Probability and Computing Notes and Questions

Angjoo Kanazawa

August 23, 2011

1 Chapter 6 The Probabilistic Method

1.1 Notes

- To prove teh existence of an object with specific properties, construct an approximate probability space of objects S, and show that the probability that an object (in S with the specific proprties) is selected is > 0. Strictly.
-

1.2 Exercises

6.1 Consider an instance of SAT with m clauses, where every clause has exactly k literals.

(a) Give a Las Vegas algorithm that finds an assignment that satisfies at least $m(1 - 2^{-k})$ clauses, anayze its expected running time:

The first part is straight from the book (6.2.2). Note:

$$P(\text{a clause is satisfied}) = 1 - P(\text{all literals are false}) = 1 - 2^{-k}$$

Let $i = 1 \ldots m$, $X_i = 1$ if ith clause is satisfied, $X_i = 0$ otherwise. Let $X = \sum_i^m X_i$, the total number of satisfied clauses. So we get

$$E(X) = \sum_i^m X_i P(X_i = 1) = m(1 - 2^{-k}) = \mu$$

Now given this, the LV algorithm goes like this:

Repeat until $X \geq \mu$:

- assign values to all boolean variables independently and uniformly.
- Check the value of X.

What is the expected number of runs until LV finishes? We're done when $X \geq \mu$. Note that X has a binomial distribution*. Let Y be the number of runs needed for the LV to terminate, i.e. number of trials before the first sucess. So Y has a geometric distribution and let $\hat{p}$ be the probability of success at each trial i.e. $\hat{p} = P(X \geq \mu)$. Then, $E(Y)$, what we want, is*

$$E(Y) = \sum_{I=1}^{\infty} i\hat{p}(1 - \hat{p})^{i-1} = 1/p \tag{6.1}$$

So for each run, what is $\hat{p} = P(X \geq \mu)$? Using $p = P(\text{a clause is satisfied}) = 1 - 2^{-k}$:

$$\begin{aligned}P(X \geq \mu) &= 1 - P(X < \mu) \\ &= 1 - [P(X=0) + P(X=1) + \cdots + P(X = \mu - 1)] \\ &= 1 - [(1-p)^m + \binom{m}{1} p(1-p)^{m-1} + \cdots + \binom{m}{\mu-1} p^{\mu-1}(1-p)^{m-(\mu-1)}] \\ &= 1 - \sum_{i=0}^{\mu-1} \binom{m}{i} p^i (1-p)^{m-i}\end{aligned}$$

Now recall $\sum_i^m \binom{m}{i} = 2^m$, so $\sum_i^{\mu-1} \binom{m}{i} \leq \frac{2^m}{2} = 2^{m-1}$.
Also, notice that here $p^i(1-p)^{m-i} = (1-2^{-k})^i(1-2^{-k})^{m-i} = (1-2^{-k})^m$.
So we can continue the above inequalitywith:

$$\begin{aligned}P(X \geq \mu) &= 1 - \sum_{i=0}^{\mu-1} \binom{m}{i} p^i (1-p)^{m-i} \\ &\geq 1 - [2^{m-1}(1-2^{-k})^m] = 1 - 2^{-k}\end{aligned}$$

As an example, with $k = 1$ $\hat{p}$ is 1/2. Using 6.1, we get $E[Y] = 2$.
.. If $E[Y] = \frac{1}{1-2^{-k}}$, the bigger the k, the faster the algorithm.. But that makes sense.
Can one always say $P(X \geq \mu) \geq 1/2$?

(b) Give a derandomization of the randomized algorithm using the method of conditional expectations: This I also just followed the book. Maybe too closely.

We know setting variables independently and uniformly gives us $E(X) \geq m(1-2^{-k})$. Now set the boolean variables $x_1, x_2, \ldots$ up to r deterministically one at a time.

Consider the expected total # of satisfied clauses if the remaining boolean variables are selected independently and uniformly. Write this as $E(X|x_1, x_2, \ldots, x_r)$. We want a away o set the next variable s.t.

$$E(X|x_1, \ldots, x_r) \leq E(X|x_1, \ldots, x_r, x_{r+1}) \tag{6.2}$$

Inductively, the base case is $E(X|x_1) = E(X)$. Now, consider setting x_{r+1} randomly to true or false. Each has probability 1/2. So $E(X|x_1, \ldots, x_r) = \frac{1}{2}E(X|x_1, \ldots, x_{r+1} = 1) + \frac{1}{2}E(X|x_1, \ldots, x_{r+1} = 0)$ From this we can deduce

$$\max(E(X|x_1, \ldots, x_r, x_{r+1} = 1), E(X|x_1, \ldots, x_r, x_{r+1} = 0)) \geq E(X|x_1, \ldots, x_r)$$

So we just have to chose the assignment that increases the conditional expectation the most.. we only have two options x_{r+1} is T or F, so look at clauses that contain the x_{r+1} variable twice and see how the expectation changes based on the assignment and take the better one? Something like that.

2 Chapter 7 Markov Chains and Random Walks

2.1 Markov Chains: Definitions and Representations

- **Stochastic process** $\mathbf{X} = \{X(t) : t \in T\}$ is a collection of random variables. Often t represents time, where $\mathbf{X}$ models the value of a r.v. X that changes over time.

- $X(t)$, or X_t is the *state* of the process at time t.
- If X_t assumes values from a countably infinite set, $\mathbf{X}$ is a **discrete space process**. If X_t assumes values from a finite set, then $\mathbf{X}$ is a **finite process**. If T is a countably finite set, then $\mathbf{X}$ is a **discrete *time* process**.
- Here we focus on a process where X_t depends on X_{t-1}, but no other states.
- **Definition:** a *discrete time stochastic process* $X_0, X_1, \ldots$ is a Markov chain if

$$\begin{aligned} Pr(X_t = a_t | X_{t-1} = a_{t-1}, X_{t-2} = a_{t-2}, \ldots, X_0 = a_0) &= P(X_t = a_t | X_{t-1} a_{t-1}) \\ &= P_{a_{t-1}, a_t}. \end{aligned}$$

  What we said on the point above. X_t depends on X_{t-1} but not on any other past. How we got to X_{t-1} is irrelevant, because all the dependency of X_t is captured by X_{t-1}, and this is the *Markov property.*
- the *transition probability* $P_{i,j} = P(X_t = j | X_{t-1} = i)$ is the probability that the process moves from i to j in *one step.* The markov property implies that the Markov chain is uniquely defined by the one step *transition matrix*, $\mathbf{P}$, made up of $P_{i,j}$s. $\forall i, \sum_{j \geq 0} P_{i,j} = 1$.
- The transition matrix is used to computing the distribution of future states of the process. if $p_i(t)$ denotes hte probability that the process is at state i at time t, and $\bar{p}(t)$ be the vector of the distribution of the state of the chain at time t (future), summing over al possible states at time $t-1$ we have

$$\bar{p}(t) = \bar{p}(t-1)\mathbf{P}$$

- for any n, the *n-step transition probability* $P^n_{i,j} = P(X_{t+n} = j | X_t = i)$ is the probability that the chain moves from state i to state j in exactly n steps.
- using the first transition from i, we get

$$P^n i, j = \sum_{k \geq 0} P_{i,k} P^{n-1}_{k,j}$$

- So $P^{(n)}$ is the n-step transition matrix, where the entry in the ith row and jth col is $P^n_{i,j}$. Then from the above we have $\mathbf{P^{(n)}} = \mathbf{P} \cdot \mathbf{P^{(n-1)}}$. Keep on doing this, we see that $\mathbf{P^{(n)}}$ is just $\mathbf{P^n}$.
- So, $\forall t \geq 0$ and $n \geq 1$,

$$\bar{p}(t+n) = \bar{p}(t)\mathbf{P}^n$$

- People represent a Markov chain by a directed, weighted graph $D = (V, E, w)$, where V is the set of states of the chain. $\exists$ an edge $(i,j) \in E$ iff $P_{i,j} > 0$, where the weight, $w(i,j)$ of that edge is $P_{i,j}$. Self-loops are okay. For each i (a state), $\sum_{(i,j) \in E} w(i,j) = 1$. A sequence of visited states is a directed path on D.
- Example, to calculate the probability of going from state 0 to state 3 *in 3 steps* is the $(0,3)$ entry in $\mathbf{P}^3$. The probability of ending in state 3 after 3 steps where the beginning state if chosen uniformly at random from n states is $(1/n, 1/n, \ldots, 1/n)\mathbf{P}^3$. where the $|V|$ long row vector $(1/n, 1/n, \ldots, 1/n)$ is the $\bar{p}(0)$ here. (Remember it's right multiplication). That was just $\bar{p}(0+3) = \bar{p}(0)\mathbf{P}^n$.