8^{TH} TUTORIAL ON RANDOMIZED ALGORITHMS

Eigenvalues of adjacency matrices III. + DNF counting.

1. *Graph powers and eigenvalues.* Let G be an undirected graph. For $k \geq 1$, consider the k-th power of G, denoted $G^{(k)}$, defined as having the same set of vertices and an edge for every walk of length exactly k in G (the graph will have loops and parallel edges). Express the eigenvalues of $G^{(k)}$ in terms of the eigenvalues of G.

2. *Monte Carlo estimation of π.* Consider a circle of diameter 1 enclosed within a square with sides of length 1. We will sample N points (uniformly and independently) from the square and set the indicator variable $X_t = 1$ if the t-th point is inside the circle, and set $X_t = 0$ otherwise. It is clear that $E[X] = N \cdot \pi/4$, where X is the sum of N of these indicator variables.

Give an upper bound on the value of N for which $4X/N$ gives an estimator of π that is accurate to d digits, with probability at least $1 - \delta$.

3. *Naïve sampling for DNF counting.* Suppose we have a class of instances of the DNF satisfiability problem, i.e., for each n, a formula with n variables, such that there are $\alpha(n)$ satisfying truth assignments for some polynomial α. Suppose we apply the naïve approach of sampling assignments and checking whether they satisfy the formula. Show that, after sampling $2^{n/2}$ assignments, the probability of finding even a single satisfying assignment for a given instance is exponentially small in n.

4. Consider the following variant of the *Coverage algorithm* for approximating the DNF counting problem. For $t = 1, \dots, N$,

- select a clause C_t at random with probability proportional to the number of satisfying truth assignments (recall how to count these numbers),
- select a satisfying truth assignment a for C_t uniformly at random (how?), and
- define random variable $X_t = 1/|\text{cov}(a)|$, where $\text{cov}(a)$ denotes the set of clauses that are satisfied by a (there's always at least one).

Our estimator for $\#F$ (the number of satisfying assignments for the DNF formula) is

$$Y = \frac{\sigma}{N} \cdot \sum_{t=1}^{N} X_t \,,$$

where σ is the sum of the sizes of the coverage sets $\text{cov}(a)$ over all satisfying assignments a (how to calculate σ?). Prove that Y is an (ε, δ)-approximation for $\#F$ for a sufficiently large N.

Chernoff Bounds

Theorem 1 (Multiplicative Chernoff Bound – Upper Tail)**.** *Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables (i.e., $X_i \in \{0,1\}$). Let $X = \sum_{i=1}^n X_i$ and let $\mu = \mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i]$. Then, for any $\delta > 0$,*

$$\Pr\big[X \geq (1+\delta)\mu\big] \;\; \leq \;\; \exp\Big(-\tfrac{\delta^2\mu}{3}\Big).$$

Theorem 2 (Multiplicative Chernoff Bound – Lower Tail)**.** *Under the same assumptions as Theorem 1, for $0 < \delta < 1$,*

$$\Pr\big[X \leq (1-\delta)\mu\big] \;\; \leq \;\; \exp\Big(-\tfrac{\delta^2\mu}{2}\Big).$$

Theorem 3 (Additive Chernoff/Hoeffding Bound)**.** *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $[0,1]$. Let $X = \sum_{i=1}^n X_i$ and let $\mu = \mathbb{E}[X]$. Then for any $t > 0$,*

$$\Pr\big[X - \mu \geq t\big] \;\; \leq \;\; \exp\Big(-2t^2/n\Big),$$

and similarly

$$\Pr\big[X - \mu \leq -t\big] \;\; \leq \;\; \exp\Big(-2t^2/n\Big).$$

Hoeffding Bound (General Form)

Although the name "Hoeffding bound" is sometimes used interchangeably with the additive Chernoff bound above, the more general Hoeffding bound is as follows:

Theorem 4 (Hoeffding's Inequality)**.** *Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables where Y_i takes values in an interval of length R_i. Suppose $\mathbb{E}[Y_i] = \mu_i$. Let $S = \sum_{i=1}^n Y_i$ and $\mathbb{E}[S] = \sum_{i=1}^n \mu_i$. Then, for any $t > 0$,*

$$\Pr\big[|S - \mathbb{E}[S]| \geq t\big] \;\; \leq \;\; 2\exp\Big(-\tfrac{2t^2}{\sum_{i=1}^n R_i^2}\Big).$$