MC that represents sample space for uniform sampler



Construction of MC
➔ Should know all indep. sets of G_{i-1} ➔ Impossible
➔ Dynamic transitions on imaginary MC

- Example
  - Consider Independent set of G_4
  - A state is an instance of independent set
  - Neighbor states: States that are differ in only one vertex



MC^2

- Given that an MC is irreducible and ergodic, its stationary distribution ≡ long-term probability of states

- Irreducible. Why?
  - Again, consider Independent set of G_4
    - Finite # states
    - Any two states are communicating

- Aperiodic
  - Add a self-loop to each state

- Uniform sampling
  - The visiting probabilities to all states are the same
  - Uniform stationary probabilities $(\pi_x = \pi_y)$

Uniform Distribution MC^2

- Assuming random walk over MC, how to define transition probabilities to obtain uniform stationary probabilities?
- Recall stationary prob. of RW is $\pi_u = \frac{d_u}{2|E|}$

  ➔ All states must have the same degree

Problem: Degrees (# neighbor states) of states are different
Solution: Equal transition probabilities to all neighbor states
Add self-loops

- Claim:
  - Let M is the largest degree and define transition probability as

$$P_{x,y} = \begin{cases} 1/M, & \text{if } x \neq y \text{ and } y \text{ is a neighbor of } x \\ 0, & \text{if } x \neq y \text{ and } y \text{ is not a neighbor of } x \\ 1-N(x)/M, & \text{if } x = y \end{cases}$$

  Then the stationary distribution is uniform distribution
- Proof:
  - If $\pi_x = \pi_y$, then $\pi_x P_{x,y} = \pi_y P_{y,x}$ since $P_{x,y} = P_{y,x} = 1/M$ ➔ Time reversible and $\pi_x = \pi_y = 1/|\Omega|$ is the stationary distribution

MC^2

- Generally, it is impossible (or impractical) to enumerate all states

→ Instead of pre-defining the entire MC, make impromptu transitions from the current state

→ Randomly select a neighbor state from the current state

- Let $X_0, X_1, \ldots, X_n$ be a sequence of transitions
- For large r, X_t $(t \geq r)$ distributed like the stationary distribution
- Sample at $X_r, X_{2r}, X_{3r}, \cdots$ transitions
- Efficiency of sampler
  - How large is r?
  - Easy of transitions

Example: Uniform Distribution MC^2

- Apply MC^2 to independent set sampling

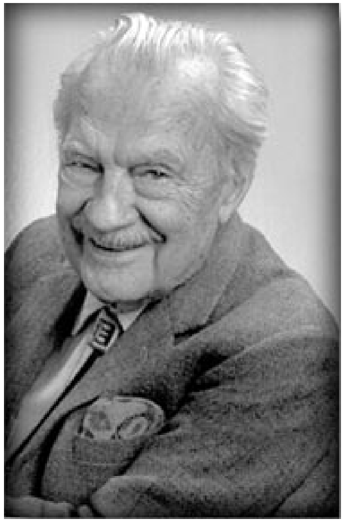
Start from arbitrary independent set X_0

1. From state X_i, find the next state X_{i+1} as follows
   a) Choose a vertex (v) randomly from V
   b) If $v \in X_i$, then $X_{i+1} \leftarrow X_i$ −{v}
   c) else if $v \notin X_i$ and X_i +{v} is still an independent set, then $X_{i+1} \leftarrow X_i$ +{v}
   d) else $X_{i+1} \leftarrow X_i$

- Properties of the MC
  - Irreducible?
  - Aperiodic?
  - Transition probability $P_{x,y}$? (Or what is the value of M?)

Metropolis Algorithm

Nicholas Metropolis (1915~1999) was an American Physicist, Mathematician who developed Monte Carlo method with his team (including von Neumann) at LANL



- Want to assign **Non-uniform distribution**
- Claim:
  - Let $M \geq \max_{x \in \Omega} N(x)$ and let π_x be the desired stationary probability of state x
  - Define MC such as

$$P_{x,y} = \begin{cases} (1/M) \cdot \min\left(1, \pi_y / \pi_x\right), & if\ x \neq y\ and\ y\ is\ a\ neighbor\ of\ x \\ 0, & if\ x \neq y\ and\ y\ is\ not\ a\ neighbor\ of\ x \\ 1 - \sum_{y \neq x} P_{x,y}, & if\ x = y \end{cases}$$

Metropolis Algorithm

- Proof
  - If $\pi_x < \pi_y$, then $P_{x,y} = 1/M$ and $P_{y,x} = (1/M) \cdot \pi_x/\pi_y$

    ➔ $\pi_x P_{x,y} = \pi_y P_{y,x}$
  - Similarly, $\pi_x P_{x,y} = \pi_y P_{y,x}$ for $\pi_x > \pi_y$

- Application: Independent set
  - Want to assign larger (or smaller) probability in proportion to the independent set size

    ➔ $\boldsymbol{\pi_x \propto \lambda^{|I_x|}}$

Start from arbitrary independent set X_0

1. From state X_i, find the next state X_{i+1} as follows
   a) Choose a vertex (v) randomly from V
   b) If $v \in X_i$, then $X_{i+1} \leftarrow X_i - \{v\}$ w/ probability min(1,1/λ)
   c) else if $v \notin X_i$ and $X_i + \{v\}$ is still an independent set, then $X_{i+1} \leftarrow X_i + \{v\}$ with probability min(1, λ)
   d) else $X_{i+1} \leftarrow X_i$

Review–Probabilistic Method

- A new field of mathematics originated by Erdos in 1940s
- Prove the existence of events with certain properties
  - Some methods are constructive
- Very useful (Powerful) in CS
  - Many CS (optimization) problems are NP-Hard ➔ We developed heuristic solutions? ➔ How good is the solutions?
- Methods
  - Basic counting
  - Expectation
  - Derandomization using conditional expectation
  - Sample & Modify
  - Second moment
  - Conditional expectation inequality
  - LLL

Review–MCRW

- Many (or most) CS problems are concerned with dynamics of systems rather than static phenomena

  ➔ Modeled as stochastic (Random) process

- Markov process
  - A stochastic process with the memoryless property
- Transition probability and stationary distribution
  - Conditions to have a stationary distribution
    - Irreducible
    - Ergodic (Positive recurrent, aperiodic)
- Computation of stationary distribution
- Random Walk
  - Evidence of transitions but transition probabilities are not known

Review–Cont. Distribution and Poisson Process

- Continuous distribution
  - Uncountable sample space
- Like the discrete case, we have
  - Joint distribution
  - Conditional probability
    - Marginal distribution
- Examples of continuous distribution
  - Uniform
  - Exponential
- Stochastic counting process
- Poisson process
  - Number of arrivals in a time interval has the Poisson distribution

Review–Cont. Distribution and Poisson Process

- Interarrival time of Poisson process
  - Exponential distribution
  - Memoryless property
- Combining and splitting of Poisson process
- CTMC (Continuous Time Markov Chain)
  - Transitions at each state is Poisson
  - M/M/1
- Queueing theory
  - Performance of queueing (= shared) systems
  - Little's Theorem: $N=\lambda T$