The Probabilistic Method

1. If $E[X] = C$, then there are values $c_1 \leq C$ and $c_2 \geq C$ such that $Pr(X = c_1) > 0$ and $Pr(X = c_2) > 0$.
2. If a random object in a set satisfies some property with positive probability then there is an object in that set that satisfies that property.

Theorem

Given any graph $G = (V, E)$ with n vertices and m edges, there is a partition of V into two disjoint sets A and B such that at least $m/2$ edges connect vertex in A to a vertex in B.

Proof.

Construct sets A and B by randomly assign each vertex to one of the two sets.
The probability that a given edge connect A to B is $1/2$, thus the expected number of such edges is $m/2$.
Thus, there exists such a partition. $\square$

Maximum Satisfiability

Given m clauses in CNF (Conjunctive Normal Form), assume that no clause contains a variable and its complement.

Theorem

For any set of m clauses there is a truth assignment that satisfy at least $m/2$ of the clauses.

Proof.

Assign random values to the variables. The probability that a given clause (with k literals) is not satisfied is 2^{-k}, so the probability that it is satisfied is

$$1 - 2^{-k} \geq \frac{1}{2}.$$

$\square$

Monochromatic Complete Subgraphs

Given a complete graph on 1000 vertices, can you color the edges in two colors such that no clique of 20 vertices is monochromatic?

Theorem

If $n \leq 2^{k/2}$ then it is possible to edge color the edges of a complete graph on n points (K_n), such that is has no monochromatic K_k subgraph.

Proof:

Consider a random coloring.

For a given set of k vertices, the probability that the clique defined by that set is monochromatic is bounded by

$$2 \times 2^{-\binom{k}{2}}.$$

There are $\binom{n}{k}$ such cliques, thus the probability that **any** clique is monochromatic is bounded by

$$\binom{n}{k} 2 \times 2^{-\binom{k}{2}} \leq \frac{n^k}{k!} 2 \times 2^{-\binom{k}{2}}$$

$$\leq 2^{(k)^2/2 - k(k-1)/2 + 1} \frac{1}{k!} < 1.$$

$$= 2^{k/2} + 1/k! < 1$$

Thus, there is a coloring with the required property.
When $n = 1000 \leq 2^{10} = 2^{k/2}$ we get that there exists a 2-colouring of K_{1000} with no monochromatics K_{20}.

Sample and Modify

An *independent set* in a graph G is a set of vertices with no edges between them.
Finding the largest independent set in a graph is an NP-hard problem.

Theorem

Let $G = (V, E)$ be a graph on n vertices with $dn/2$ edges. Then G has an independent set with at least $n/2d$ vertices.

Algorithm:

1. Delete each vertex of G (together with its incident edges) independently with probability $1 - 1/d$.
2. For each remaining edge, remove it and one of its adjacent vertices.

$X =$ number of vertices that survive the first step of the algorithm.

$$E[X] = \frac{n}{d}.$$

$Y =$ number of edges that survive the first step.
An edge survives if and only if its two adjacent vertices survive.

$$E[Y] = \frac{nd}{2}\left(\frac{1}{d}\right)^2 = \frac{n}{2d}.$$

The second step of the algorithm removes all the remaining edges, and at most Y vertices.
Size of output independent set:

$$E[X - Y] = \frac{n}{d} - \frac{n}{2d} = \frac{n}{2d}.$$

Conditional Expectation

Definition

$$E[Y \mid Z = z] = \sum_{y} y \Pr(Y = y \mid Z = z),$$

where the summation is over all y in the range of Y.

Lemma

For any random variables X and Y,

$$E[X] = \sum_{y} \Pr(Y = y) E[X \mid Y = y],$$

where the sum is over all values in the range of Y.

Derandomization using Conditional Expectations

Given a graph $G = (V, E)$ with n vertices and m edges, we showed that there is a partition of V into A and B such that at least $m/2$ edges connect A to B.
How do we find such a partition?

$C(A, B) =$ number of edges connecting A to B.
If A, B is a random partition $E[C(A, B)] = \frac{m}{2}$.
Algorithm:

1. Let $v_1, v_2, \dots, v_n$ be an arbitrary enumeration of the vertices.
2. Let x_i be the set where v_i is placed ($x_i \in \{A, B\}$).
3. For $i = 1$ to n do:
    1. Place v_i such that

$$E[C(A, B) \mid x_1, x_2, \dots, x_i]$$
$$\geq E[C(A, B) \mid x_1, x_2, \dots, x_{i-1}] \geq m/2.$$

Lemma

For all $i = 1, \ldots, n$ *there is an assignment of* v_i *such that*

$$E[C(A,B) \mid x_1, x_2, \ldots, x_i]$$
$$\geq E[C(A,B) \mid x_1, x_2, \ldots, x_{i-1}] \geq m/2.$$

Proof.

By induction on i.
For $i = 1$, $E[C(A,B) \mid x_1] = E[C(A,B)] = m/2$
For $i > 1$, if we place v_i randomly in one of the two sets,

$$
\begin{aligned}
& E[C(A,B) \mid x_1, x_2, \ldots, x_{i-1}] \\
= \ & \frac{1}{2} E[C(A,B) \mid x_1, x_2, \ldots, x_i = A] \\
& + \frac{1}{2} E[C(A,B) \mid x_1, x_2, \ldots, x_i = B].
\end{aligned}
$$

$$
\begin{aligned}
& \max(E[C(A,B) \mid x_1, x_2, \ldots, x_i = A], \\
& E[C(A,B) \mid x_1, x_2, \ldots, x_i = B]) \\
\geq \ & E[C(A,B) \mid x_1, x_2, \ldots, x_{i-1}] \\
\geq \ & m/2
\end{aligned}
$$

How do we compute

$$\begin{aligned} & \max(E[C(A,B) \mid x_1, x_2, \ldots, x_i = A], \\ & E[C(A,B) \mid x_1, x_2, \ldots, x_i = B]) \\ \geq \quad & E[C(A,B) \mid x_1, x_2, \ldots, x_{i-1}] \end{aligned}$$

We just need to consider edges between v_i and $v_1, \ldots, v_{i-1}$.
Simple Algorithm:

1. Place v_1 arbitrarily.
2. For $i = 2$ to n do
    1. Place v_i in the set with smaller number of neighbors.

Dense graphs with no short cycles

Theorem

For every integer $k \geq 3$ there exists a graph G with n vertices, at least $\frac{1}{4}n^{1+\frac{1}{k}}$ edges and no cycle of length less than k.

Proof: Consider a random graph $G \in \mathcal{G}_{n,p}$ with $p = n^{\frac{1}{k}-1}$ and let the random variable X denote the number of edges in the graph. Then

$$\begin{aligned}
\mathbf{E}[X] &= p\binom{n}{2} \\
&= n^{\frac{1}{k}-1}\frac{1}{2}n(n-1) \\
&= \frac{1}{2}\left(1-\frac{1}{n}\right)n^{\frac{1}{k}+1}
\end{aligned}$$

Dense graphs with no short cycles

Let Y be the random variable whose value (for the given graph G) is number of cycles of length at most $k-1$ in G.
Each i-cycle occurs with probability p^i and there are $\binom{n}{i}\frac{(i-1)!}{2}$ possible cycles of length i. Thus

$$\begin{aligned}\mathbf{E}[Y] = \sum_{i=1}^{k-1}\binom{n}{i}\frac{(i-1)!}{2}p^i &\leq \sum_{i=1}^{k-1} n^i p^i \\ &= \sum_{i=1}^{k-1} n^{\frac{i}{k}} \\ &< kn^{\frac{k-1}{k}}\end{aligned}$$

Dense graphs with no short cycles

Hence

$$\begin{aligned}\mathbf{E}[X-Y] &\geq \frac{1}{2}\left(1-\frac{1}{n}\right)n^{\frac{1}{k}+1} - kn^{\frac{k-1}{k}} \\ &\geq \frac{1}{4}n^{\frac{1}{k}+1}\end{aligned}$$

So, if we delete one edge from every cycle of length at most $k-1$ in G the expected number of edges in the resulting graph G' is at least $\frac{1}{4}n^{\frac{1}{k}+1}$. This means that there exists a graph that has at least $\frac{1}{4}n^{\frac{1}{k}+1}$ and no cycles with less than k vertices.

High chromatic number and no triangles

The **Chromatic** number, $\chi(G)$ of a graph $G = (V, E)$ is the minimum integer k so that we can partition V into disjoint sets $V_1, V_2, \ldots, V_k$ with the property that no edge is inside any V_i.

Theorem

For every $k \geq 1$ there exists a graph with no clique of size 3 (triangle-free) and chromatic number at least k.

High chromatic number and no triangles

Proof Let $G \in \mathcal{G}_{n,p}$ where $p = n^{-\frac{2}{3}}$
To prove that $\chi(G) > k$ it suffices to show that G has no independent set of size $\lceil \frac{n}{k} \rceil$. In fact we prove that with high probability G no has independent set of suze $\lceil \frac{n}{2k} \rceil$.
Let the random variable I count the number of independent sets of size $\lceil \frac{n}{2k} \rceil$ in G. Let $\mathcal{S}$ be the set of all $S \subseteq V$ of size $\lceil \frac{n}{2k} \rceil$. Let the indicator variable I_S be one if S is an independent set and 0 otherwise. So $I = \sum_{\{S \in \mathcal{S}\}} I_S$.

Then we have $E[I_s] = (1-p)^{\binom{\lceil \frac{n}{2k} \rceil}{2}}$

High chromatic number and no triangles

$$
\begin{aligned}
\mathbf{E}[I] &= \sum_{\{S \in \mathcal{S}\}} \mathbf{E}[I_s] \\
&= \binom{n}{\lceil \frac{n}{2k} \rceil} (1-p)^{\binom{\lceil \frac{n}{2k} \rceil}{2}} \\
&< \binom{n}{\lceil \frac{n}{2k} \rceil} (1-p)^{\binom{\frac{n}{2k}}{2}}
\end{aligned}
$$

Using that $\binom{n}{r} \leq 2^n$ for all $0 \leq r \leq n$ and $1 - x < e^{-x}$ when $x > 0$, we get

$$
\begin{aligned}
\mathbf{E}[I] &< 2^n e^{-\frac{pn(n-2k)}{8k^2}} \\
&< 2^n e^{-\frac{n^{\frac{4}{3}}}{16k^2}} \\
&< \frac{1}{2},
\end{aligned}
$$

when $n > 2^{12}k^6$

High chromatic number and no triangles

When $n \geq 2^{12}k^6$ we have $E[I] < \frac{1}{2}$.
By Markov's inequality $Pr(I > 0) < \frac{1}{2}$ when $n \geq 2^{12}k^6$.
Let T be the number of triangles in G. Now we need to show that $E[T]$ is also much less than one, BUT that is not true!

$$\mathbf{E}[T] = \binom{n}{3} p^3 < \frac{n^3}{3!}(n^{-\frac{2}{3}})^3 = \frac{n}{6} \tag{1}$$

High chromatic number and no triangles

We found that $E[T] = \frac{n}{6}$.

By Markov's inequality, $Pr(T \geq \frac{n}{2}) \leq \frac{\frac{n}{6}}{\frac{n}{2}} = \frac{1}{3}$ for large n

Now we have $Pr(I \geq 1) + Pr(T \geq \frac{n}{2}) < \frac{1}{2} + \frac{1}{3} < 1$ so there exists a graph G with $I = 0$ and $T \leq \frac{n}{2}$.

High chromatic number and no triangles

Choose a set M of at most $\frac{n}{2}$ vertices which meets all triangles in g and let $G' = G - M$.
Then G' is triangle-free and has at least $\frac{n}{2}$ vertices. Also G' has no independent set of size $\lceil \frac{n}{2k} \rceil$ (because G has no such set) so $\chi(G') > \frac{\frac{n}{2}}{\frac{n}{2k}} = k$.

Randomization as a Resource

Complexity is usually studied in terms of resources, TIME and SPACE.

We add a new resource, RANDOMNESS, measured by the number of independent random bits used by the algorithm (= the entropy of the random source).

Example: Packet Routing

We proved:

Theorem

There is an algorithm for permutation routing on an $N = 2^n$-cube that uses a total of $O(nN)$ random bits and terminates with high probability in cn steps, for some constant c.

Can we achieve the same result with fewer random bits?

Theorem

There is an algorithm for permutation routing on an $N = 2^n$-cube that uses a total of $O(n)$ random bits and terminates with high probability in cn steps, for some constant c.

Proof

Let $A(X)$ be a randomized algorithm with input x that uses (up to) s random bits.
Let $A(x, r)$ be the execution of algorithm A with input x and a fixed sequence r on s bits.
We can write $A(X)$ as

1. Choose r uniformly at random in $[0, 2^s - 1]$.
2. Run $A(X, r)$.

In the two phase routing algorithm $s = \log(N^N) = nN$ (it chooses a random destination independently for each packet).
Let $\mathcal{B} = \{B_1, \ldots, B_r\}$ be the a collection of 2^s deterministic algorithms $A(I, r)$.
We proved:

Lemma

For a given input permutation π and a deterministic algorithm B_i chosen uniformly at random from $\mathcal{B}$, the probability that B_i fails to route π in cn steps is bounded by $1/N$.

Choose a random set $\mathcal{D} = \{D_1, \ldots, D_{N^3}\}$ of N^3 elements in $\mathcal{B}$. Let $X_i^\pi = 1$ if algorithm D_i does NOT route permutation π in cn steps, else $X_i^\pi = 0$

$$E[\sum_{i=1}^{N^3} X_i^\pi] \leq N^2$$

$$Prob(\sum_{i=1}^{N^3} X_i^\pi \geq 2N^2) \leq e^{-N^2/3}$$

$$Prob(\exists\pi \quad \sum_{i=1}^{N^3} X_i^\pi \geq 2N^2) \leq N! e^{-N^2/3} < 1$$

$$Prob(\exists\pi, \quad \sum_{i=1}^{N^3} X_i^{\pi} \geq 2N^2) \leq N! e^{-N^2/3} < 1$$

Theorem

There exists a set $\mathcal{D}$ of N^3 deterministic algorithms, such that for any given permutation π and an algorithm D chosen uniformly at random from $\mathcal{D}$, algorithm D routes π in cn steps with probability $1 - 1/N$. The random choice requires $O(n)$ random bits.

Can we do better?

Do we need any random bits?

Definition

A routing algorithm is **oblivious** if the path taken by one packet is independent of the source and destinations of any other packets in the system.

Theorem

Given an N-node network with maximum degree d the routing time of any deterministic oblivious routing scheme is

$$\Omega(\sqrt{\frac{N}{d^3}}).$$

Theorem

For any deterministic oblivious algorithm for permutation routing on the $N = 2^n$ cube there is an input permutation that requires $\Omega(\sqrt{N}/n^3)$ steps.

Theorem

Any randomized oblivious routing algorithm for permutation routing on the $N = 2^n$ cube must use $\Omega(n)$ random bits to route an arbitrary permutation in $O(n)$ expected time.

proof

Assume that the algorithm uses k random bits.
It can choose between no more than 2^k possible deterministic executions.
There is a deterministic execution $\tilde{A}$ that is chosen with probability $\geq 1/2^k$.
Let π be an input permutation that requires $\Omega(\sqrt{N}/n^3)$ steps in $\tilde{A}$.
The expected running time of this input permutation on the randomized algorithm is $\Omega(\sqrt{N}/(2^k n^3))$
So, if we want this to be $O(n)$ we must take k roughly $\log N = n$.