Set cover problem: definition

Given: A finite set (*universe*) U of n elements, a collection of its subsets $S_1, \ldots, S_k$. Each set S_i has a nonnegative cost $cost(S)$.
Find: A minimum-cost collection of these subsets whose union is U.

Greedy algorithm

Idea: cover most elements per unit cost in each step.

1. $C \leftarrow \emptyset$
2. While $C \neq U$:
3. Let S be the most cost-effective set in the current iteration
4. Let $\alpha = \frac{cost(S)}{|S \cap \overline{C}|}$, the cost-effectiveness of S.
5. Pick S.
6. For each $e \in S \setminus C$, set price$(e) = \alpha$.
7. Output the picked sets.

Set cover greedy: analysis

Suppose the elements are picked in order $e_1, e_2, \ldots, e_n$.

Lemma

For every $k = 1, \ldots, n$, price$(e_i) \leq \frac{OPT}{n-k+1}$.

Proof.

At any time during greedy, compare what is left to be covered and the sets of the optimal solution that have not been used. These sets together cost at most OPT and cover at least $|\overline{C}|$. Therefore, the most cost-effective set has $\alpha \leq OPT/|\overline{C}|$.

In iteration that covered e_k, $|\overline{C}| \geq n - k + 1$. Thus

$$\text{price}(e_k) \leq \frac{OPT}{|\overline{C}|} \leq \frac{OPT}{n-k+1}$$

□

Greedy set cover analysis (2)

Theorem

Greedy is an H_n approximation algorithm for set cover ($H_n = 1 + 1/2 + \cdots + 1/n$).

Proof.

The cost of each chosen set is the sum of its elements' prices. The total cost of the set cover is

$$\sum_{k=1}^{n} \text{price}(e_k) \leq OPT(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}).$$

□

Is the analysis tight?

There's an example in the book....

Can we do better?

Actually, NO!

It has been shown that for any $c > 0$, if there is an approximation algorithm for set cover with guarantee at most $(1 - c) \ln n$, then (well, not actually P=NP, but something just as unlikely: every problem in NP could then be solved in time $O(n^{\text{polylog}(n)})$, which is not thought to be possible...)

A completely different approach: linear programming

How to formulate set cover as an ILP?
Solution: some sets.
So, variables: tell us which sets have been selected.
$x_S = 1$ iff S is part of the solution, $x_S = 0$ otherwise.

Cost: $\sum_{S \in \mathcal{S}} c(S) x_S$.
Constraints say we have to cover every element e: $\sum_{S\colon e \in S} x_S \geq 1$.

Set cover (I)LP

$$\min \sum_{s \in \mathcal{S}} c(S) x_S$$

s.t.

$$\sum_{S\colon e \in S} x_S \geq 1, \ \forall e \in U$$

$$x_S \in \{0, 1\} \ \forall S \in \mathcal{S}$$

To get a polynomially solvable relaxation, replace the 0-1 constraints by $0 \leq x_S \leq 1$ for all S.

Set cover rounding

How to round? Deterministically, it's not easy to get a good bound. However, randomized rounding here is very simple.

Set cover randomized rounding procedure:
Suppose x_S^* for every S are the values of x_S in the optimal LP solution.
For each S independently, flip a coin biased by x_S^*, so that S is included in the cover with probability exactly x_S^*.

Write $\overline{x}_S$ for the rounded value of x_S.
Easy: the expected cost of this rounding procedure is

$$\sum_S c(S)\,\mathrm{P}[\overline{x}_S = 1] = \sum_S c(S) x_S^* = OPT_{LP} \leq OPT.$$

How about the success probability?

Set cover rounding: success probability

For any $e \in U$, let k_e be the number of sets containing e. Then,

$$\begin{aligned} \mathrm{P}[e \text{ is covered}] &= 1 - \prod_{S\colon e \in S} (1 - x_S^*) \\ &\geq 1 - \Big(\frac{k_e - (\sum_{S\colon e\in S} x_S^*)}{k_e}\Big)^{1/k_e} \\ &\geq 1 - (1 - \frac{1}{k_e})^{1/k_e} \geq 1 - \frac{1}{e}. \end{aligned}$$

(First line to second line: geometric-arithmetic inequality)
(Second line to third: $\sum_{S\colon e\in S} x_S \geq 1$.)

Set cover rounding: success probability (2)

From previous slide: For every e, the probability e is covered is at least $1/2$.

Repeat the rounding (independently) $2 \lg n$ times, take the union of all (partial) covers found. Then

1. The expected total cost is at most $2 \lg n OPT$
2. The probability e is not covered is at most $(1/2)^{2 \lg n} < 1/(2n)$. The probability there is an uncovered element is (by union bound) at most $1/2$.

Is this tight?