EFFICIENT COMPUTING THROUGH RANDOM ALGORITHMS

Prof. Dan A. Simovici

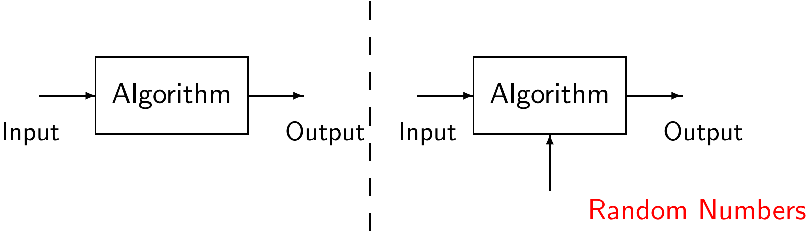
Doctoral Summer School
Iasi, Romania, June 2013

Deterministic vs. Randomized Algorithms

The common paradigm in algorithm design is that of deterministic algorithm.

- For a deterministic algorithm the input completely determines the sequence of computations performed by the algorithm.
- The behavior of random algorithms is determined not only on the input but also on several random choices.
- The same randomized algorithm, given the same input multiple times, may perform different computations in each invocation.
- The running time of a randomized algorithm on a given input is a random variable.

Deterministic vs. Random Algorithms



Determininstic Algorithm

Random Algorithm

Deterministic vs. Random Algorithm Design

- for deterministic algorithms, good behavior means that time requirements are polynomial in the size of the input;
- for random algorithms we need proof that it is highly likely that the behavior of the algorithm will be good on any input.

Probabilistic Analysis of Algorithms

- probabilistic Analysis of algorithms is an entirely distinct pursuit;
- random inputs having a given probability distributions are applied;
- goal is to show that the algorithm requires polynomial time on most inputs;

Las Vegas vs. Monte Carlo

- A Las Vegas algorithm provides a solution with a probability larger than $\frac{1}{2}$ and never gives an incorrect solution
- A Monte Carlo algorithm applies in situations when the algorithm makes a decision or a classification and provides a yes/no answer; if the answer is yes, then it confirms it with the probability larger than $\frac{1}{2}$, but if the answer is no, the algorithm will never give a definite result.
- The failure of the algorithm to return yes in a long series of trials gives evidence that the answer is no.

Example

Let A be an array on n components, where $n \geq 2$; suppose that half of the components of A are 1^{s} and the other half are 0^{s}. Find an 1 in the array. Consider the algorithms

```
LV(A,n)
begin
   repeat
     randomly select one out of n elements;
   until 1 is found
end
```

```
MC(A,n,k)
begin
   i = 1;
   repeat
     randomly select one out of n elements;
     i = i + 1;
   until i == k or an 1 is found;
end
```

The Las Vegas Algorithm

```
LV(A,n)
begin
   repeat
      randomly select one out of n elements;
   until an 1 is found;
end
```

- the algorithm succeeds with probability 1;
- the algorithm always outputs the correct answer;
- the running time is a random variable and arbitrarily large but the expected running time is finite.

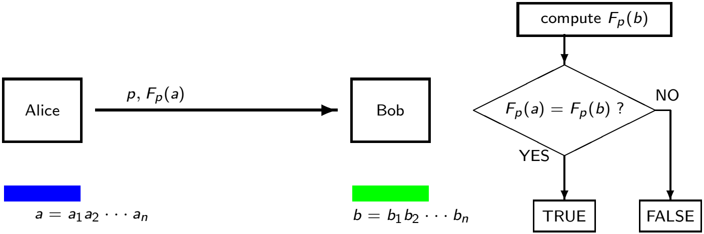
The Monte Carlo Algorithm

```
MC(A,n,k)
begin
   i = 1;
   repeat
      randomly select one out of n elements;
      i = i + 1;
   until i == k or an 1 is found;
end
```

- no guarantee of success;
- run time is fixed.

The Relationship between LV and MC

- a LV algorithm can be converted into a MC algorithm by having it output an arbitrary (possibly erroneous) output if it fails to complete under a specified time;
- a MC can be converted in a LV algorithm if there exists an efficient checking the correctness of the answer by repeatedly running the MC until it produces a correct answer.


Alice
p, $F_p(a)$
Bob
compute $F_p(b)$
$F_p(a) = F_p(b)$?
NO
YES
TRUE
FALSE
$a = a_1 a_2 \cdots a_n$
$b = b_1 b_2 \cdots b_n$

Comparing Binary Strings

Let $a = a_0 a_1 \cdots a_{n-1}$ and $b = b_0 b_1 \cdots b_{n-1}$ be two binary strings, where a is the binary representation of some natural number t.

A Monte Carlo algorithm:

- Alice choses a uniformly random random prime p, $2 \leqslant p \leqslant T$, where $t \leqslant T$. The fingerprint of a is $F_p(a) = a \mod p$.
- Alice sends $F_p(a)$ and p to Bob.
- Bob computes $F_p(b)$. If Bob sees $F_a(p) = F_b(p)$, then the algorithm outputs TRUE; otherwise, the algorithm outputs FALSE.

- There are no false negatives, since $a = b$ implies $F_p(a) = F_p(b)$.
- If $a \neq b$, we may still have $F_p(a) = F_p(b)$, which is a false positive.
- We claim that the probability of an error is small.

ALGORITHM OUTPUT

	TRUE	**FALSE**
$a = b$	$\surd$	~~**FALSE NEGATIVE**~~
$a \neq b$	**FALSE POSITIVE**	$\surd$

Let

$$\mathrm{prime}(x) = |\{p \mid p \text{ is prime and } p \leqslant x\}|.$$

Theorem

A non-zero n-bit integer has at most n distinct prime divisors.

Proof: Each prime divisor is at least 2 and the integer is not larger than $2^n - 1$. By the unique factorization theorem, there are no more than n prime divisors.

Since $|a - b|$ is a non-zero n-bit integer, there are at most n prime numbers that divide $|a - b|$. Therefore, the probability of an error is not larger than $\frac{n}{\text{prime}(T)}$.
The prime number theorem states that prime(x) is close to $\frac{x}{\ln x}$ as $x \to \infty$. Thus, the probability of error is less or equal to $n\frac{\ln T}{T}$.

To limit error choose $T = cn \ln n$. In this case

$$P(\text{error}) \leqslant n \frac{\ln(cn \ln n)}{cn \ln n} = \frac{1}{c} \left(1 + \frac{\ln(cn \ln n)}{\ln n} \right).$$

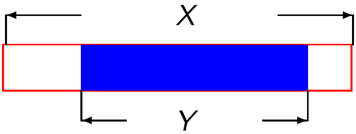
Since $\frac{\ln(cn \ln n)}{\ln n} = o(1)$ we have

$$P(\text{error}) = \frac{1}{c} + o(1).$$

Thus, with c large enough $P(\text{error})$ can be made as small as desired.

Pattern Matching Problem

- Problem: Given two input strings $X = x_0 x_1 \cdots x_{n-1}$ and $Y = y_0 y_1 \cdots y_{m-1}$, is Y a contiguous substring of X?
- Equivalent formulation: Is there a j, $1 \leqslant j \leqslant n - m$ such that for $X(j,m) = x_j x_{j+1} \cdots x_{j+m-1}$, $X(j,m) = Y$?



The Algorithm

- regard X and Y as binary integers;
- choose a random prime p, where $2 \leqslant p \leqslant T$;
- compute the fingerprints $F_p(Y)$ and $F_p(X(j,m))$ for $0 \leqslant j \leqslant n-m$;
- if there is some j such that $F_p(Y) = F_p(X(j,m))$, then output MATCH, otherwise output NO MATCH.

- there are no false negatives, but there may be false positives, when strings do not match, but the algorithm returns MATCH;
- if $X(j, m) \neq Y$ for $0 \leqslant j \leqslant n - m$, then, by the union bound,

$$P(\text{error}) \leqslant n \frac{\text{prime}(m)}{\text{prime}(T)}.$$

A Tighter Bound

- if $F_p(X(j,m)) = F_p(Y)$, then p divides $|X(j) - Y|$;
- if there is an error, then p divides the product $\prod_{j=0}^{n-m} |X(j,m) - Y|$.
- since $|X(j,m) - Y|$ is an m bit number and we multiply these, $\prod_{j=0}^{n-m} |X(j,m) - Y|$ is a most an mn-bit integer;
- therefore,

$$P(\text{error}) \leqslant \frac{\text{prime}(mn)}{\text{prime}(T)}.$$

- if $T = cmn$ we have $P(\text{error}) \leqslant \frac{\text{prime}(mn)}{\text{prime}(cmn)} = \frac{1}{c}\left(1 + \frac{\ln c}{\ln mn}\right)$.

Example

The size of a human chromosome ranges from $50 \cdot 10^6$ to $250 \cdot 10^6$ base pairs.

If we are looking for a string of length $m = 2^8$ in a DNA string of length $n = 2^{27}$ (within the ballpark of chromosome length), then by choosing $T = 2^{64}$ (so p is a 64-bit integer) gives

- $c = \frac{T}{mn} = \frac{2^{64}}{2^{35}} = 2^{29}$;
- $P(\text{error}) \leqslant \frac{1}{2^{29}} \left(1 + \frac{29}{35}\right)$, which is minuscule!

Polynomial Identity Testing

A polynomial $P(x_1, \ldots, x_n)$ over a field $\mathbb{F}$ can be written as a sum of monomials of the form $cx_1^{k_1} \cdots x_n^{k_n}$.
For example, for $p = (x + y)(2x + z^2)$ we can write

$$p(x, y, z) = 2x^2 + 2xy + xz^2 + yz^2.$$

Two Problems on Polynomials

- The "Evaluates to Zero Everywhere" (EZE) problem: Given a polynomial $D(x_1, \ldots, x_n)$ over $\mathbb{F}$, decide whether, for every choice of $y_1, \ldots, y_n$ in $\mathbb{F}$ the value of $D(y_1, \cdots, y_n)$ is 0.
- The Polynomial Identity Testing (PIT) Problem: Given a polynomial $D(x_1, \ldots, x_n)$, we can write it as a sum of monomials. If, upon expanding p to a sum of monomials, each coefficient is 0, then we say that p is the zero polynomial, or that it is identically zero.

- if p is identically zero then it evaluates to zero everywhere;
- if $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$ the converse is true;
- if $\mathbb{F}$ is some finite field the converse is false; for example if $p(x) = x^2 + x$ is a polynomial over $GF(2)$, then p is not identically zero, but $p(x) = 0$ for $x \in \{0, 1\}$.

The brute force approach is unfeasible. If we explicitly expand p and $\partial(p) = d$, then there could by $\binom{n+d}{d}$ monomials (which is exponential in d).

The Degree of a Multivariate Polynomial

A monomial is an expression of the form $\mu = ax_1^{b_1} \cdots x_n^{b_n}$ where $a \in \mathbb{F}$ and $b_1, \cdots, b_n \in \mathbb{N}$.
The degree of μ is $\sum_{i=1}^{n} b_i$.
The degree of a polynomial is the largest degree of any of its monomials.

Polynomial Representations

- Polynomials can be represented explicitly, as sums of monomials.
- Other forms are possible. For example, $V(x_1, \ldots, x_n)$ defined by

$$V(x_1, \ldots, x_n) = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \cdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix} = \prod_{i<j}(x_i - x_j)$$

is the Vandermonde polynomial of degree $\frac{n(n-1)}{2}$.

Given a polynomial $D(x_1, \ldots, x_n)$ of n variables and of degree d is D identical to the 0 polynomial?

Basic assumption: there is an efficient way of computing the values of D.

Algorithm:

Let $S \subseteq \mathbb{R}$ be a finite set. Pick at random uniformly and independently $r_1, \ldots, r_n$ from S. If $D(r_1, \ldots, r_n) = 0$ return YES; otherwise, return NO.

If $D \equiv 0$, the algorithm returns YES with probability 1, so the possible error is one-sided.

Theorem

The Inequality of Schwartz-Zippel If D is a polynomial of degree d and $D \not\equiv 0$, then

$$P(D(r_1, \dots, r_n) = 0) \leqslant \frac{d}{|S|}$$

Proof

The argument is by induction on n.
The base step $n = 1$ follows from the fact that there are at most d roots, so $P(D(r_1) = 0) \leqslant \frac{d}{|S|}$.
The inductive step: Note that D can be written as

$$D(x_1, \ldots, x_n) = \sum_{i=1}^{k} x_1^i Q_i(x_2, \ldots, x_n),$$

where k is the largest power of x_1 in a monomial of D.
By our choice of k the polynomial $Q_k(x_2, \ldots, x_n)$ is not identically 0 and its degree is no larger than $d - k$.

Proof (cont'd)

By the inductive hypothesis

$$P(Q_k(x_2, \ldots, x_n) = 0) \leqslant \frac{d-k}{|S|}.$$

Let K be the event "$Q_k(x_2, \ldots, x_n) = 0$".
Let us now randomly choose the values of $y_2, \ldots, y_n$ and assume that the event K did not occur. Define $\Delta(x_1)$ to be the univariate polynomial

$$\Delta(x_1) = \sum_{i=1}^{k} x_1^i Q_i(y_2, \ldots, y_n),$$

Since K did not occur, the degree of Δ is k. Thus,

$$P(\Delta(y_1) = 0|\bar{K}) \leqslant \frac{k}{|S|}.$$

Let L be the event "$\Delta(y_1) = 0$", clearly equivalent to $D(y_1, y_2, \ldots, y_n) = 0$.
We have

$$\begin{aligned} P(L) &= P(L \cap K) + P(L|\bar{K})P(\bar{K}) \\ &\leqslant P(K) + P(L|\bar{K})P(\bar{K}) \\ &\leqslant \frac{k}{|S|} + \frac{d-k}{|S|} = \frac{d}{|S|}. \end{aligned}$$

Recall

Markov's Inequality:

If X is a non-negative random variable having the expected value $E[X]$, then $P(X \geqslant a) \leqslant \frac{E[X]}{a}$.

Proof: (for the discrete case). Let $X : \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$, where $x_1 > x_2 > \cdots > x_n$, $p_i \geqslant 0$ for $1 \leqslant i \leqslant n$ and $\sum_{i=1}^{n} p_i = 1$. If $x_i \geqslant a > x_{i+1}$, we have

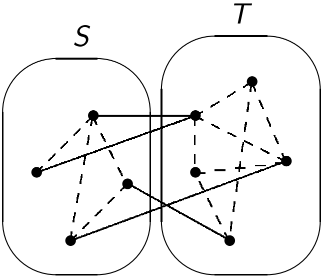
$$P(X \geqslant a) = P((X = x_1) \cup (X = x_2) \cup \cdots \cup (X = x_i)) = p_1 + \cdots + p_i.$$

On other hand,

$$\begin{aligned} E[X] &= x_1 p_1 + \cdots + x_i p_i + x_{i+1} p_{i+1} + \cdots + x_n p_n \\ &\geqslant x_1 p_1 + \cdots + x_i p_i \\ &\geqslant a(p_1 + \cdots + p_i) = aP(X \geqslant a). \end{aligned}$$

Maximal Cut

Given a graph $\mathcal{G} = (V, E)$ find a partition $\{S, T\}$ of the set V of vertices (called a cut) such that the number of edges between S and T is maximal. The set of edges that join a vertex in S with a vertex in T is denoted by $E(S, T)$ and the size of the cut is $|E(S, T)|$.



Theorem

In any graph $\mathcal{G} = (V, E)$ there exists a cut with at least half of the edges crossing it.

Proof

Let S be a random subset of V. A vertex belongs to S with probability 0.5. The indicator of an edge e is the random variable X_e, where

$$X_e = \begin{cases} 1 & \text{if } e \in E(S,T), \\ 0 & \text{otherwise} \end{cases}$$

For $X = |E(S,T)|$ we have $X = \sum_{e \in E} X_e$ and

$$E[X_e] = P(e \in E(S,T)) \cdot 1 + P(e \notin E(S,T)) \cdot 0 = \frac{1}{2}$$

because the probability that an end of e is in S and the other is not is $\frac{1}{2}$. Therefore $E[X] = \sum_{e \in E} E[X_e] = \frac{|E|}{2}$. There exits an event where X takes a value at least $E[X]$, so there is a cut with at least half the edges.

The Random Assignment of Vertices

Suppose that in order to build a cut (S, T) we assign vertices at random to S or T.
Let $Y = |E| - X$ be the number of edges that to not cross from S to T.
We have $Y \geqslant 0$ and $E[Y] = |E| - E[X] = \frac{|E|}{2}$. By Markov's inequality

$$P(Y \geqslant aE[Y]) = P(Y \geqslant \frac{a|E|}{2}) \leqslant \frac{1}{a}.$$

For $a = 1.5$ we have

$$P(Y \geqslant \frac{3}{4}|E|) \leqslant \frac{2}{3}, \text{ or } P(Y < \frac{3}{4}|E|) > \frac{1}{3}.$$

Therefore, $P(X > \frac{1}{4}|E|) \geqslant \frac{1}{3}$, which shows that a random cut will have at least a quarter of the edges with a probability of at least $\frac{1}{3}$.

The 3-SAT Problem

The 3-SAT problem starts with a formula in conjunctive normal form

$$\varphi = C_1 \wedge C_2 \wedge \cdots \wedge C_m,$$

where each clause C_i is disjunction of three distinct literals of the form $C_i = \ell_j \vee \ell_k \vee \ell_h$, and seeks to determine if there exists a truth assignment that satisfies all φ.
Here ℓ is either a propositional variable x_i or its negation $\bar{x}_i$.

Example

Consider the clause $C = x_1 \vee \bar{x}_2 \vee \bar{x}_3$ and the list of truth assignments on the set of variables of this clause:

v_1	0	0	0	$\surd$
v_2	0	0	1	$\surd$
v_3	0	1	0	$\surd$
v_4	0	1	1	$-$
v_5	1	0	0	$\surd$
v_6	1	0	1	$\surd$
v_7	1	1	0	$\surd$
v_8	1	1	1	$\surd$

One out of every eight truth assignments fails to satisfy the clause!

Example

Let φ be the formula

$$(\bar{x}_1 \vee \bar{x}_2 \vee x_3) \wedge (x_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (x_1 \vee x_2 \vee x_3).$$

The truth assignment v on $\{x_1, x_2, x_3\}$ given by $v(x_1) = 1$, and $v(x_2) = v(x_3) = 0$ satisfies φ.

Instead of solving SAT let's seek a truth assignment that satisfies the maximum number of clauses.

Theorem

For every formula φ there exists a truth assignment that satisfies $\frac{7m}{8}$ clauses.

Proof

Choose randomly a truth assignment $v : \{x_1, \ldots, x_n\} \longrightarrow \{\mathbf{T}, \mathbf{F}\}$. Define

$$Y_i = \begin{cases} 1 & \text{if } C_i \text{ is satisfied,} \\ 0 & \text{otherwise.} \end{cases}$$

The number of satisfied clauses is $Y = \sum_{i=1}^{m} Y_i$.
Among 8 truth assignments to the variables of C_i only one fails to satisfy C_i. Thus, we have $E[Y_i] = P(C_i \text{ is satisfied}) = \frac{7}{8}$, so $E[Y] = \sum_{i=1}^{m} E[Y_i] = \frac{7m}{8}$. Since there exists an event in the probability space such that Y is greater than $E[Y]$, there exists an assignment that satisfies $\frac{7}{8}$ of the clauses.

How to Get a Good Truth Assignment

Let $Z = m - Y$ be the number of unsatisfied clauses, $Z \geqslant 0$. We have

$$E[Z] = m - E[Y] = \frac{m}{8}.$$

By Markov's Inequality,

$$P\left(Z \geqslant aE[Z]\right) = P\left(Z \geqslant \frac{am}{8}\right) \leqslant \frac{1}{a},$$

so for $a = 2$, $P(Z \geqslant m/4) \leqslant 1/2$, which implies

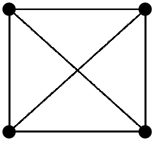
$$P\left(Y > \frac{3m}{4}\right) > \frac{1}{2}.$$

Thus, a randomly chosen assignment satisfies at least three quarters of clauses with at least 0.5 probability!

Let $\Gamma_{n,p}$ the distribution of random undirected graphs with n vertices such that an edge exists with probability p. We say that $G \sim \Gamma_{n,p}$ is $G = (V, E)$ belongs to this distribution.

- a graph in $\Gamma_{n,p}$ with a given set of m edges has the probability $p^m(1-p)^{\binom{n}{2}-m}$;
- a graph in $\Gamma_{n,p}$ can be generated by considering each of the $\binom{n}{2}$ edges and then, independently add each edge to the graph with probability p; the expected number of edges is $\binom{n}{2}p$ and each vertex has expected degree $(n-1)p$.

Studying $\Gamma_{n,p}$ yields interesting a powerful results. For example, for $G \sim \Gamma_{n,p}$ does G contain a clique having four vertices?



Define for every set C of 4 vertices in $\mathcal{G}$, an indicator variable I_C by

$$I_C = \begin{cases} 1 & \text{if } C \text{is a clique} \\ 0 & \text{otherwise.} \end{cases}$$

There are $\binom{n}{4}$ sets C, so there is this number of indicator variables.
If X_n is the number of 4-cliques in a graph with n vertices, $X_n = \sum\{I_C | C \subseteq V, |C| = 4\}$. There are six edges in a 4-clique, and each is chosen independently, hence

$$E[I_C] = P(I_C = 1) = p^6,$$

because each of the six edges are chosen independently. This implies

$$E[X_n] = \binom{n}{4} p^6 = \Theta(n^4 p^6).$$

- Note that $P(X_n > 0) = P(X_n \geqslant 1)$ Thus, if $\lim_{n\to\infty} n^4 p^6 = 0$ (written as $p \ll n^{-\frac{2}{3}}$), then $\lim_{n\to\infty} P(X_n > 0) = 0$.
- We claim that if $p \gg n^{-\frac{2}{3}}$, then $\lim_{n\to\infty} P(X_n > 0) = 1$.

Recall

- For a random variable X, the variance is $var(X)$ is $E[(X - E[X])^2]$. We also have $var(X) = E[X^2] - (E[X])^2$.
- For any two random variables X and Y , the covariance $cov(X, Y)$ is $E[XY] - E[X]E[Y]$. If X and Y are independent, then $cov(X, Y) = 0$.
- Chebyshevs Inequality: $P(|X - E[X]| \geqslant a) \leqslant \frac{var(X)}{a^2}$.
  Proof: Let $Y = (X - E[X])2$. Y is a non-negative random variable, so by applying Markov Inequality,

$$P(|X - E[X]| \geqslant a) = P(Y \geqslant a^2) \leqslant \frac{E[Y]}{a^2} = \frac{var(X)}{a^2}.$$

Proof that $p \gg n^{-\frac{2}{3}}$, implies $\lim_{n\to\infty} P(X_n > 0) = 1$

Note that $X_n = 0$ implies $|X_n - E[X_n]| = |E[X_n]| \geqslant E[X_n]$.
Therefore,

$$P(X_n = 0) \leqslant P\left(|X_n - E[X_n]| \geqslant E[X_n]\right) \leqslant \frac{var(X_n)}{E[X_n]^2} = \frac{E[X_n^2] - E[X_n]^2}{E[X_n]^2}.$$

We claim that $E[X_n^2] - E[X_n]^2$ is small compared to $E[X_n]^2$.

Proof (cont'd)

$$
\begin{aligned}
var(X_n) &= E[X_n^2] - E[X_n]^2 = E\left[\left(\sum_C I_C\right)^2\right] - E\left[\sum_C I_C\right]^2 \\
&= E\left[\sum_C I_C^2 - \sum_{C \neq D} I_C I_D\right] - \left(\sum_C E[I_C]\right)^2 \\
&= \sum_C E[I_C^2] - \sum_{C \neq D} E[I_C I_D] - \sum_C E[I_C]^2 + \sum_{C \neq D} E[I_C]E[I_D] \\
&= \sum_C var(I_C) + \sum_{C,D} cov(I_C, I_D).
\end{aligned}
$$

Evaluation of $cov(I_C, I_D)$

Cases to consider:

- if $|C \cap D| \leqslant 1$, no common edges exist, so I_C, I_D are independent, which implies $cov(I_C, I_D) = 0$;
- if $|C \cap D| = 2$, one pair of vertices is shared, so we need only 11 edges to be present; thus, $cov(I_C, I_D) = E[I_C I_D] - p^{12} = p^{11} - p^{12} \leqslant p^{11}$; this can happen $\binom{n}{6}$ times, so the total contribution is less than $\binom{n}{6} p^{11} = \Theta(n^6 p^{11})$;
- if $|C \cap D| = 3$, three pairs of vertices are shared, so three fewer edges are needed; thus, $cov(I_C, I_D) = E[I_C I_D] - p^{12} = p^9 - p^{12} \leqslant p^9$; this may happen $\binom{n}{5}$ times, so the total contribution to the sum is $\binom{n}{5} p^9 = \Theta(n^5 p^9)$;

Evaluation of $cov(I_C, I_D)$ (cont'd)

We have

$$var(I_C) = E[I_C^2] - E[I_C]^2 = p^6 - p^{12} = \Theta(p^6),$$

which implies

$$\begin{aligned}
var(X_n) &= \sum_C var(I_C) + \sum_{C \neq D} cov(I_C, I_D) \\
&\leqslant \Theta(n^4 p^6) + \Theta(n^6 p^{11}) + \Theta(n^5 p^9) \\
&= \Theta(n^4 p^6) + \Theta(n^6 n^{-\frac{22}{3}}) + \Theta(n^5 n^{-6}) \\
&\quad \text{(taking into account that } p \gg n^{-\frac{2}{3}}) \\
&= \Theta(n^4 p^6)
\end{aligned}$$

Evaluation of $cov(I_C, I_D)$ (cont'd)

Finally,

$$\frac{var(X_n)}{(E[X])^2} = \frac{\Theta(n^4 p^6)}{\Theta(n^4 p^6)^2} = \frac{1}{\Theta(n^4 p^6)},$$

so $\lim_{n\to\infty} \frac{var(X_n)}{(E[X])^2} = 0$ because $p \gg n^{-\frac{2}{3}}$.

Given three matrices $A \in \mathbb{R}^{m \times p}$, $B \in \mathbb{R}^{p \times n}$ and $C \in \mathbb{R}^{m \times n}$ determine if $AB = C$.

Freivalds' Monte Carlo algorithm:

begin
    $i = 1$;
    repeat
        choose $\mathbf{r} = (r_1, \dots, r_n)' \in \{0, 1\}^n$ at random;
        compute $C\mathbf{r}$ and $A(B\mathbf{r})$;
        if $C\mathbf{r} \neq A(B\mathbf{r})$;
            return FALSE;
        endif;
        $i = i + 1$;
    until $i = k$;
    return TRUE
end

Theorem

Freivalds' algorithm is correct with a probability at least equal to $1 - 2^{-k}$.

Proof: We show that if $AB \neq C$, then $P(A(B\mathbf{r}) = C\mathbf{r}) \leqslant 1/2$.
If $AB \neq C$, then $D = AB - C \neq 0$. Without loss of generality we may assume that $d_{11} \neq 0$. Note that $A(B\mathbf{r}) = C\mathbf{r}$ is equivalent to $D\mathbf{r} = \mathbf{0}$ and this implies $\sum_{j=1}^{n} d_{1j} r_j = 0$.
Since $d_{11} \neq 0$, we have $r_1 = -\frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}$. This equality holds for at most one of the two choices we have for r_1, so $P(AB\mathbf{r} = C\mathbf{r}) \leqslant 0.5$.
If $C = AB$ the algorithm is always correct; if $C \neq AB$ the probability of a correct answer is $1 - 0.5^k$ because the loop is run for k times.

Finding the nearest pair of points

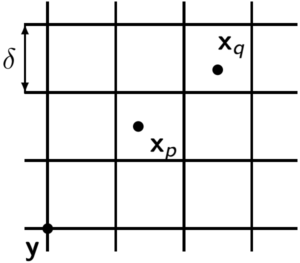
This is the first probabilistic algorithm by M. Rabin, published in 1976!
Problem Statement: given n points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in the unit square $[0, 1]^2$ in $\mathbb{R}^2$ find two points that are the closest with respect to Euclidean distance.
To simplify the presentation assume that there is a unique closest pair. If there are several with the same minimum distance the algorithm still works.
The problem can clearly be solved in $O(n^2)$, but randomness allows a better result!

Outline of Rabin's Algorithm

Let S be a set of points in $\mathbb{R}^2$ and let

$$\delta(S) = \min\{d(\mathbf{u},\mathbf{v}) \mid \mathbf{u},\mathbf{v} \in S \text{ and } \mathbf{u} \neq \mathbf{v}\}$$

Consider a mesh of squares $\mathcal{M}$ having the size δ.



Key remark: Even if $\delta(S) \leqslant \delta$ we cannot be certain that the nearest pair is in the same square of the mash. However, we are sure that at worst the closest pair lies in squares with a common vertex.

Lemma

If $\delta(S) \leqslant \delta$, where δ is the mesh size, then there exists a mesh point $\mathbf{y}$ such that the nearest pair lies in a quadruple of squares situated at the north and east of $\mathbf{y}$.

S can be partitioned in a union of sets, $S = S_1 \cup \cdots \cup S_k$ such that each S_i consists of all the points of S within one square of $\mathcal{M}$.
Define $N(\mathcal{M}) = \sum_{i=1}^{k} \frac{n_i(n_i-1)}{2}$.

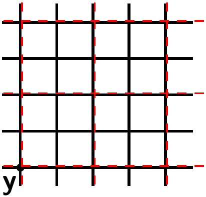
- if we know that the nearest pair is within one of the sets S_i, then it can be discovered by performing $N(\mathcal{M})$ computations;
- under the previous assumption, the nearest pair will be discovered after $N(\mathcal{M}) - 1$ comparisons between the computed distances;
- thus, we are interested in finding a mesh $\mathcal{M}$ for which $N(\mathcal{M}) = O(n)$.

The effect of increasing the mesh size δ

Lemma

Let $\mathcal{M}$ be a mesh of size δ. Construct a mesh $\mathcal{M}_1$ by choosing a fixed mesh point **y** *of $\mathcal{M}$ as origin and forming a mesh of size 2δ and lines parallel to those of $\mathcal{M}$. Then, for a fixed set S we have*

$$N(\mathcal{M}_1) \leqslant 16N(\mathcal{M}) + 24n.$$



Proof

The squares of $\mathcal{M}_1$ are quadruples of squares of $\mathcal{M}$ and yield the partition $S = T_1 \cup \cdots \cup T_q$.

- each T_i is the union of at most four of the sets S_j;
- if $|T_i| = m_i$ and $T_i = S_{j_1} \cup \cdots \cup S_{j_4}$, then $m_i \leqslant n_{j_1} + \cdots + n_{j_4}$;
- if $k_i = \max\{n_{j_1}, \ldots, n_{j_4}\}$, then

$$\frac{m_i(m_i - 1)}{2} \leqslant \frac{4k_i(4k_i - 1)}{2} = \frac{16k_i(k_i - 1)}{2} + 6k_i.$$

- since $k_1, \ldots, k_m$ are a subset of $n_1, \ldots, n_k$ and $\sum n_i \leq 4n$ because every $\mathbf{x}_i$ in S is in at most four S_js, the conclusion follows.

Theorem

There exists a constant c_1 so that for every S, $\mathcal{M}$ and $\mathcal{M}_1$ as above, if $N(\mathcal{M}) \leqslant cn$, then $N(\mathcal{M}_1) \leqslant c_1 cn$.

This holds (with an appropriate c_1) for any fixed linear blowup of the mesh size of $\mathcal{M}$.

Theorem

For any set $S \subseteq \mathbb{R}^2$, where $|S| = n$, there exists a mesh $\mathcal{M}_0$ so that $\mathcal{M}_0$ creates a partition $\{S_1, \ldots, S_k\}$ such that $n \leqslant N(\mathcal{M}_0) \leqslant c_1 n$, where c_1 is the previous constant.

Proof

Choose a pair of perpendicular directions ℓ_1, ℓ_2 intersecting at $\mathbf{y}$ such that $S \cap (\ell_1 \cup \ell_2) = \emptyset$;

- form a mesh $\mathcal{M}$ using ℓ_1, ℓ_2 and $\mathbf{y}$ with a small enough size such that each square contains one point of S and no point of S is located on the grid lines;
- by successively doubling the mesh size we reach *for the first time* a mesh $\mathcal{M}_0$ for which $n \leqslant N(\mathcal{M}_0)$;
- when this occurs for the first time we have $N(\mathcal{M}) \leqslant c_1 n$.

Outline of Probabilistic Algorithm

- randomly select a sample of m points, $S_1 = \{\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_m}\}$ of the set $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$; find $\delta(S_1)$;
- construct a mesh $\mathcal{M}$ with mesh size $\delta_1(S_1)$;
- if $m = m(n)$ is appropriately chosen, then with high probability we have $N(\mathcal{M}) = O(n)$;
- since $\delta(S) \leqslant \delta$, by a previous lemma, the nearest pair in S will be located in a square of one of the meshes of size 2δ.

Success on a Partition

- Let $\pi = \{S_1, \ldots, S_k\}$ be a partition of S. If $t : \{1, \ldots, m\} \longrightarrow S$ is an injection, that is, a choice of m elements of S, then t is an (m, π)-success if there is a block S_i of π that contains at least two elements in the range of t. Otherwise we call t an (m, π)-failure.
- If σ is another partition, $\sigma = \{H_1, \ldots, H_\ell\}$ of S, then we say that π dominates σ if for every m, the probability of an (m, π)-success is at least equal to the probability of an (m, σ) success on σ.
- t is an (m, π)-failure if no block of π contains more than one element in the range of t.

How Many Failures and Successes?

$t : \{1, \ldots, m\} \longrightarrow S$ is an (m, π)-failure iff the function $T : \{1, \ldots, m\} \longrightarrow S/\pi$ given by $T(p) = [t(p)]$ for $1 \leqslant p \leqslant m$ is an injection. If $|S/\pi| = k$, the number of injections T is

$$\begin{cases} \frac{k!}{(k-m)!} & \text{if } m \leqslant k, \\ 0 & \text{if } k < m \leqslant n. \end{cases}$$

Therefore, the number of (m, π)-successes is

$$\begin{cases} \frac{n!}{(n-m)!} - \frac{k!}{(k-m)!} & \text{if } m \leqslant k, \\ \frac{n!}{(n-m)!} & \text{if } k < m. \end{cases}$$

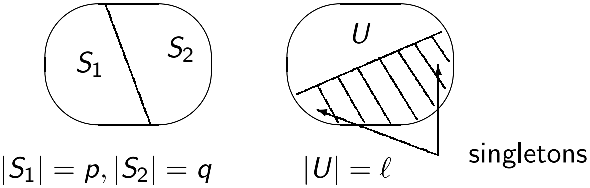
The probability of an (m, π)-success is

$$P(m, \pi, k, n) = \begin{cases} 1 - \frac{\frac{k!}{(k-m)!}}{\frac{n!}{(n-m)!}} & \text{if } m \leqslant k, \\ 1 & \text{if } k < m. \end{cases}$$

Example

$\tau = \{S_1, S_2\}$ be a partition of a set $S = S_1 \cup S_2$ with $|S_1| = p > 1$ and $|S_2| = q > 1$.
Claim: τ dominates the partition σ of S that consists of a block U with $|U| = \ell$ and $p + q - \ell$ singletons if and only if

$$\ell(\ell - 1) \leqslant p(p-1) + q(q-1)$$



Example (cont'd)

We have

$$P(m,\tau,k,n) = \begin{cases} 1 - \frac{\frac{2!}{(2-m)!}}{\frac{n!}{(n-m)!}} & \text{if } m \leqslant 2, \\ 1 & \text{if } 2 < m \end{cases} = \begin{cases} 1 - \frac{2}{n} & \text{if } m = 1 \\ 1 & \text{if } m \geqslant 2, \end{cases}$$

and

$$P(m,\sigma,n) = \begin{cases} 1 - \frac{\frac{(p+q-\ell+1)!}{((p+q-\ell+1)-m)!}}{\frac{n!}{(n-m)!}} & \text{if } m \leqslant p+q-\ell+1, \\ 1 & \text{if } k < p+q-\ell+1. \end{cases}$$

Therefore, we must justify the inequality $P(m,\tau,n) \geqslant P(m,\sigma,n)$ only for $m = 1$ and $m = 2$.

Example

Any partition π of a set of six elements into three two-element sets dominates any partition σ of the same set into a 3-element set and three singletons.
The probability of a success in the first case is

$$P(3,\pi,6) = \begin{cases} 1 - \frac{\frac{3!}{(3-m)!}}{\frac{6!}{(6-m)!}} & \text{if } m \leqslant 3, \\ 1 & \text{if } 3 < m. \end{cases}$$

In the second case, the probability is

$$P(4,\sigma,6) = \begin{cases} 1 - \frac{\frac{4!}{(4-m)!}}{\frac{6!}{(6-m)!}} & \text{if } m \leqslant 4, \\ 1 & \text{if } 4 < m. \end{cases}$$

Clearly, $\frac{3!}{(3-m)!} \leqslant \frac{4!}{(4-m)!}$ if $m \leqslant 3$. For $m = 4$, $P(3,\pi,6) = 1 \geqslant P(4,\pi,6)$, so π dominates σ.

Example

Any partition π of a set of six elements into two three-element sets dominates any partition σ of the same set into a 4-element set and two singletons.

We have

$$P(2,\pi,6) = \begin{cases} 1 - \frac{\frac{2!}{(2-m)!}}{\frac{n!}{(n-m)!}} & \text{if } m \leqslant 2, \\ 1 & \text{if } 2 < m. \end{cases}$$

and

$$P(3,\sigma,6) = \begin{cases} 1 - \frac{\frac{3!}{(3-m)!}}{\frac{n!}{(n-m)!}} & \text{if } m \leqslant 3, \\ 1 & \text{if } 3 < m. \end{cases}$$

Since $\frac{2!}{(2-m)!} \leqslant \frac{3!}{(3-m)!}$ it follows, as before, that π dominates σ.

Exercise: Prove that if $\pi = \{B_1, B_2\}$ is a partition with $|B_1| = |B_2| = 4$ of a set S with $|S| = 8$ and $\sigma = \{H_1, H_2, H_3, H_4\}$ with $|H_1| = 5$, $|H_2| = |H_3| = |H_4| = 1$, then π dominates σ.

The Sum of Two Partitions

Definition

Let S, S' be two disjoint sets and let $\pi = \{B_1, \ldots, B_k\}$ be a partition of S and $\sigma = \{H_1, \ldots, H_\ell\}$ be a partition of S'. The sum of π and σ is the partition $\pi + \sigma$ of $S \cup S'$ given by

$$\pi + \sigma = \{B_1, \ldots, B_k, H_1, \ldots, H_\ell\}.$$

Note that $N(\pi + \pi') = N(\pi) + N(\pi')$.

Theorem

Let S, S' be two disjoint sets, π be a partition of S and σ_1, σ_2 be two partitions of S'. If σ_1 dominates σ_2, then $\pi + \sigma_1$ dominates $\pi + \sigma_2$.

Proof is left as exercise.

Partition Transformations

- Any pair of blocks B_i, B_j in π such that $|B_i| = |B_j| = 3$ can be replaced with a triplet of blocks H_1, H_2, H_3 such that $|H_1| = 4$, $|H_2| = H_3| = 1$ to yield a partition σ such that π dominates σ and $N(\pi) = N(\sigma)$.
- Any triplet of blocks B_i, B_j, B_k in π such that $|B_i| = |B_j| = |B_k| = 2$can be replaced with a quadruple of sets H_1, H_2, H_3, H_4 with $|H_1| = 3$, $|H_2| = |H_3| = |H_4| = 1$ to yield a partition σ such that π dominates σ and $N(\pi) = N(\sigma)$.

A Special Partition Transformation

- Any pair of blocks B_i, B_j in π such that $|B_i| = |B_j| = 4$ can be replaced with a quadruple of blocks H_1, H_2, H_3, H_4 such that $|H_1| = 5$, $|H_2| = H_3| = |H_4| = 1$ to yield a partition σ such that π dominates σ and $N(\sigma) \geqslant \frac{5}{6}N(\pi)$.

Indeed, $N(\sigma) = \cdots + 10$, $N(\pi) = \cdots + 12$ and $\frac{T+10}{T+12} \geqslant \frac{10}{12}$ when $T > 0$.

Theorem

There exists a constant λ, $\lambda > 0$, such that for every partition π of a finite set S there exists another partition σ of S such that

- *π dominates σ,*
- *$\lambda N(\pi) \leqslant N(\sigma)$, and*
- *all blocks of σ, with one exception are singletons.*

Proof

Let $\pi = \{S_1, \ldots, S_k\}$. We may assume that each block that is not a singleton contains at least five elements. Further, suppose initially that k is even and non-singletons can be arranged in pairs.
Let (S_1, S_2) be such a pair with $|S_1| = p \geqslant 5$ and $|S_2| = q \geqslant 5$. $\{S_1, S_2\}$ is a partition of $S_1 \cup S_2$ and this partition dominates a partition of $S_1 \cup S_2$ that consists of a block U with $|U| = \ell$ and the remaining blocks being g singletons, where $p + q = \ell + 1 + \cdots + 1$ if and only if

$$\ell(\ell - 1) \leqslant p(p-1) + q(q-1)$$

If ℓ is the largest number with this property then

$$\ell(\ell - 1) \leqslant p(p-1) + q(q-1) \leqslant (\ell + 1)\ell,$$

Proof (cont'd)

The second inequality implies

$$\left(1 - \frac{2}{\ell+1}\right)\left[\frac{p(p-1)}{2} + \frac{q(q-1)}{2}\right] \leqslant \frac{\ell(\ell-1)}{2}$$

Let π be a partition of S and let
$m(\pi) = 1 + \min\{|B| \mid B \in \pi \text{ and } |B| > 1\}$.
If each of the paired sets is replaced in the manner previously described, using ℓ that satisfies the double inequality, then we obtain a partition π_1 that is dominated by π. Since $m(\pi_1) \leqslant \ell + 1$ holds for each pair in π we have

$$\left(1 - \frac{2}{m(\pi_1)}\right) N(\pi) \leqslant N(\pi_1).$$

Proof (cont'd): Estimation of a lower bound for $m(\pi_1)$

- if $p \leqslant q$, then the inequality $p(p-1)+q(q-1) \leqslant (\ell+1)\ell$ implies
$$(p+1)^2 \frac{2p(p-1)}{(p+1)^2} \leqslant (\ell+1)^2;$$
- since for $p \geqslant 5$, we have
$$\sqrt{\frac{40}{36}} \leqslant \sqrt{\frac{2p(p-1)}{(p+1)^2}} \leqslant \sqrt{2},$$
it follows that $s(p+1) \leqslant \ell+1$ where $\sqrt{\frac{40}{36}} \leqslant s$;
- from σ_1 we can obtain a partition σ_2 using the same process that allowed us to obtain σ_1 from σ;
- repeating this sufficiently many times (about $\log_{\sqrt{2}} k$ times, where $k = |\pi|$ we obtain a partition σ' of S which is dominated by π, and all the blocks of σ' except 1 are singletons.

Proof (cont'd)

For the constant λ we have

$$\lambda = \frac{5}{6}\left(1 - \frac{2}{6s}\right)\left(1 - \frac{2}{6s^2}\right)\cdots \tag{1}$$

Since the series $\frac{2}{6s} + \frac{2}{6s^2} + \cdots$ converges we have $\lambda > 0$ and $N(\pi) \leqslant N(\pi')$.

Theorem

Let π be a partition of the set S, $|S| = n$ such that $n \leqslant N(\pi)$. If $n^{\frac{2}{3}}$ pairwise distinct points are drawn at random from S, then the probability of success, i.e., the probability that two elements will be chosen from the same block of π is at least $\mu(n) = 1 - 2e^{-cn^{\frac{1}{6}}}$, where $c = \sqrt{2\lambda}$ for λ defined in Equality (1).

Proof

By Theorem given on slide 79, π dominates a partition $\sigma = \{H_1, \dots, H_m\}$, where $|H_i| = 1$ for $2 \leqslant i \leqslant m$ and $\lambda n \leqslant N(\sigma)$. Thus, for $p = |H_1|$, we have $2\lambda n \leqslant p(p-1)$ so that $c\sqrt{n} \leqslant p$ for $c \approx \sqrt{2\lambda}$.

- the probability that in one choice from S we miss H_1 is $1 - \frac{p}{n}$, so smaller than $1 - \frac{c}{\sqrt{n}}$;
- for $n^{\frac{2}{3}}$ choices the probability of all missing H_1 is smaller than

$$\left(1 - \frac{c}{\sqrt{n}}\right)^{\sqrt{n}\cdot n^{\frac{1}{6}}} \approx e^{-cn^{\frac{1}{6}}};$$

- the probability of success (at least two hits in H_1) is greater than $1 - 2e^{-cn^{\frac{1}{6}}}$.

Theorem

There exists a constant c_2 *so that if we choose at random* $S_1 = \{\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_m}\}$, $m = n^{\frac{2}{3}}$, *out of* $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ *and draw any mesh* $\mathcal{M}$ *of size* $\delta(S_1)$, *then the probability that* $N(\mathcal{M}) \leqslant c_2 n$ *is greater than* $\mu(n)$, *where* $\mu(n)$ *was defined in the Theorem given on slide 84.*

Proof

For the set S consider the mesh $\mathcal{M}_0$ of size δ_0 given in the Theorem of slide 65, by which S is partitioned so that $n \leqslant N(\mathcal{M}_0) \leqslant c_1 n$. Since $|S_1| = n^{\frac{2}{3}}$, the probability that two points of S_1 fall with one square of $\mathcal{M}_0$ is greater than $\mu(m)$.

- there are 16 meshes $\mathcal{M}_1, \ldots, \mathcal{M}_{16}$ derived from $\mathcal{M}_0$ by quadrupling the mesh size δ_0; the basic square of each consists of 16 basic squares of $\mathcal{M}_0$;
- if $\delta(S_1) \leqslant \delta_0\sqrt{2}$, then for any square mesh with mesh size δ_1 each of its squares will be a subset of a square of one of the $\mathcal{M}_i$, $1 \leqslant i \leqslant 16$; thus, $N(\mathcal{M}) \leqslant \sum_{i=1}^{16} N(\mathcal{M}_i)$;
- since $N(\mathcal{M}_0) \leqslant c_1 n$, by the Theorem on slide 64, $N(\mathcal{M}_i) \leqslant c_1^3 n$ for $1 \leqslant i \leqslant 16$.

Thus, with probability greater than $\mu(n)$, $N(\mathcal{M}) \leqslant 16 c_1^3 n$.

Theorem

For any set S, $|S| = n$, if S_1 is a subset of S such that $|S| = n^{\frac{2}{3}}$ is chosen at random and a mesh $\mathcal{M}$ of size $\delta(S_1)$ is formed, then the expected value of $N(\mathcal{M})$ is smaller than $c_2 n$.

Proof

By the Theorem on slide 86

$$P(S_1 \mid c_2 n \leqslant N(\mathcal{M})) \leqslant 1 - mu(n) = 2e^{-cn^{\frac{1}{6}}}.$$

Since $N(\mathcal{M}) \leqslant n(n-1)$, the expected contribution from the choices of S_1 for which mesh size $\delta(S_1)$ leads to a mesh $\mathcal{M}$ with $c_2 n \leqslant N(\mathcal{M})$, is smaller than $n(n-1)e^{-cn^{\frac{1}{6}}}$, which tends to 0 when n grows. Hence, the expected value of $N(\mathcal{M})$ is smaller than $(c_2 + \epsilon)n$ for $\epsilon > 0$ and n sufficiently large.

Conclusions

- The main benefits of random algorithms: simplicity and speed.
- A wide variety of applications.
- Beautiful mathematics!

Thank you for your attention!

Sildes are available at www.dsim at cs.umb.edu