Randomized Algorithms 2023A – Lecture 10
Probabilistic Embedding into Dominating Trees (cont'd) and Importance Sampling*

Robert Krauthgamer

1 Probabilistic Embedding into Dominating Trees (cont'd)

We completed the proof of the following theorem (from last time).

Theorem 2 [Bartal'96, Fakcharoenphol-Rao-Talwar'03]: Every n-point metric admits a probabilistic embedding into dominating trees with distortion $O(\log n)$.

Recall that Algorithm B (described last time) computes a hierarchical decomposition $P_L, \ldots, P_1, P_0$, which in turn defines a dominating tree T, whose leaves are the points of X.

Analysis: We say that a center w *separates* a pair $\{x, y\} \subset X$ at level i if x, y are in the same cluster of P_{i+1} but in different clusters of P_i, and w is the center that "caused" this, i.e., the first point (according to π) that "captures" exactly one of x, y (at level i).

Lemma 6: For every $x, y \in X$,

$$\mathbb{E}[d_T(x,y)] \leq \sum_{i=0}^{L} \sum_{w \in X} \Pr[w \text{ separates } \{x,y\} \text{ at level } i] \cdot 2^{i+2}.$$

Proof: As seen in class, it follows easily form Lemma 3.

Lemma 7 (contribution of a single center): Fix $x, y \in X$. Arrange $X = \{w_1, \ldots, w_n\}$ in order of increasing distance from the set $\{x, y\}$ (breaking ties arbitrarily). Then

$$\forall s \in [n], \quad \sum_i \Pr[w_s \text{ separates } \{x,y\} \text{ at level } i] \cdot 2^{i+2} \leq O(\tfrac{1}{s}) \cdot d(x,y).$$

Completing the proof of Theorem 2: Plugging Lemma 7 into Lemma 6,

$$\forall x, y \in X, \quad \mathbb{E}[d_T(x,y)] \leq \sum_{s=1}^{n} O(\tfrac{1}{s}) \cdot d(x,y) \leq O(\log n) \cdot d(x,y).$$

*These notes summarize the material covered in class, usually skipping proofs, details, examples and so forth, and possibly adding some remarks, or pointers. The exercises are for self-practice and need not be handed in. In the interest of brevity, most references and credits were omitted.

QED.

Proof of Lemma 7: Was seen in class by carefully breaking the event into two events, roughly one about β and one about the ordering π.

Exer: Prove that Algorithm A outputs a partition P of X where all clusters have diameter at most 2^{i+1}, and for x, y with $d(x,y) \le 2^{i-3}$,

$$\Pr[x, y \text{ are in different clusters of } P] \le O\big(\log \tfrac{|B(\{x,y\},2^{i-1})|}{|B(\{x,y\},2^{i-3})|}\big) \cdot \tfrac{d(x,y)}{2^{i+1}},$$

where $B(\{x,y\}, r)$ is the set of points within distance at most r from $\{x,y\}$ (analogous to a ball of some radius around a point).

Observe that this gives the bound needed in Lemma 5 with $\alpha = O(\log n)$.

2 Importance Sampling

Sampling is often used to estimate a sum. When the variance is too large, this method can offer reduce the variance. The idea is to sample, instead of uniformly, in a "focused" manner that roughly imitates the contributions, but one of course has to "factor out" the bias in this sampling.

Setup: We want to estimate $z = \sum_{i\in[n]} z_i$ without reading all the z_i values. The main concern is that the z_i are unbounded, and thus most of the contribution might come from a few unknown elements. If we happen to have a "good enough" lower bound on each element z_i, then we can sample with probability $p_i \ge \Omega(\frac{z_i}{z})$.

Theorem 1 [Importance Sampling]: Let $z = \sum_{i\in[n]} z_i$, and $\lambda \ge 1$. Let $\hat{Z}$ be an estimator obtained by sampling a single index $\hat{i} \in [n]$ according to distribution $(p_1, \ldots, p_n)$ where $\sum_{i\in[n]} p_i = 1$ and each $p_i \ge \frac{z_i}{\lambda z}$, and setting $\hat{Z} = z_{\hat{i}}/p_{\hat{i}}$. Then

$$\mathbb{E}[\hat{Z}] = z \quad \text{and} \quad \sigma(\hat{Z}) \le \sqrt{\lambda}\,\mathbb{E}[\hat{Z}].$$

Proof: was seen in class.

Exer: Show that averaging $t = O(\lambda/\varepsilon^2)$ independent repetitions of the above approximates z within factor $1 \pm \varepsilon$ with success probability at least $3/4$.

Hint: use Chebyshev's inequality.

Exer: Prove a variant of Theorem 1, where each z_i is read independently with probability $q_i \ge \min\{1, t\frac{z_i}{z}\}$, in which case it contributes $\frac{z_i}{q_i}$ (and otherwise contributes 0). Show that with high probability, the number of values read is $O(\sum_i q_i)$ and the estimate is $(1 \pm O(1/\sqrt{t}))z$.

Hint: The difference is here we decide whether to read each z_i independently, while in Theorem 1 we read at each step exactly one value z_i.

2.1 Counting DNF solutions via Importance Sampling

Problem definition: The input is a DNF formula f with m clauses $C_1, \ldots, C_m$ over n variables $x_1, \ldots, x_n$, i.e., $f = \vee_{i=1}^m C_i$ where each C_i is the conjunction of literals like $x_2 \wedge \bar{x}_5 \wedge x_n$.

The goal is the estimate the number of Boolean assignments that satisfy f.

Theorem 2 [Karp and Luby, 1983]: Let $S \subset \{0,1\}^n$ be the set of satisfying assignments for f. There is an algorithm that estimates $|S|$ within factor $1+\varepsilon$ in time that is polynomial in $m + n + 1/\varepsilon$.

2.2 Counting: A first attempt

Random assignments: Sample t random assignments, and let Z count how many of them are satisfying. We can estimate $|S|$ by $Z/t \cdot 2^n$.

Formally, we can write $Z = \sum_{i=1}^t Z_i$ where each Z_i is an indicator for the event that the i-th sample satisfies f. We can see it is an unbiased estimator:

$$\mathbb{E}[Z \cdot 2^n/t] = \sum_{i=1}^t \mathbb{E}[Z_i] \cdot 2^n/t = |S|.$$

Observe that $\mathrm{Var}(Z) = \frac{1}{t^2} \sum_i \mathrm{Var}(Z_i \cdot 2^n) = \frac{1}{t} \mathrm{Var}(Z_1 \cdot 2^n)$. But even though we can use Chernoff-Hoeffding bounds since Z_i are independent, it's not very effective because the variance could be exponentially large than the expectation.

Exer: Show that the standard deviation of Z (for $t = 1$) could be exponentially large relative to the expectation.

2.3 Counting: A second attempt

Idea: We can bias the probability towards the assignments that are satisfying, but then we will need to "correct" the bias.

Let $S_i \in \{0,1\}^n$ be all the assignments that satisfy the i-th clause, hence $|S_i| = 2^{n-\mathrm{len}(C_i)}$.

Remark: The naive approach does not use the DNF structure at all. We can use this structure by writing $S = \cup_i S_i$, which can be expanded using the inclusion-exclusion formula, but it would be too complicated to estimate efficiently.

Algorithm E:

1. Choose a clause C_i with probability proportional to $|S_i|$ (namely, $|S_i|/M$ where $M = \sum_i |S_i|$).
2. Choose at random an assignment $a \in S_i$.
3. Compute the number y_a of clauses satisfied by a.
4. Output $Z = \frac{M}{y_a}$.

Claim 2a: $\mathbb{E}[Z] = |S|$ (i.e., this algorithm's output is unbiased).

Claim 2b: $\sigma(Z) \leq m \cdot \mathbb{E}[Z]$.

The proofs are straightforward and were seen in class.

Exer: Show that $|S|$ can be approximated within factor $1 \pm \varepsilon$ with success probability at least 3/4, by averaging $O(m^2/\varepsilon^2)$ independent repetitions of the above.

Hint: use Chebyshev's inequality.

Exer: Show how to improve the success probability to $1-\delta$ by increasing the number of repetitions by an $O(\log \frac{1}{\delta})$ factor.

Exer: Explain this DNF counting algorithm using the importance sampling theorem.

Hint: Think about the relative contribution of each assignment $\hat{a}$ to $|S|$.