Probability and Algorithms

Randomized algorithms - algorithm that perform random steps.

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

Applications: Communication networks, Cryptography, Search engines, Fast data structures, Scheduling, Optimization algorithms, AI reasoning, Learning, Quantum computing...

Verifying Polynomial Identities

Problem: Verify $P(x) \equiv Q(x)$.

Example: Check if

$$(x+11)(x-2)(x+3)(x-4)(x+5)(x-6) \equiv x^6-7x^3+25$$

A Randomized Solution

Choose a random integer r between $[0, 600]$.

Compute $H(r)$ and $G(r)$.

If $H(r) = G(r)$ output CORRECT else output FALSE.

Example: Assume $r = 2$: then $H(2) \neq G(2)$. Thus $H(x) \not\equiv G(x)$.

A Bad Scenario

Check if $x^2 + 7x + 1 \equiv (x + 2)^2$.

If $r = 1$ is chosen then, LHS $=$ RHS although the polynomials are not equivalent.

A little algebra

Assume that $G(x) \not\equiv H(x)$, and that the sum of the degrees of x in H and G is bounded by d.

$F(x) \equiv G(x) - H(x)$ is a polynomial of degree bounded by d.

Theorem 1. *The equation $F(x) = 0$ has no more than d roots.*

Analysis of the Algorithm

If the identity is correct, the algorithm always returns the correct answer.

If the identity is NOT correct, the algorithm outputs a WRONG answer only if we pick r which is a root of the polynomial $F(x) = G(x) - H(x) = 0$.

Events and Probability

Consider an experiment with finite or infinite number of outcomes.

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

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

An **event** E is a union of simple events - a subset of the sample space.

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

The probability of an event is the sum of the probabilities of its atomic events.

Probability Space

A σ-field (Ω, F) consists of a sample space Ω and a collection of subsets F satisfying the following conditions.

- $\phi \in F$.
- $E \in F \Rightarrow \bar{E} \in F$.
- $E_1, E_2, \ldots \in F \Rightarrow E_1 \cup E_2 \cup \ldots \in F$.

Given a σ-field (Ω, F), a probability measure $\mathrm{Pr} : F \rightarrow R^+$ is a function that satisfies the following conditions:

- For all $A \in F$, $0 \leq Pr(A) \leq 1$.
- $Pr\{\Omega\} = 1$.

- For any sequence of mutually exclusive events $A_1, A_2, \ldots$:

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

A **Probability Space** $(\Omega, F, \Pr)$ consists of a σ-field (Ω, F) with a probability measure $\Pr$ defined on it.

Conditional Probability and Independence

$$\Pr\{A|B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

Two events are **independent** if

$$\Pr\{A \cap B\} = \Pr\{A\} \Pr\{B\}$$

or equivalently (if $\Pr\{B\} \neq 0$)

$$\Pr\{A|B\} = \Pr\{A\}$$

Independent events need not be related to independent physical processes.

The Birthday Paradox

What is the probability that among m people no two have the same birthday?

Assumptions:

1. All birthdays are equally likely.

2. Birthdays are independent events.

The sample space is the set of all vectors $S = \{(b_1, \ldots, b_m) | b_i \in [1, \ldots, N]\}$.

We need to compute $Pr(E)$ where $E = \{(b_1, \ldots, b_m | b_i \neq b_j$ for all $i \neq j\}$.

How many different atomic events are counted in E?

The number of possible m different birthdays is $N.(N-1).(N-2)\ldots(N-m+1)$

$$Pr(E) = \frac{N.(N-1).(N-2)...(N-m+1)}{N^m}$$

$$= \Pi_{i=0}^{m-1}(1-i/N)$$

$$\leq \Pi_{i=0}^{m-1} e^{-i/N} = e^{-\sum_{i=0}^{m-1} i/N} = e^{-m(m-1)/2N}$$

For $m = \sqrt{2N} + 1 \leq 28$,

$\Pr(E) < 1/e < 1/2.$

Alternate Analysis

Assume that we choose one birthday after the other independently and uniformly at random from $[1 \ldots N]$.

Let E_i: "the ith choice is different from the first $i-1$ choices".

$\Pr(\cap_{i=1}^{m} E_i) = \Pr(E_1)\Pr(E_2|E_1)\Pr(E_3|E_2 \cap E_1)\ldots\Pr(E_m|\cap_{i=1}^{m-1} E_i) = \Pi_{i=1}^{m}(1-\frac{i-1}{N})$