Advanced Mathematics Topics in Computer Science

LECTURE 2: PROBABILISTIC ANALYSIS AND RANDOMIZED ALGORITHMS

Roadmap

- Sample Space and Events
  - Properties and Propositions
- Probabilistic Analysis
  - The hiring problem

Sample Space

- Definition: The sample space S of an experiment (whose outcome is uncertain) is the set of all possible outcomes of the experiment.
- Example (child): Determining the sex of a newborn child in which case
  - S = {boy, girl}.
- Example (horse race): Assume you have an horse race with 12 horses. If the experiment is the order of finish in a race, then
  - S = {all 12! permutations of (1, 2, 3, ..., 11, 12)}

Events

- Any subset E of the sample space S is known as an event; i.e. an event is a set consisting of possible outcomes of the experiment.
- If the outcome of the experiment is in E, then we say that E has occurred.
- Example (child): The event E = {boy} is the event that the child is a boy.
- Example (horse race): The event E = {all outcomes in S starting with a 7} is the event that the race was won by horse 7.

Axioms of Probability

- Consider an experiment with sample space S. For each event E, we assume that a number P (E), the probability of the event E, is denied and satisfies the following 3 axioms.

- Axiom 1
  - 0 <= P (E) <= 1

- Axiom 2
  - P (S) = 1

- Axiom 3. For any sequence of mutually exclusive events $\{E_i\}_{i>=1}$, i.e. E_i intersects E_j = Ø when i ≠ j, then
  - P (Union of E_i) = Sum of P(E_i)

Properties

- Proposition: $P(E^c) = 1 - P(E)$.
- Proposition: If $E \subset F$ then $P(E) \leq P(F)$.
- Proposition: We have $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Example: Matching Problem

- You have n letters and n envelopes and randomly stu¤ the letters in the envelopes. What is the probability that at least one letter will match its intended envelope?

- The sample space is the space of permutations of {1, 2, …, n} and thus has n! outcomes.

- Let Ei =“letter i matches its intended envelop”. We are interested in P (E1 ∪ E2 ∪ …∪ En).

- Consider the event $E_{i1} \cap \ldots \cap E_{ir}$ the event that each of the r letters $i_1, \ldots, i_r$ match their intended envelopes. There are (n - r) (n - r - 1) … 1 such outcomes corresponding to the number of ways the remaining r envelopes can be matched. Assuming all outcomes equi-probable, we have

- $P(Ei1 \cap \ldots \cap Eir) = (n-r)! / n!$

Matching problem (cont.)

- Apply the formula in Proposition 3
- Each term is equal to $-1^{(r+1)}$ x (n choose r) x (n-r)!/n! = 1/r!
- Final probability = $\sum_{r=1}^{n} (-1)^{r+1} \frac{1}{r!}$ = 1 – e^{-1} when n → ∞

Example: Three children with same birthday

- A recent news story in the Vietnam featured a family whose three children had all been born on the same day. But is this so remarkable?

- The sample space is S = ((i , j, k) ; i in {1, ..., 365} , j in {1, ..., 365} , j in {1, ..., 365}) so assuming each day is equally likely, the probability the three days coincides is
  - 1 / 365 x 365 ~= 7.5 / 1, 000, 000.

- This is quite small but much higher that winning at the lottery.

- There are 24,000,000 households in Vietnam, and 1,000,000 of them are made up of a couple and 3 or more dependent children. Therefore we would expect around 7 or 8 families in Vietnam to have three children all born on the same day, and so this family is unlikely to be unique in this country.

The hiring problem

```
HIRE-ASSISTANT(n)
best←0
candidate 0 is a least-qualified dummy candidate
for i←1 to n
    do interview candidate i
        if candidate i is better than candidate best
          then best←i
               hire candidate i
```

Cost Analysis

- We are not concerned with the running time of HIRE-ASSISTANT, but instead with the cost incurred by interviewing and hiring.

- Interviewing has low cost, say c_i, whereas hiring is expensive, costing c_h. Let m be the number of people hired. Then the cost associated with this algorithm is $O(nc_i + mc_h)$. No matter how many people we hire, we always interview n candidates and thus always incur the cost nc_i, associated with interviewing.

Worst-case analysis

- In the worst case, we actually hire every candidate that we interview. This situation occurs if the candidates come in increasing order of quality, in which case we hire n times, for a total hiring cost of $O(nc_h)$.

Probabilistic analysis

- *Probabilistic analysis* is the use of probability in the analysis of problems. In order to perform a probabilistic analysis, we must use knowledge of the distribution of the inputs.

- For the hiring problem, we can assume that the applicants come in a random order.

Randomized algorithm

- We call an algorithm *randomized* if its behavior is determined not only by its input but also by values produced by a *random-number generator*.