Divide and Conquer Strategy

- Split the problem into several **smaller independent** subproblems.
- Recursively solve the subproblems.
- Combine the solutions of the subproblems to get the solution for the original problem.

Quicksort

Procedure Q_S(S);

Input: A set S.

Output: The set S in sorted order.

1. If $|S| \leq 1$ then return S, else

2. (a) Let y be the first element in S.
   (b) Compare all elements of S to y. Let

$$S_1 = \{x \in S - \{y\} \mid x \leq y\}$$

$$S_2 = \{x \in S - \{y\} \mid x > y\}.$$

   (Elements in S_1 and S_2 are in the same order as in S.)
   (c) Return the list:

$$Q\_S(S_1), y, Q\_S(S_2).$$

Time Analysis

Let $T(n)$ be the running time of Q_S on a set S of size n.

Theorem 1. $T(n) = O(n^2)$.

Theorem 2. $T(n) = \Omega(n \log n)$.

How does Quicksort perform on the "average" ?

Can we make Quicksort run fast on *all* inputs ?

Randomized Quicksort

Procedure Q_S(S); **Input:** A set S.

Output: The set S in sorted order.

1. If $|S| \leq 1$ then return S, else

2.(a) Choose a random element y uniformly from S.

(b) Compare all elements of S to y. Let

$$S_1 = \{x \in S - \{y\} \mid x \leq y\}$$

$$S_2 = \{x \in S - \{y\} \mid x > y\}.$$

(Elements in S_1 and S_2 are in the same order as in S.)

(c) Return the list:

$$Q\_S(S_1), y, Q\_S(S_2).$$

Probability and Algorithms

Randomized algorithms - algorithm that perform random steps.

Probabilistic analysis of algorithms - the performance of an algorithm on a randomly generated input.

Randomized algorithms are typically faster and simpler to implement than their deterministic counterparts.

Probabilistic algorithms are useful in analyzing the performance of an algorithm on a "typical" input; also called as *average case analysis*.

Randomized Algorithm

We augment the standard RAM computation model with a new operation:

Choose a random number uniformly from the set $\{a_1, a_2, ..., a_k\}$.

We assume that this operation takes 1 step.

Finding the Repeated element

Consider an array of n numbers (assume n is even) that has $n/2$ distinct elements and $n/2$ copies of another element. We want an algorithm to identify the repeated element.

Any deterministic algorithm will need at least $n/2+2$ steps in the worst case. Consider an adversary who has full knowledge of the algorithm. The adversary can make sure that the first $n/2+1$ elements examined by the algorithm are all distinct.

Consider the following randomized algorithm:

1. Randomly pick two array elements and check whether they come from different cells and have the same value.

2. If they do, output **true** else output **false**.

What is the probability that the algorithm will find the repeated element?

Events and Probability

Consider an experiment with a finite (or countably infinite) number of outcomes.

Each outcome is a simple event (or a sample point).

The **sample space** is the set of all possible simple (elementary) events.

An **event** $\mathcal{E}$ is a union of simple events - a subset of the sample space.

Two events are **mutually exclusive** if $A \cap B = 0$.

Probability Space

A probability distribution Pr on a sample space S is a mapping from events of S to real numbers such that

- $Pr(A) \geq 0$ for any event A.
- $Pr\{S\} = 1$.
- For any (finite or countably infinite) sequence of pairwise mutually exclusive events ,$A_1, A_2, \ldots$:

$$\Pr\{\cup_i A_i\} = \sum_i \Pr\{A_i\}$$

The pair (S, Pr) is called a discrete **probability space**.

$$Pr(\mathcal{E}) \;=\; \sum_{s \in \mathcal{E}} Pr(s).$$

Examples:

Consider the random process defined by the outcome of rolling a dice.

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$$

We assume that all "facets" have equal probability, thus

$$Pr(1) = Pr(2) = ...Pr(6) = 1/6.$$

The probability of the event "odd outcome"

$$= Pr(\{1, 3, 5\}) = 1/2$$

An Infinite Sample Space

Flip an unbiased coin until HEADS appears for the first time. Here the sample space is

$$\{H, TH, TTH, TTTH, \ldots\}.$$

The event that "the number of TAILS seen is odd" is give by the infinite set

$$\{TH, TTTH, TTTTTH, \ldots\}.$$

Assume that we roll two dice:

$\mathcal{S} =$ all ordered pairs $\{(i,j),\ 1 \le i,j \le 6\}$.

We assume that each (ordered) combination has probability $1/36$.

Probability of the event "sum $= 2$" $=$

$$Pr((1,1)) = 1/36.$$

Probability of the event "sum $= 3$"

$$Pr(\{(1,2),(2,1)\}) = 2/36.$$

Let $E_1 =$ "sum bounded by 6",

$$E_1 = \{(1,1),(1,2),(1,3),(1,4),(1,5),(2,1),(2,2),$$

$$(2,3),(2,4),(3,1),(3,2),(3,3),(4,1),(4,2),(5,1)\}$$
$$Pr(E_1) = 15/36$$

Let $E_2 =$ "both dice have odd numbers", $Pr(E_2) = 1/4$.

$$Pr(E_1 \cap E_2) =$$
$$Pr(\{(1,1),(1,3),(1,5),(3,1),(3,3),(5,1)\}) =$$
$$6/36 = 1/6.$$

Back to Finding the Repeated Element

There are n^2 possible combinations of picking the two elements.

Out of these, $\frac{n}{2}(\frac{n}{2} - 1)$ combinations will result in success.

Since every combination is equally likely, the probability that the algorithm will succeed is $\frac{n/2(n/2-1)}{n^2} = \frac{1}{2}(\frac{1}{2} - \frac{1}{n}) = \frac{1}{4} - \frac{1}{2n} \geq \frac{1}{5}$ for $n \geq 10$.

Thus the probability that the algorithm fails is $< 4/5$.

How to reduce the failure probability?

Conditional Probability

What is the probability that a random person born in Indiana is a student at Purdue.

$E_1 =$ the event that a random person in the world is born in Indiana.

$E_2 =$ the event that a random person in the world is a student at Purdue.

The conditional probability that a random person born in Indiana is a student at Purdue is denoted

$$Pr(E_2 \mid E_1).$$

Computing Conditional Probabilities

$$Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)}$$

By conditioning on B we restrict the sample space to the set B.

Thus we are interested in $Pr(A \cap B)$ "normalized" by $Pr(B)$.

Example

What is the probability that in rolling two dice the sum is 8 given that the sum was even?

$E_1 =$ "sum is 8",

$Pr(E_1) = Pr((2,6),(3,5),(4,4),(5,3),(6,2)) = 5/36$

$E_2 =$ "sum even",

$Pr(E_2) = 1/2 = 18/36.$

$Pr(E_1 \mid E_2) = \frac{Pr(E_1 \cap E_2)}{Pr(E_2)} = \frac{5/36}{1/2} = 5/18.$

Independence

Two events A and B are independent if

$$Pr(A \cap B) = Pr(A) \times Pr(B),$$

or (when $Pr(B) > 0$)

$$Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} = Pr(A).$$