Randomized approximation algorithms: CS60023

Epsilon nets: Spring 2024: S P Pal

Copyrights reserved

S P Pal

Department of Computer Science and Engineering

Indian Institute of Technology, Kharagpur

The method of random sampling

- Take any *set system* (hypergraph) $G(V, S)$ where $V = \{v_1, \cdots, v_n\}$ and $S = \{e_1, \cdots, e_m\}$; here, $e_i \subseteq V$ for all $1 \leq i \leq m$.

The method of random sampling

- Take any *set system* (hypergraph) $G(V, S)$ where $V = \{v_1, \cdots, v_n\}$ and $S = \{e_1, \cdots, e_m\}$; here, $e_i \subseteq V$ for all $1 \leq i \leq m$.
- Given any integer $1 \leq r \leq n$, we wish to find a subset $N \subseteq V$ that intersects every e_i of size greater than $\frac{n}{r}$.

The method of random sampling

- Take any *set system* (hypergraph) $G(V, S)$ where $V = \{v_1, \cdots, v_n\}$ and $S = \{e_1, \cdots, e_m\}$; here, $e_i \subseteq V$ for all $1 \leq i \leq m$.
- Given any integer $1 \leq r \leq n$, we wish to find a subset $N \subseteq V$ that intersects every e_i of size greater than $\frac{n}{r}$.
- We can assume that $|e_i| > \frac{n}{r}$, for any i, and $m > 1$.

Binomial random sampling of the set of vertices

- Let $p = \frac{cr(\log m)}{n}$, for some large enough constant c.

Binomial random sampling of the set of vertices

- Let $p = \frac{cr(\log m)}{n}$, for some large enough constant c.
- Sample the set V by Binomial distribution, that is, construct the set N by including in it v_i with probability p.

Binomial random sampling of the set of vertices

- Let $p = \frac{cr(\log m)}{n}$, for some large enough constant c.
- Sample the set V by Binomial distribution, that is, construct the set N by including in it v_i with probability p.
- Then, $N \cap e_i = \phi$ with probability less than $(1-p)^{\frac{n}{r}}$ for a single value of i.

Binomial random sampling of the set of vertices

- Let $p = \frac{cr(\log m)}{n}$, for some large enough constant c.
- Sample the set V by Binomial distribution, that is, construct the set N by including in it v_i with probability p.
- Then, $N \cap e_i = \phi$ with probability less than $(1-p)^{\frac{n}{r}}$ for a single value of i.
- The probability that N does not intersect some e_i is less than $m(1-p)^{\frac{n}{r}}$.

Binomial random sampling of the set of vertices

- We can make this probability smaller than any constant. Why?

Binomial random sampling of the set of vertices

- We can make this probability smaller than any constant. Why?
- The probability can be shown to be bounded by $\frac{1}{m^{c-1}}$, which can be made smaller than any given limiting constant as m grows beyond a certain value for each valid choice of the constant c.

Binomial random sampling of the set of vertices

- We can make this probability smaller than any constant. Why?
- The probability can be shown to be bounded by $\frac{1}{m^{c-1}}$, which can be made smaller than any given limiting constant as m grows beyond a certain value for each valid choice of the constant c.
- So, the sample N of expected size $np = cr \log m$ intersects every set e_i of size $\frac{n}{r}$; note also that the random sample N is of size $O(r \log m)$ with high probability.

Binomial random sampling of the set of vertices

- We can make this probability smaller than any constant. Why?
- The probability can be shown to be bounded by $\frac{1}{m^{c-1}}$, which can be made smaller than any given limiting constant as m grows beyond a certain value for each valid choice of the constant c.
- So, the sample N of expected size $np = cr \log m$ intersects every set e_i of size $\frac{n}{r}$; note also that the random sample N is of size $O(r \log m)$ with high probability.
- Therefore, we have a way of getting random samples such as N of size $O(r \log m)$ with high probability, so that N intersects all the sets e_i of size greater than $\frac{n}{r}$.

Binomial random sampling of the set of vertices

- This was a simple randomized construction of such a set N (with high probability), that intersects all the sets of size not lesser than $\frac{n}{r}$.

Binomial random sampling of the set of vertices

- This was a simple randomized construction of such a set N (with high probability), that intersects all the sets of size not lesser than $\frac{n}{r}$.
- Can we construct such a set N by a deterministic method?

Binomial random sampling of the set of vertices

- This was a simple randomized construction of such a set N (with high probability), that intersects all the sets of size not lesser than $\frac{n}{r}$.
- Can we construct such a set N by a deterministic method?
- We discuss a greedy deterministic method, and later we consider another method, which is a derandomization of the above randomized sampling technique.

Binomial random sampling of the set of vertices

- This was a simple randomized construction of such a set N (with high probability), that intersects all the sets of size not lesser than $\frac{n}{r}$.
- Can we construct such a set N by a deterministic method?
- We discuss a greedy deterministic method, and later we consider another method, which is a derandomization of the above randomized sampling technique.

Deterministic construction of ϵ-nets

- We state a greedy and deterministic way of generating an ϵ-net N as follows.

Deterministic construction of ϵ-nets

- We state a greedy and deterministic way of generating an ϵ-net N as follows.
- Find a vertex v_i contained in most sets $e_j \subseteq S$.

Deterministic construction of ϵ-nets

- We state a greedy and deterministic way of generating an ϵ-net N as follows.
- Find a vertex v_i contained in most sets $e_j \subseteq S$.
- Remove this vertex from further consideration and add it to N.

Deterministic construction of ϵ-nets

- We state a greedy and deterministic way of generating an ϵ-net N as follows.
- Find a vertex v_i contained in most sets $e_j \subseteq S$.
- Remove this vertex from further consideration and add it to N.
- Then, remove all sets containing v_i from future consideration.

Deterministic construction of ϵ-nets

- We state a greedy and deterministic way of generating an ϵ-net N as follows.
- Find a vertex v_i contained in most sets $e_j \subseteq S$.
- Remove this vertex from further consideration and add it to N.
- Then, remove all sets containing v_i from future consideration.
- Repeat these steps until all hyperedges from S are removed.

Deterministic construction of ϵ-nets

- We state a greedy and deterministic way of generating an ϵ-net N as follows.
- Find a vertex v_i contained in most sets $e_j \subseteq S$.
- Remove this vertex from further consideration and add it to N.
- Then, remove all sets containing v_i from future consideration.
- Repeat these steps until all hyperedges from S are removed.
- Show that this algorithm can be made to run in $O(mn)$ time.

Deterministic construction of ϵ-nets

- Does N turn out to be quite small as required?

Deterministic construction of ϵ-nets

- Does N turn out to be quite small as required?
- Let m_k be the number of hyperedges remaining after k iterations. Clearly, $m_0 = m$.

Deterministic construction of ϵ-nets

- Does N turn out to be quite small as required?
- Let m_k be the number of hyperedges remaining after k iterations. Clearly, $m_0 = m$.
- Sticking to the assumption that each set e_i has at least $\frac{n}{r}$ elements, we now have m_k sets left after the kth iteration with at least $\frac{n}{r}$ elements.

Deterministic construction of ϵ-nets

- Does N turn out to be quite small as required?
- Let m_k be the number of hyperedges remaining after k iterations. Clearly, $m_0 = m$.
- Sticking to the assumption that each set e_i has at least $\frac{n}{r}$ elements, we now have m_k sets left after the kth iteration with at least $\frac{n}{r}$ elements.
- We can select any of the $n - k$ remaining vertices in the next iteration.

Deterministic construction of ϵ-nets

- Does N turn out to be quite small as required?
- Let m_k be the number of hyperedges remaining after k iterations. Clearly, $m_0 = m$.
- Sticking to the assumption that each set e_i has at least $\frac{n}{r}$ elements, we now have m_k sets left after the kth iteration with at least $\frac{n}{r}$ elements.
- We can select any of the $n - k$ remaining vertices in the next iteration.
- We have a distribution of $n - k$ distinct vertices in at least $m_k \times \frac{n}{r}$ instances over the m_k sets.

Deterministic construction of ϵ-nets

- Does N turn out to be quite small as required?
- Let m_k be the number of hyperedges remaining after k iterations. Clearly, $m_0 = m$.
- Sticking to the assumption that each set e_i has at least $\frac{n}{r}$ elements, we now have m_k sets left after the kth iteration with at least $\frac{n}{r}$ elements.
- We can select any of the $n - k$ remaining vertices in the next iteration.
- We have a distribution of $n - k$ distinct vertices in at least $m_k \times \frac{n}{r}$ instances over the m_k sets.
- So, the most frequent vertex of the m_k sets must be in at least $\frac{m_k \times \frac{n}{r}}{n-k} \geq \frac{m_k}{r}$ sets. Why?

Deterministic construction of ϵ-nets

- Hint: What is the expected number of vertices in these sets?

Deterministic construction of ϵ-nets

- Hint: What is the expected number of vertices in these sets?
- Thus, $m_{k+1} \leq m_k(1 - \frac{1}{r})$.

Deterministic construction of ϵ-nets

- Hint: What is the expected number of vertices in these sets?
- Thus, $m_{k+1} \leq m_k(1 - \frac{1}{r})$.
- So, $m_k \leq m(1 - \frac{1}{r})^k$.

Deterministic construction of ϵ-nets

- Hint: What is the expected number of vertices in these sets?
- Thus, $m_{k+1} \leq m_k(1 - \frac{1}{r})$.
- So, $m_k \leq m(1 - \frac{1}{r})^k$.
- For a large enough constant $c > 0$, and any $k \geq cr \log m$, we have $m_k < 1$, and therefore $m_k = 0$.

Deterministic construction of ϵ-nets

- Hint: What is the expected number of vertices in these sets?
- Thus, $m_{k+1} \leq m_k(1 - \frac{1}{r})$.
- So, $m_k \leq m(1 - \frac{1}{r})^k$.
- For a large enough constant $c > 0$, and any $k \geq cr \log m$, we have $m_k < 1$, and therefore $m_k = 0$.
- In other words, picking any sufficiently large number $k \geq cr \log m$ of vertices we can ensure that we hit all the hyperedges that have at least $\frac{n}{r}$ vertices.

Deterministic construction of ϵ-nets

- Hint: What is the expected number of vertices in these sets?
- Thus, $m_{k+1} \leq m_k(1 - \frac{1}{r})$.
- So, $m_k \leq m(1 - \frac{1}{r})^k$.
- For a large enough constant $c > 0$, and any $k \geq cr \log m$, we have $m_k < 1$, and therefore $m_k = 0$.
- In other words, picking any sufficiently large number $k \geq cr \log m$ of vertices we can ensure that we hit all the hyperedges that have at least $\frac{n}{r}$ vertices.
- So, we can deterministically compute a $\frac{1}{r}$-net N of size $O(r \log m)$, greedily as above.