Derandomisation

How to get a deterministic algorithm from a probabilistic one?

1) Method of conditional probabilities:

Example with Ramsey numbers:

Proposition: For every n, there exist an edge colouring of K_n such that the number of monochromatic K_4 is at most $\binom{n}{4} 2^{-5}$.

Proof: X = number of monochromatic copies of K_4 in a random uniform colouring of K_n.

$$\mathbb{E}[X] = \binom{n}{4} 2 . 2^{-6} = \binom{n}{4} 2^{-5}.$$

$$\mathbb{P}\left(X \leq \binom{n}{4} 2^{-5}\right) > 0 \quad \text{by F.M.M.} \qquad \blacksquare$$

How to find such colouring deterministically? Maximise conditional E.

Consider an ordering $e_1, \dots, e_m$ of the edges of K_n
Consider a colouring α of the edges $e_1, \dots, e_{i-1}$.

Let β be a random uniform colouring extending α.

$$W_i = \mathbb{E}\left[X(\beta) \mid \beta_{|e_1, \dots e_{i-1}} = \alpha\right] = \sum_{R \in \binom{[n]}{4}} w(R)$$

where $w(R) = \begin{cases} 0 & \text{if } \alpha \text{ colours two edges of } R \neq. \\ 2^{-5} & \text{if no edge of } R \text{ is coloured in } \alpha. \\ 2^{r-6} & \text{if } \alpha \text{ colours } r \text{ edge of } R \text{ identically.} \end{cases}$$

Let $W_i^{(Red)} = \mathbb{E}\left[X(\beta) \mid \beta_{|e_1,\dots e_{i-1}} = \alpha \ , \ \beta(e_i) = red\right]$.

$W_i^{(Blue)} = $ ______________________________ $blue]$.

We have $W_i = \dfrac{W_i^{(red)} + W_i^{(Blue)}}{2} \leq W_{i+1} = \max\left(W_i^{(Red)}, W_i^{(blue)}\right)$

we color e_i in order to maximize W_{i+1}.

calculating W_{RED} and W_{Blue} can be done in time $O(n^4)$:

2) Pessimistic estimators

More generally, consider independent random variable $X_1, \dots X_n \in \{0,1\}$ and let $A_1, \dots A_m$ be a collection of events determined by the $X_1, \dots X_n$.

Let $k = \sum \mathbb{P}(A_i)$ the expectation of the number of A_i.

By Markov, there exists $x = (x_1, \dots x_n) \in \{0,1\}^n$ such that at most k events A_i hold at x. We look for x algorithmically. For every , $j \in [n]$, for every $x_1, \dots x_{j-1}$,

$$\sum_i \mathbb{P}(A_i \mid X_1 = x_1, \dots X_{j-1} = x_{j-1}) = \mathbb{P}(X_j = 0) \sum_i \mathbb{P}(A_i \mid X_1 = x_1, \dots X_j = 0)$$
$$+ \mathbb{P}(X_j = 1) \sum_i \mathbb{P}(A_i \mid X_1 = x_1, \dots X_j = 1)$$
$$\geq \min\left\{ \sum_i \mathbb{P}(A_i \mid X_1 = x_1, \dots X_j = 0), \sum_i \mathbb{P}(A_i \mid X_1 = x_1 \dots X_j = 1) \right)$$

So we can choose the value of each x_j progressively, keeping the conditional expectation low

For this be tractable, we need to be able to compute this conditional expectation efficiently.

What do we do when this is not the case?

Suppose that for each A_i, for each $0 \leq j \leq l$, we have a function $f_j^i(x_1, \dots x_j)$ that can be efficiently computed s.t.

$$f_{j-1}^i(x_1, \dots x_{j-1}) \geqslant P(X_j = 0)\, f_j^i(x_1, \dots x_{j-1}, 0)$$
$$+ P(X_j = 1)\, f_j^i(x_1, \dots x_{j-1}, 1)$$

and that $f_j^i(x_1, \dots x_j) \geqslant P(A_i) X_1 = x_1, \dots X_j = x_j)$.

If in the beginning $\sum_{i=0}^{m} f_0^i \leq t$ then we can find efficiently $x = (x_1, \dots x_n)$ such the number of A_i that hold in x is

$$\sum_{i=0}^{n} P(A_i \mid X_1 = x_1, \dots X_n = x_n) \leq \sum_{i=0}^{m} f_n^i(x_1, \dots x_n)$$
$$\leq \sum_{i=0}^{m} f_0^i \leq t$$

Theorem: Let $A = (a_{ij})_{1 \leq i,j \leq n}$ be the $n \times n$ matrix of reals where $-1 \leq a_{ij} \leq 1$ for all i,j. Then one can find in polynomial time $(x_1, \dots x_n) \in \{0,1\}^n$ s.t. $\forall i \in [n]$,

$$\left\| A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\|_\infty = \left| \sum_{j=1}^{n} x_i a_{ij} \right| \leq \sqrt{2n \ln(2n)}.$$

Proof: Let $X_1, \dots X_n$ be uniform independent variables in $\{-1, 1\}$.

Denote $\beta = \sqrt{2m \ln(2n)}$ and $\alpha = \beta / m$.

Let A_i be the event $\left|\sum_{j=1}^{m} X_i a_{ij}\right| > \beta$.

Let $G(x) = \cosh(\alpha x) = \dfrac{e^{\alpha x} + e^{-\alpha x}}{2}$

We have $G(x) \leq e^{\frac{\alpha^2 x^2}{2}}$ (Expand the Taylor series to get $\cosh(y) \leq e^{y^2/2}$)

with equality if $x = 0$ or $\alpha = 0$.

We have $G(x)G(y) = \frac{1}{4}\left(e^{\alpha(x+y)} + e^{\alpha(x-y)} + e^{\alpha(y-x)} + e^{-\alpha(x+y)}\right)$

$$= \frac{G(x+y) + G(x-y)}{2}.$$

Let $f_p^i(x_1, \dots x_p) = 2e^{-\alpha\beta} G\left(\sum_{j=1}^{p} x_j a_{ij}\right) \prod_{j=p+1}^{m} G(a_{ij})$

1. These functions can be computed efficiently ✓

2. $\forall i \in [n]$, $\forall x_1, \dots x_{p-1} \in \{-1, 1\}$,

$$f_{p-1}^i(x_1, \dots x_{p-1}) = \frac{f_p^i(x_1, \dots, x_{p-1}, -1) + f_p^i(x_1, \dots x_{p-1}, 1)}{2}$$

3. $\forall i \in [n]$, $\forall x_1, \dots x_{p-1} \in \{-1, 1\}$,

$$f_{p-1}^i(x_1, \dots x_{p-1}) \geq \mathbb{P}(A_i \mid X_1 = x_1, \dots X_{p-1} = x_{p-1})$$

4. $\sum_{i=0}^{m} f_0^i < 1$

2. $f^i_{p-1}(x_1, \dots x_{p-1}) = 2e^{-\alpha\beta} G\left(\sum_{j=1}^{p-1} x_j a_{ij}\right) G(a_{ip}) \prod_{j=p+1}^{m} G(a_{ij})$

$$= 2e^{-\alpha\beta}\left[\frac{G\left(\sum_{j=1}^{p-1} x_j a_{ij} + a_{ip}\right) + G\left(\sum_{j=1}^{p-1} x_j a_{ij} - a_{ip}\right)}{2}\right] \prod_{j=p+1}^{m} G(a_{ij})$$

$$= \frac{f^i_p(x_1, \dots x_{p-1}, -1) + f^i_p(x_1, \dots x_{p-1}, 1)}{2}$$

3. $\mathbb{P}(A_i \mid X_1 = x_1, \dots X_p = x_{p-1}) = \mathbb{P}\left(\sum_{j=1}^{p-1} x_j a_{ij} + \sum_{j=p}^{m} X_j a_{ij} > \beta\right)$

$$+ \mathbb{P}\left(\underbrace{-\sum_{j=1}^{p-1} x_j a_{ij}}_{=v} - \sum_{j=p}^{m} X_j a_{ij} > \beta\right)$$

$$= \mathbb{P}\left(e^{\alpha(v + \sum_{j=p}^{m} X_j a_{ij})} > e^{\alpha\beta}\right) + \mathbb{P}\left(e^{-\alpha(v + \sum X_j a_{ij})} > e^{\alpha\beta}\right)$$

$$\underset{\text{Markov}}{\leq} e^{\alpha v} e^{-\alpha\beta} \mathbb{E}\left[e^{\alpha \sum_{j=p}^{m} X_j a_{ij}}\right] + e^{-\alpha v} e^{-\alpha\beta} \mathbb{E}\left[e^{-\alpha \sum_{j=p}^{m} X_j a_{ij}}\right]$$

$$= 2e^{-\alpha\beta} G(v) \prod_{j \geq p} G(a_{ij}) = f^i_{p-1}(x_1, \dots x_{p-1}).$$

4. $\sum_{i=1}^{m} f^i_0 = 2e^{-\alpha\beta} \sum_{i=1}^{m} \prod_{j=1}^{m} G(a_{ij}) \underset{(1)}{\leq} \sum_{i=1}^{m} 2e^{-\alpha\beta} \prod_{j=1}^{m} e^{\alpha^2 a_{ij}^2/2}$

$$\underset{(2)}{\leq} \sum_{i=1}^{m} 2e^{-\alpha\beta} e^{\alpha^2 m/2}$$

because $\alpha = \beta/m$ $\longrightarrow$ $= 2m\, e^{\alpha^2 m/2 - \alpha\beta}$

$= 2m e^{-\alpha^2 m/2}$

because $\alpha = \sqrt{2m \ln(2m)}$ $\longrightarrow$ $= 1$

(1) is strict unless all $a_{ij} = 0$

(2) is strict unless all $|a_{ij}| = 1$.

so $\sum_{i=1}^{m} f^i_0 < 1$. ∎

3) Method of small probability space.

Suppose we have an algorithm that uses $O(\log n)$ random bits on instances of size n.

Then the probability space has size $n^{O(1)}$ an one can in polynomial time go through all possibilities.

Idea: → reduce the number of random bits
→ reduce the size of the probability space by taking d-wise independent r.v. instead of mutually independent

$X_1, \dots X_n$ are d-wise independent if $\forall I \subseteq [n]$ with $|I| \leq d$, $(x_i)_{i \in I}$, $\mathbb{P}\left(\bigcap_{i \in I} \{X_i = x_i\}\right) = \prod_{i \in I} \mathbb{P}(X_i = x_i)$.

Example:

	s	X_1	X_2	X_3
size of the probability space	00	0	0	0
	01	0	1	1
	10	1	0	1
	11	1	1	0

are pairwise independent ($d = 2$).

$X = (X_1, X_2, X_3)$ has support only four elements of $\{0,1\}^{[3]}$ instead of eight for the uniform distribution on $\{0,1\}^{[3]}$.

Application: to find a 2-coloring of K_n with at most $\binom{n}{4} 2^{-5}$ monochromatic K_4, we don't need mutual independence.

Let $X_1, \dots X_{\binom{n}{2}}$ be a family of 6-wise independent uniform $\{0,1\}$ random variables.

Consider a random coloring of K_n, where the i^{th} edge of K_n is colored with X_i.

$$\mathbb{E}[\#\text{ monochromatic } K_4] \underset{L.E}{=} \sum_{R \in \binom{[n]}{4}} P(R \text{ is monochromatic})$$

$$\underset{\text{by 6-wise indep}}{=} \sum_{\binom{[n]}{4}} 2 \cdot \frac{1}{2^{-6}} = \binom{n}{4} 2^{-5}$$

Can we find such variables in a small probability space?

Theorem: Suppose $n = 2^k - 1$ and $d = 2t+1$.
There exists a symmetric probability space Ω of size $2(n+1)^t$ and d-wise independent uniform $\{0,1\}$ random variables $X_1, \dots X_n$.
(The space and the variables can be constructed).

Proof: Consider $F_m = GF(2^k)$ the Galois Field of size 2^k.
Let $x_1, \dots x_n$ be the n non-zero elements of F_m, represented as columns vectors of length k over $GF(2)^k$.

Let H be the $(1+kt) \times n$ matrix over $GF(2)$.

$$H = \underbrace{\begin{pmatrix} 1 & \text{———} & 1 \\ x_1^1 & \text{———} & x_n^1 \\ x_1^3 & \text{———} & x_n^3 \\ \vdots & & \\ x_1^{2t-1} & \text{———} & x_n^{2t-1} \end{pmatrix}}_{n} \Big\updownarrow 1+kt$$

(BCH : Bose - Chaudhuri - Hocquenghem code : error correcting code)

Lemma: Any set of $d = 2t+1$ columns of H is linearly independent over $GF(2)$.

Proof: Let $J \subseteq [n]$ of size $2t+1$.
Denote H_j the columns of H.
Suppose $\sum_{j \in J} z_j H_j = 0$ where $z_j \in \{0,1\} = GF(2)$.

$$(*) \qquad \sum_{j \in J} z_j x_j^i = 0 \qquad \text{for } i = 0 \text{ and } 1 \leq i \leq 2t-1 \text{ odd}.$$

Let $\Omega = [2(n+1)^t]$ and let $(A_{ij})_{\substack{i \in \Omega \\ j \in [n]}}$ be the $\{0,1\}$ matrix whose $2(n+1)^t = 2^{kt+1}$ rows are all the linear combinations over $GF(2)$ of the rows of H.

Consider the uniform probability measure and let X_j be the random variable defined by $X_j(i) = a_{ij}$ $\forall i \in \Omega$.

Let $J \subseteq [n]$ of size at most d. Let H_J and A_J be the submatrices of H and A formed by the rows indexed with $j \in J$.

Let $x_J \in \{0,1\}^J$ and $X_J = (X_j)_{j \in J}$.

$\mathbb{P}(X_J = x_J)$ is the proportion of linear combinations of rows of A_J that sum up to x_J.

The number of such rows is the number of solutions of a system of $|J|$ linearly independent equations with $kt+1$ variables, i.e. the number of free variables: $2^{kt+1-|J|}$

so $\mathbb{P}(X_J = x_J) = 2^{-|J|}$.

$\blacksquare$

This gives a way of constructing a sample space of size $O(n^{\lfloor d/2 \rfloor})$ of d-wise independent uniform r.v. (instead of 2). this is optimal $\Omega_d(n^{\lfloor d/2 \rfloor})$

The class NC all problems that can be decided in polylogarithmic time by a Turing machine with a polynomial number of processors.

Equivalently: problems decidable by a uniform boolean circuit with depth polylogarithmic and a polynomial number of gates of maximum arity 2.

Application: derandomisation of thm 1:

Construct with an NC-algorithm a 2-edge-colouring of K_n with at most $\binom{n}{4} 2^{-5}$ monochromatic K_4.

$\rightarrow$ number of variables: $m = \binom{n}{2}$, take k s.t.

$$m = \binom{n}{2} \leq 2^k - 1 < 2\binom{n}{2} \qquad \text{i.e. } k = \lceil \log(m+1) \rceil$$

$\rightarrow$ $d \geq 6$. Take $t = 3$, we have $d = 7$.

By the previous theorem, we construct m 7-wise independent uniform random variables on a probability space of size at most $2(\lceil \log(m+1) \rceil + 1)^3 = O((\log n)^c)$

and count the number of monochromatic K_4 in $\|$ for each of them.

$(X_1, \dots X_n)$ are hard to distinguish from random uniform.
link with hash table.