CS 590R
Lecture Notes: Jan 16, 2003
Probability and Algorithms

Lecturer: Gopal Pandurangan
Notes taken by: Maleq Khan

Probability and Algorithms

Randomized algorithms: Algorithms that perform random steps and does not assume any specific property of input, i.e., input is considered arbitrary, are called randomized algorithms. Randomized algorithms tend to be simpler and more efficient (which is desired), comparing with the deterministic algorithms, and produce comparatively good results.

Probabilistic algorithms: On the other hand, probabilistic algorithms assume some properties of input. Input is chosen according to some probabilistic distribution.

Philosophically, there is no deterministic algorithm - at every step of an algorithm, there is a chance that the machine may crash.

Probabilistic analysis: Probabilistic analysis of the algorithms performs average case analysis considering a randomly generated input. Probabilistic analysis can be applied on both the randomized algorithms and deterministic algorithms in various field of computer science such as communication networks, cryptography, search engines, fast data structures, scheduling, optimizing algorithms, AI reasoning, learning, quantum computing.

A randomized algorithm to verify the polynomial idnetities and its probabilistic analysis is given below.

A Randomized Algorithm to Verify Polynomial Identities

Verify that $P(x) \equiv Q(x)$ where $P(x)$ and $Q(x)$ are two polynomials.
Example: Check if $(x+11)(x-2)(x+3)(x-4)(x+5)(x-6) \equiv x^6 - 7x^3 + 25$

Solution 1: Expand the expression, bring it in to the form $\sum c_i x_i$, and compare the co-efficients.

Solution 2: A randomized solution. Let the polynimial identity is $H(x) \equiv G(x)$. To verify its correctness perform the following steps.

a) Choose a random integer r between [0, 600].
b) Compute $H(r)$ and $G(r)$
c) If $H(r) = G(r)$ output TRUE, else output FALSE

Example: Assume $r = 2$. Then $H(2) \neq G(2)$, thus $H(x) \not\equiv G(x)$

The algorithm above may not produce a correct answer. There is a chance of giving output TRUE, even if $H(x) \not\equiv G(x)$. It happens when the chosen r is a root of the equation $H(x) - G(x) = 0$. For example, consider, we want to check if $x^2 + 7x + 1 \equiv (x + 2)^2$. If $r = 1$ is chosen, LHS = RHS although the polynomials are not equivalent.

What is the probability that the above algorithm (solution 2) produces a correct answer?

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 equation $F(x) = G(x) - H(x) = 0$. In other words, if the algorithm outputs 'FALSE', it is always a correct answer but when the algorithm outputs 'TRUE', it might be a wrong answer. This is called one-sided error. Such a algorithm is called one-side error Monte-Carlo algorithm.

Assume that $G(x) \not\equiv H(x)$, and that the degrees of the polynomials $H(x)$ and $G(x)$ is bounded by d. Then $F(x) \equiv G(x) - H(x)$ is also a polynomial of degree bounded by d. The equation $F(x) = 0$ has no more than d roots. Assume that r is chosen from the interval $[L, U]$, i.e. $r \in [L, U]$. The length of the interval, $l = U - L + 1$.

Now, $\Pr\{\text{wrong answer}\} = \Pr\{r \text{ is a root of the eq. } F(x) = 0\} \leq \frac{d}{l}$.

The probability of getting wrong answer can be reduced by repeating the algorithm. If in any iteration, $H(r) \neq G(r)$ output 'FALSE', otherwise output 'TRUE'. If we repeat the algorithm k times, the probability of getting a wrong answer is $(\frac{d}{l})^k$. For example, if we choose $r \in [0, 100d]$ and repeat 3 times, $\Pr\{\text{wrong answer}\} \leq (d/100d)^3 = 10^{-6}$.

Events and Probability

Consider an experiment with finite or infinite number of outcomes. Each outcome is a *simple event* (or a *sample point*). Remember the previous example (verifying polynomial identities), where we choose $r \in [0, 100d] = [0, 600]$ (considering $d = 6$). Every integer in the interval [0,600] is a sample point.

The *sample space* Ω is the set of all possible simple (elementary) events. In the previous example Ω = [0,600]. Here [0,600] is a discrete sample space. If r is allowed to take any real value from [0,600], then [0,600] is a continuous sample space. The sample space for tossing a fair die is the set of its six faces, that is, Ω = {1,2,3,4,5,6}. Note that Pr{Ω} = 1

If the size (number of all possible simple events) of a sample space is finite (infinite), it is called a *finite (infinite) sample space*. In the above example, the discrete sample space, [0, 600], is a finite sample space and the continuous sample space is an infinite sample space. A discrete sample space can also be infinite. Consider tossing a fair coin until we get a 'tail', where the sample space is $\{T, HT, HHT, \ldots\}$, which is infinite (here H and T represents head and tail of a coin respectively).

An *event* E is a set of some simple events. Every subset of the sample space is an event. Consider tossing a fair die. {1,2,3,4}, {1,3,5}, and {2,4,6} are three events.

Two events are *mutually exclusive* if $A \cap B = 0$. That is, if A occurs, B never occur and vice versa. {1,2,3} and {2,4,6} are two mutually exclusive events.

The probability of an event is the sum of the probabilities of its atomic (simple events) events. For example, Pr{2,4,6} = Pr{2} + Pr{4} + Pr{6}.

Probability Space

When sample space is infinite, it is not feasible to consider all subsets of the sample space. Instead, we consider a collection F of subsets of sample space Ω. Another difficulty of dealing with events in a infinite sample space is that the probability of occurring a particular event tends to be 0. The following formulation helps to deal with the infinite sample spaces as well as the finite sample spaces.

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

- $\phi \in F$.
- $E \in F \Rightarrow \bar{E} \in F$, i.e., F is closed under complement.

- $E_1, E_2, \cdots \in F \Rightarrow E_1 \cup E_2 \cup \cdots \in F$, i.e., F is closed under union.

Given a *σ-field* (Ω, F), a *probability measure* $\Pr : F \to R^+$ 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 $A_1, A_2, \ldots$:

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

  This is known as countable additivity property.

For a finite discrete sample space, usually, F is a collection of all possible subsets of Ω. There are $2^{|\Omega|}$ such subsets.

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

Example 1: Consider rolling of a fair die. $\Omega = \{1, 2, 3, 4, 5, 6\}$; an event E_1 = "outcome is even" = {2,4,6}; $\Pr\{E_1\} = \frac{3}{6}$.

Example 2: Consider rolling of two fair dice together. Sample space $\Omega = \{1, 2, \ldots, 6\} \times \{1, 2, \ldots, 6\} = \{(1,1), (1,2), \ldots, (2,1), (2,2), \ldots, (6,6)\}$; an event E_2 = "sum is less than 4" = {(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)}; $\Pr\{E_2\} = \frac{6}{36}$.

Conditional Probability and Independence of Events

Let A and B be any two events in some sample space. Probability of occurring A given that B occurred is given by

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

Two events are *independent* if

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

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

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

In other words, two events A and B are independent if the probability of occurring A when it is given that B occurred is same as the probability of occurring A when it is given that B did NOT occur. Independent events need not be related to independent physical processes.

Consider tossing a fair die. Two events E_1 = "outcome is ≤ 4" and E_2 = "outcome is even" are independent.

$\Pr\{E_1\} = \Pr\{1,2,3,4\} = \frac{4}{6}$.

$\Pr\{E_2\} = \Pr\{2,4,6\} = \frac{3}{6}$.

$\Pr\{E_2 \cap E_2\} = \Pr\{2,4\} = \frac{2}{6} = \Pr\{E_1\} \times \Pr\{E_2\}$.

The events E_1 = "outcome is ≤ 3" and E_2 = "outcome is even" are NOT independent. Another example of independent events are the successive tossing of a coin.

This theory of independent event is a powerful tool which makes theoretical analysis of many systems possible and/or easier.

The Birthday Paradox

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

The following two assumptions are made to find the answer to the above question.

1. All birthdays are equally likely.
2. Birthdays are independent events.

Let, there are N possible birthdays, i.e., number of days in a year. Then the size of the sample space $|S| = N^m$, where m is the number of people. 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 $E = \{(b_1, \ldots, b_m|b_i \neq b_j \text{ for all } i \neq j\}$.

How many different atomic events are counted in E?

Let us order the people as first person, second person, third person, etc. The birthday for the first person can be chosen in N ways. To have a different birthday, the birthday for second person can be chosen in $(N-1)$ ways, and so on. Thus the number of possible m different birthdays is $N.(N-1).(N-2)\ldots(N-m+1)$.

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

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

Using the fact that for all real-valued x, $1 + x \leq e^x$,

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

For $m = \sqrt{2N} + 1 \leq 28$, $\Pr(E) < 1/e < 1/2$

An Alternate Analysis: Assume that we choose one birthday after the other independently and uniformly at random from $[1 \ldots N]$. Let E_i denotes the event "the ith choice is different from the first $i-1$ choices".
After $i-1$ choices have been made, if it is given that all of these $(i-1)$ birthdays are distinct, the probability that the ith birthday is different from these $(i-1)$ birthdays, is given by

$$\Pr\{E_i | E_1 \cap E_2 \cap \ldots \cap E_{i-1}\} = \frac{N - (i-1)}{N} = 1 - \frac{i-1}{N}$$

The probability that all m people have different birthdays,

$$\Pr\{E\} = \Pr\{\cap_{i=1}^{m} E_i\}$$
$$= \Pr\{E_1\} \Pr\{E_2|E_1\} \Pr\{E_3|E_2 \cap E_1\} \ldots \Pr\{E_m| \cap_{i=1}^{m-1} E_i\}$$
$$= \Pi_{i=1}^{m}(1 - \frac{i-1}{N}) = \Pi_{i=0}^{m-1}(1 - \frac{i}{N})$$

The Principle of Deferred Decision: The alternate analysis use this principle: instead of choosing the m birthdays apriori according to the uniform distribution (as was effectively done in the first analysis) we choose them in a "step by step" fashion i.e., so to speak only when needed (we defer the choosing). Notice that the two approaches are equivalent, but in many situations, the deferred decisions approach leads to a much easier analysis.