CMSC 858S: Randomized Algorithms

Fall 2001

Handout 5: Pessimistic estimators

Please Note: As usual, the references at the end are given for extra reading if you are interested in exploring these ideas further. You are not required to read these references for the purposes of this course. The material for this handout is basically adapted from [1]. It is meant to supplement and clarify the discussion in class on this topic; we do not reproduce all of the discussion from class, but only some of the more intricate parts of it.

1 The basic method

The method of pessimistic estimators generalizes the method of conditional probabilities, and is used when exact conditional probabilities are hard to compute efficiently. We develop the method in the following setting. Suppose $X_1, X_2, \ldots, X_n$ are *independent* random variables, with X_i taking values in some finite set A_i. (The method works in more general settings, but we choose this one for simplicity and to see the main idea.) Suppose we have a random variable $Y = g(X_1, X_2, \ldots, X_n)$ such that $\mathbf{E}[Y] \leq y$. The issue we address is: what are some sufficient conditions for us to efficiently and *deterministically* find a setting of values for $(X_1, X_2, \ldots, X_n)$ under which we have $Y \leq y$? We start with a basic theorem, and then present a useful corollary which is applicable in many situations.

Theorem 1.1 *Suppose there is a family of functions $\phi_0, \phi_1, \ldots, \phi_n$, where ϕ_i's domain is the set $A_1 \times A_2 \times \cdots \times A_i$, and range is the set of reals; suppose further that the following properties hold. (Recall that A_i is the domain of random variable X_i; also, ϕ_0 is just a constant.)*

(C1) Each function ϕ_i is computable in deterministic polynomial time;

(C2) for all i and for all $(a_1, a_2, \ldots, a_i) \in A_1 \times A_2 \times \cdots \times A_i$, $\mathbf{E}[Y \mid (X_1 = a_1, \ldots, X_i = a_i)] \leq \phi_i(a_1, a_2, \ldots, a_i)$;

(C3) for all $i \leq n-1$ and for all $(a_1, a_2, \ldots, a_i)$, there exists some $a_{i+1} \in A_{i+1}$ such that $\phi_{i+1}(a_1, a_2, \ldots, a_i, a_{i+1}) \leq \phi_i(a_1, a_2, \ldots, a_i)$; and

(C4) $\phi_0 \leq y$.

Then, we can find in deterministic polynomial time, a tuple $(b_1, \ldots, b_n)$ such that $g(b_1, \ldots, b_n) \leq y$ (i.e., such that $Y \leq y$).

Convince yourself that the following deterministic polynomial-time algorithm produces the required tuple $(b_1, \ldots, b_n)$:

For $i = 1, 2, \ldots, n$:
  $b_i =$ any element of A_i such that $\phi_i(b_1, b_2, \ldots, b_i) \leq \phi_{i-1}(b_1, b_2, \ldots, b_{i-1})$.

The main issue now is: how to find such a class of functions ϕ_i? A commonly applicable solution to this problem is given by the following corollary.

Corollary 1.1 *Suppose there exists a random variable Z such that we have the following three properties.*

(P1) For all i and for all $(a_1, a_2, \ldots, a_i) \in A_1 \times A_2 \times \cdots \times A_i$, $\mathbf{E}[Y \mid (X_1 = a_1, \ldots, X_i = a_i)] \leq \mathbf{E}[Z \mid (X_1 = a_1, \ldots, X_i = a_i)]$;

(P2) $\mathbf{E}[Z] \leq y$; and

(P3) suppose we modify the distribution of $(X_1, X_2, \ldots, X_n)$ in any manner such that the X_i are still independent, and where each X_i is sampled from some arbitrary given distribution D_i; then, we can compute $\mathbf{E}[Z]$ in deterministic polynomial time.

Then, we can find in deterministic polynomial time, a tuple $(b_1, \ldots, b_n)$ such that $g(b_1, \ldots, b_n) \leq y$ (i.e., such that $Y \leq y$).

Proof. Define $\phi_i(a_1, a_2, \ldots, a_i) = \mathbf{E}[Z \mid (X_1 = a_1, \ldots, X_i = a_i)]$. We now verify that properties (C1)–(C4) hold. (C2) follows from (P1), and (C4) follows from (P2). Let us verify that (C1) holds; we need to show that for all i and for all $(a_1, a_2, \ldots, a_i)$, $\mathbf{E}[Z \mid (X_1 = a_1, \ldots, X_i = a_i)]$ is efficiently computable. But recall that the X_j are *independent* random variables; suppose each X_j is chosen from some distribution D'_j. Then, conditioning on "$X_1 = a_1, \ldots, X_i = a_i$" is the same as generating all the X_j independently, such that:

- for $j \leq i$, X_j is chosen from the distribution D_j that places all the probability mass on the element a_j; and
- for $j > i$, X_j is chosen from the distribution D'_j.

Thus, hypothesis (P3) shows that (C1) is true.

Finally, let us verify (C3). We have

$$\mathbf{E}[Z \mid (X_1 = a_1, \ldots, X_i = a_i)] = \sum_{a_{i+1} \in A_{i+1}} \Pr[X_{i+1} = a_{i+1}] \cdot \mathbf{E}[Z \mid (X_1 = a_1, \ldots, X_{i+1} = a_{i+1})],$$

since all the X_i are independent. So, since $\sum_{a_{i+1} \in A_{i+1}} \Pr[X_{i+1} = a_{i+1}] = 1$, it is easy to see that

$$\exists a_{i+1} \in A_{i+1} : \ \mathbf{E}[Z \mid (X_1 = a_1, \ldots, X_{i+1} = a_{i+1})] \leq \mathbf{E}[Z \mid (X_1 = a_1, \ldots, X_i = a_i)],$$

which shows that (C3) holds. □

2 An application to low-congestion routing

Recall this problem from our earlier discussion on randomized rounding. Let $[t]$ denote the set $\{1, 2, \ldots, t\}$. We are given a graph $G = (V, E)$ and k pairs of vertices (s_i, t_i). For each $i \in [k]$, we are given a collection

$$\mathcal{P}_i = \{P_{i,1}, P_{i,2}, \ldots, P_{i,\ell_i}\}$$

of (s_i, t_i)–paths. The objective is to choose a path from $\mathcal{P}_i$ for each i, in order to minimize the *congestion*: the maximum, over all edges f, of the total number of chosen paths that pass through f. Also recall that the LP relaxation of the natural integer programming formulation is as follows:

Minimize W subject to:

(i) $\forall i \in [k],\ \sum_{j \in [\ell_i]} z^*_{i,j} = 1$;

(ii) $\forall f \in E,\ \sum_{(i,j): f \in P_{i,j}} z^*_{i,j} \leq W$; and

(iii) $z^*_{i,j} \geq 0$.

Let $\{z^*_{i,j}\}$ denote an optimal solution that we have computed for this LP, with W^* being the corresponding optimal LP value. We wish to round the $z^*_{i,j}$ to some values in $\{0,1\}$ that satisfy the above constraints, without losing much in the objective function. Recall that the randomized rounding algorithm of [2] does the following: independently for each $i \in [k]$, choose exactly one $z_{i,j}$ to be 1, using the $z^*_{i,j}$ as probabilities in the natural way. It is also shown in [2] using the Chernoff-Hoeffding bounds that with probability at least 1/2, we get a solution with objective function value at most $W^*(1+\delta)$, where $\delta = \Delta^+(W^*, 1/(2m))$ and m is the number of edges in G.

We now derandomize this algorithm using Corollary 1.1. To do so, let us first cast our problem in the setting of Section 1. Let:

- $A_i = \mathcal{P}_i$, the given set of possible (s_i, t_i)–paths;
- X_i be the random choice for the (s_i, t_i)–path made by the above randomized rounding algorithm, and $z_{i,j}$ be the indicator random variable for the event "$X_i = P_{i,j}$";
- $\mathrm{cong}(f)$ be the random variable denoting the congestion on edge f; i.e., $\mathrm{cong}(f) = \sum_{(i,j): f \in P_{i,j}} z_{i,j}$;
- Y be the indicator random variable for "$\exists f:\ \mathrm{cong}(f) > W^*(1+\delta)$", and
- $y = 1/2$.

Then, as seen above, the analysis of [2] shows that $\mathbf{E}[Y] \leq 1/2$. If we derandomize this algorithm, we will deterministically produce a value for $(X_1, X_2, \ldots, X_k)$ for which $Y \leq 1/2$, i.e., for which $Y = 0$; this is what we need. To show that pessimistic estimators can be applied fruitfully here, let us recall the analysis of [2]:

$$\begin{aligned}
\mathbf{E}[Y] &\leq \sum_{f \in E} \Pr[\mathrm{cong}(f) > W^*(1+\delta)] \\
&\leq \sum_{f \in E} \mathbf{E}[(1+\delta)^{\mathrm{cong}(f) - W^*(1+\delta)}] \\
&\leq \sum_{f \in E} 1/(2m) \quad \text{(by the definition of } \delta) \\
&= 1/2.
\end{aligned}$$

A little reflection shows that in fact $\forall i\ \forall (a_1, a_2, \ldots, a_i)$,

$$\begin{aligned}
\mathbf{E}[Y \mid (X_1 = a_1, \ldots, X_i = a_i)] &\leq \sum_{f \in E} \mathbf{E}[(1+\delta)^{\mathrm{cong}(f) - W^*(1+\delta)} \mid (X_1 = a_1, \ldots, X_i = a_i)] \\
&= \mathbf{E}[(\sum_{f \in E} (1+\delta)^{\mathrm{cong}(f) - W^*(1+\delta)}) \mid (X_1 = a_1, \ldots, X_i = a_i)].
\end{aligned}$$

Thus, guided by Corollary 1.1, let us define

$$Z = \sum_{f \in E} (1+\delta)^{\mathrm{cong}(f) - W^*(1+\delta)}.$$

We need only prove that property (P3) holds; in turn, we need only show the following. Let $f \in E$ be an arbitrary but fixed edge. Suppose $X_1, X_2, \ldots, X_k$ are *independent* random variables where each X_i is sampled from some arbitrary given distribution D_i. Then, we just need to prove that $\mathbf{E}[(1+\delta)^{\mathrm{cong}(f) - W^*(1+\delta)}]$ can be computed efficiently. To see this, rewrite cong(f) as the sum of k terms, where the ith term is the congestion contributed by X_i:

$$\mathrm{cong}(f) = \sum_{i=1}^{k} \mathrm{cong}_i(f), \text{ where } \mathrm{cong}_i(f) = \sum_{j : f \in P_{i,j}} z_{i,j}.$$

Note that each $\mathrm{cong}_i(f)$ takes on values in $\{0, 1\}$, with

$$\Pr[\mathrm{cong}_i(f) = 1] = \Pr[X_i \in \{P_{i,j} : f \in P_{i,j}\}]$$

being efficiently computable. Also, all the $\mathrm{cong}_i(f)$ are independent, since the X_i are all independent. Thus,

$$\mathbf{E}[(1+\delta)^{\mathrm{cong}(f) - W^*(1+\delta)}] = (1+\delta)^{-W^*(1+\delta)} \cdot \prod_{i=1}^{k} \mathbf{E}[(1+\delta)^{\mathrm{cong}_i(f)}]$$

can be efficiently computed, concluding our argument.

References

[1] P. Raghavan. Probabilistic construction of deterministic algorithms: approximating packing integer programs. *Journal of Computer and System Sciences*, 37:130–143, 1988.

[2] P. Raghavan and C. D. Thompson. Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica*, 7:365–374, 1987.