Randomized Algorithms

Lecture 20: Revision Lecture

Raul Garcia-Patron

School of Informatics
University of Edinburgh

EXAM INFORMATION

- Date: Monday, 11th December 2023. Time: 1:00-3:00 p.m.
- Answer two (2) questions, each question has equal weight
- *This is a NOTES AND CALCULATORS PERMITTED examination: candidates may consult up to THREE A4 pages (6 sides) of notes. CALCULATORS MAY BE USED IN THIS EXAMINATION*
- Read all questions before choosing and be “strategic”
- Do good time management!
- Past papers: mostly last year. Previous years other instructors.
- Study Tutorials and train with past year exam, check assignments.

Lecture 11 - Stochastic processes

- A *Stochastic process* is a collection of random variables $\mathbf{X} = \{X_t : t \in T\}$ (usually $T = \mathbb{N}^0$).
- X_t is the state of the process at time $t \in T$:

Definition (Marov chain)

A discrete-time stochastic process is said to be a *Markov chain* if

$$\Pr[X_t = a_t \mid X_{t-1} = a_{t-1}, \ldots, X_0 = a_0] = \Pr[X_t = a_t \mid X_{t-1} = a_{t-1}].$$

- Matrix representation
- Graph representation
- Iterations: $\bar{p}_t = \bar{p}_0 \cdot M^t$.
- Stationary distribution: $\pi = \pi P$.

Lecture 11 - Convergence

Theorem (Convergence and detailed balance)

Consider a **finite**, **irreducible**, *and* **aperiodic** *Markov chain with transition matrix P.If there is a probability distribution π that for each pair of state i, j satisfies* **detailed balance** *(time reversible chains)*

$$\pi_i P_{i,j} = \pi_j P_{j,i},$$

then π is the unique stationary distribution corresponding to P.

Lemma (Irreducible)

A finite Markov chain is **irreducible** *if and only if its graph representation is a strongly connected graph.*

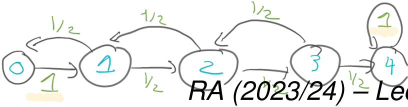
- Curing periodicity: use of self-loops.

Lecture 12 - 2-SAT Randomized Algorithm

Algorithm 2SATRANDOM$(n; C_1 \wedge C_2 \wedge \ldots \wedge C_\ell)$

1. Assign *arbitrary* values to each of the x_i variables.
2. $t \leftarrow 0$
3. **while** ($t < 2mn^2$ **and** some clause is unsatisfied) **do**
4. Choose an *arbitrary* C_h from all unsatisfied clauses;
5. Choose one of the 2 literals in C_h *uniformly at random* and flip the value of its variable;
6. **if** (we end with a satisfying assignment) **then**
7. **return** this assignment to the $x_1, \ldots x_n$ **else**
8. **return** FAILED.

- **Key idea:** instead of quantifying success by number of clauses satisfied by distance (Hamming wight) to a given solution.
- Analyse its performance by analogy to a *Markov chain*.



Lecture 12 - Probability of failure

Theorem

The algorithm 2SATRANDOM *perform up to* $2mn^2$ *iterations of the while loop. Then, when there is a satisfying assignment for* ϕ*, the probability that* 2SATRANDOM *does not discover one, is at most* 2^{-m}*.*

Proof.

1. Modify the algorithm to be run m times in parallel over “blocks" of $2n^2$ size.
2. Markov inequality guarantees a failure of $1/2$ for $2\mathrm{E}[Z_0] = 2n^2$ iterations per block: $P(Z_0 > a) \leq \frac{\mathrm{E}[Z_0]}{2}$, choose $a = 2\mathrm{E}[Z_0]$.
3. If one of the m repetition succeeds we find the solution. We get failure overall only if all the m blocks fail, i.e., $P_f = (1/2)^m = 2^{-m}$.

□

The algorithm 2SATRANDOM run the $2mn^2$ in a single loop, but this can only reach the solution faster: instead of imputing m independent input to each block, we can feed one block with the output of the previous one.

Lecture 13 - Monte Carlo Method

Definition (Generalization (Theorem 11.1))

Let $X_1, \ldots, X_m$ be independent and identically distributed indicator random variables (ie Bernoulli with a fixed parameter), and $\mu = \mathrm{E}[X_i]$. Then if $m \geq \frac{3\ln(\frac{2}{\delta})}{\epsilon^2 \mu}$, we have

$$\Pr\left(\left| \frac{1}{m} \sum_{i=1}^{m} X_i - \mu \right| \geq \epsilon\mu \right) \leq \delta.$$

So for this m, sampling gives a (ϵ, δ)-approximation of μ.

Definition (FPRAS (Definition 11.2))

A *fully polynomial randomized approximation scheme (FPRAS)*:

- Given input x, we want $(\epsilon, \delta)-$approximation of $V(x)$.
- Achieved in time polynomial in $1/\epsilon$, in $\ln(1/\delta)$ and size of x.

Lecture 13 - The DNF counting problem

Disjunctive Normal Form (DNF):

- each *clause* is now a *conjunction* ($\wedge$, AND) literals
- we have disjunctions ($\vee$, OR) of clauses

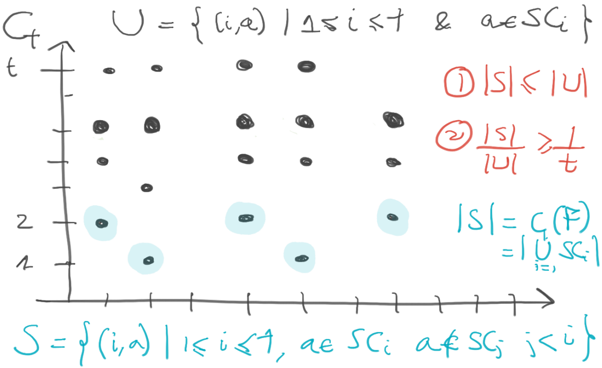
For example:

$$(x_1 \wedge \bar{x}_2 \wedge x_3) \vee (x_2 \wedge x_4) \vee (\bar{x}_1 \wedge x_3 \wedge x_4).$$

We are interested in *counting the number of satisfying assignments*.

- It is easy to find satisfying assignments or prove not satisfiable.
- It is NP-hard to compute the *exact* number of satisfying assignments for a DNF.
- However, we can approximately count them.
- Naive sampling solution over 2^n does not work!
- We need to sample from an initial non-trivial distribution

Lecture 13 - Relations between sets



- We know how to compute $|U| = \sum_{i=1}^{t} s^{n-|C_i|}$
- S is approx. of same size as U: $\frac{|S|}{|U|} \geq \frac{1}{t}$. Key to make the sampling algorithm efficient.

Lecture 14 - Markov chain Monte Carlo (MCMC)

The Markov chain Monte Carlo (MCMC) method provides a very general approach to sampling from a desired probability distribution.

- The idea is to build a *Markov chain M* on the state space Ω that we want to sample from.
- We ensure the stationary distribution of the Markov chain is unique and corresponds to the target distribution.
- We can then run M to generate a sequence of $X_0, X_1, \ldots, X_k$ of states so X_k distribution is the stationary distribution: x_k is our output sample.
- How large k has to be to have a valid sample is called mixing-time.
- Knowing the mixing-time of a Markov chain is non-trivial and will be the core of the last section of the course.

Lecture 14 - Approximate counting

Now consider the following *telescoping product*:

$$|\Omega(G)| = \frac{|\Omega(G_m)|}{|\Omega(G_{m-1})|} \times \frac{|\Omega(G_{m-1})|}{|\Omega(G_{m-2})|} \times \frac{|\Omega(G_{m-2})|}{|\Omega(G_{m-3})|} \times \ldots \times \frac{|\Omega(G_1)|}{|\Omega(G_0)|} \times |\Omega(G_0)|.$$

- $|\Omega(G_0)| = 2^n$ as every subset of V is an I.S. for G_0 (G_0 has no edges!).
- We will show how to obtain close approximate values $\tilde{r}_i$ for each ratio $r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$, for $i = 1, \ldots, m$.
- Our *estimate* for the number of I.S.s will be:

$$2^n \prod_{i=1}^{m} \tilde{r}_i.$$

We may want to sample from a nonuniform distribution.

For a finite state space Ω and neighborhood structure $\{N(X)|x \in \Omega\}$, let $N = \max_{x\in\Omega} |N(x)|$. Let M be any number such that $M > N$. For all $x \in \Omega$, let $\pi_x > 0$ be the desired probability of state x in the stationary distribution. Consider a Markov chain where

$$P_{x,y} = \begin{cases} 1/M \min(1, \pi_y/\pi_x) & \text{if } x \neq y \text{ and} y \in N(x). \\ 0 & \text{if } x \neq y \text{ and} y \notin N(x). \\ 1 - \sum_{y\neq x} P_{x,y} & \text{if } x = y \quad (\textit{rejection}) \end{cases} \tag{1}$$

MC **finite**, **irreducible** and **aperiodic**: π stationary distribution.

- We do not need to know π_x or π_y, but only their ratio π_x/π_y!

Lecture 15 - Glauber Dynamics

- Let V and S be finite sets and suppose $\Omega \subseteq S^V$.
  Ex: V vertices of a graph and S a set of colors (graph coloring).
- Let π be a probability distribution whose support is Ω.
- The Glauber chain moves from state x as follows:
  1. An element v is chosen uniformly at random from V.
  2. A new state y is chosen s.t.: $y(w) = x(w) \, \forall \, w \neq v$.

Definition

Given $x \in \Omega, v \in V : \Omega(x, v) = \{y \in \Omega : y(w) = x(w) \, \forall \, w \neq v\}$. The chain transition reads:

$$P_{x,y} = \pi(y|\Omega(x, v)) = \begin{cases} \frac{\pi(y)}{\pi(\Omega(x,v))} & \text{if } y \in \Omega(x, v) \\ 0 & \text{if } y \notin \Omega(x, v) \end{cases} \tag{2}$$

Lecture 16 - TV and Coupling

Our goal: We want to sample from a MC with stationary distribution π in time poly(n) and $\log(1/\epsilon)$.

- TV distance: $\|D_1 - D_2\| \;=\; \frac{1}{2}\sum_{x\in\Omega}|D_1(x) - D_2(x)|$
- Lower bound mixing time:
  $|D_1(A) - D_2(A)| \leq \max_{A\subseteq\Omega}|D_1(A) - D_2(A)| = \|D_1 - D_2\|$
- A coupling of two probability distributions μ and ν is a pair of random variables (X, Y) defined on a single probability space, i.e., a joint probability distribution q on $\Omega \times \Omega$ such that

$$\sum_{y\in\Omega} q(x,y) = \mu(x) \mathit{and} \sum_{x\in\Omega} q(x,y) = \nu(x) \tag{3}$$

- Upper-bounds on TV: $\|D_1 - D_2\| \leq \inf \Pr(X \neq Y)$ for a coupling (X, Y) of D_1 and D_2.
- Mixing time: we want to prove that $\|P^t(x,\cdot) - \pi\| \leq \epsilon$ fast enough
- Coupling lemma:
  $\Pr(X_T \neq Y_T | X_0 = x, Y_0 = y) \leq \epsilon \quad \Rightarrow \quad \tau(\epsilon) \leq T$

Lecture 17 - Mixing time via contraction of distance

Neighborhood: states $y \in \Omega$ reachable from x in a single step.

- Distance $d(X, Y)$: the amount of steps to reach y from x. Neighbors if $d(X, Y) = 1$. Many times $d(X, Y) \leq |V|$.
  - Distance at step t of MC: $d_t = d(X_t, Y_t)$
- $\Pr(X_T \neq Y_T | X_0 = x, Y_0 = y) \leq \max_{x,y} \Pr(d_T \geq 1) \leq \max_{x,y} E[d_T]$
- Our goal is to bound $\max_{x,y} \mathbf{E}[d_T] \leq \epsilon$.
- After some work... (see next slides) $\mathbf{E}[d_{t+1}] =\leq \beta \mathbf{E}[d_t]$, with $\beta < 1$ (Contraction of expect. distance)
- Iterate $\mathbf{E}[d_T] =\leq \beta^T d_0 \leq \beta^T |V|$

Therefore the chain is guaranteed to have mixed for all times, such that $\beta^T |V| \leq \epsilon$, leading to

$$\tau(\epsilon) = \frac{1}{\log(1/\beta)} \left(\log |V| + \log(1/\epsilon)\right).$$

Many times we can write $\beta = e^{-\alpha/|V|}$, leading to $\tau(\epsilon) = \frac{|V|}{\alpha} \left(\log |V| + \log(1/\epsilon)\right)$, where α can itself depend on parameters of the problem.

Lecture 17 - Path Coupling

- We define a distance $d(X, Y)$. Neighbors if $d(X, Y) = 1$.
- Our goal is to prove concentration of expect. of distance: $\mathbf{E}[d_{t+1}] =\leq \beta\mathbf{E}[d_t]$, with $\beta < 1$.
- Path: $X_t = Z_0, Z_1, ..., Z_{d_t} = Y_t$ where $d(Z_{i+1}, Z_i) = 1$
  - $d_t = \sum_{i=1}^{d_t} d(Z_{i+1}, Z_i)$ (by construction)
- Updated path: $X_{t+1} = Z'_0, Z'_1, ..., Z'_{d_t} = Y_{t+1}$.
  - $d_{t+1} \leq \sum_{i=1}^{d_t} d(Z'_{i+1}, Z'_i)$ (by triangle inequality)

1. For your problem of interest prove $\mathbf{E}[d(Z'_{i+1}, Z'_i)] \leq \beta d(Z_{i+1}, Z_i) = \beta$.
2. Leads to $\mathbf{E}[d_{t+1}|d_t] \leq \sum_{i=1}^{d_t} \mathbf{E}[d(Z'_{i+1}, Z'_i)] \leq \beta d_t$.
3. Then $\mathbf{E}[d_{t+1}] = \mathbf{E}[\mathbf{E}[d_{t+1}|d_t]] =\leq \beta\mathbf{E}[d_t]$.

Left to do: prove $\mathbf{E}[d(Z'_{i+1}, Z'_i)] \leq \beta$ for our problem of interest.

Lecture 18 - Ising Model

A **spin system** is a probability $p(\sigma)$ of configurations $\sigma \in \{+1, -1\}^V$, defined in a graph $G = (V, E)$.

- Interpretation as magnets: $\{+1, -1\}$ being the orientation.
- The nearest-neighbor **Ising model** is the most studied spin system. The energy of a configuration σ is defined to be:

$$H(\sigma) = \sum_{v,w \in V : (v,w) \in E} z_{v,w} \sigma(v) \sigma(w) + \sum_{z \in V} h_z \sigma(z) \tag{4}$$

- Minimization of the energy maps to NP-hard problems:
  - Ferromagnetic
  - Anti-ferromagnetic (MAX-CUT)
  - Independent sets