10. Derandomisation

- Suppose we have a (finite) probability space (Ω, Pr) and a probabilistic proof that for some interesting event $A \subseteq \Omega$, $Pr(A) > 0$ [or even $Pr(A) \geq \frac{1}{2}$ or $Pr(A) \geq 1 - \varepsilon$ for some $\varepsilon \to 0$]. [$Pr(A) = Pr_\omega(\omega \in A)$]

  When can we actually find, i.e. construct by an efficient algorithm, a sample point $\omega \in A$? This is a fundamental, challenging, and relatively little understood question.

- Note that if we have an efficient method for sampling Ω according to Pr, and $Pr(A) > \delta$ for some $\delta > 0$, we get an efficient randomised method for finding a good ω by just repeated independent sampling: the probability of k consequent bad ω's is then $\leq (1-\delta)^k \leq e^{-\delta k}$. [Of course this assumes that we also have an efficient test for whether a given $\omega \in \Omega$ satisfies $\omega \in A$.]

- But the general issue is when and how can such randomised algorithms be made deterministic, or "derandomised", efficiently.

10.1 The Method of Conditional Probabilities

- Assume for simplicity that $\Omega = \{0,1\}^n$ and Pr is uniform on Ω, so that for each binary string $\omega = (\varepsilon_1, \ldots, \varepsilon_n)$, $Pr(\omega) = 2^{-n}$.

- Suppose the event of interest is of the form A = "$X \geq k$", where $X = X(\omega)$ is some random variable; and moreover we have a proof that $E[X] \geq k$ (and hence $Pr(A) > 0$).

- A sample point (string) $\omega^* \in \{0,1\}^n$ satisfying $X(\omega^*) \geq k$ can now be "traced out" as follows: introduce binary r.v.'s $Y_1, \ldots, Y_n$ corresponding to the bits of ω and taking values 0/1 with prob. $1/2$. Then e.g.

$$E[X] = \tfrac{1}{2} E[X \mid Y_1 = 0] + \tfrac{1}{2} E[X \mid Y_1 = 1]$$

and, for general $(\varepsilon_1, \ldots, \varepsilon_j) \in \{0,1\}^j$:

$$\begin{aligned} E[X \mid \varepsilon_1, \ldots, \varepsilon_j] &= \tfrac{1}{2}\left(E[X \mid \varepsilon_1, \ldots, \varepsilon_j, 0] + E[X \mid \varepsilon_1, \ldots, \varepsilon_j, 1]\right) \\ &\vdots \\ &\leq \max\left\{E[X \mid \varepsilon_1, \ldots, \varepsilon_j, 0],\ E[X \mid \varepsilon_1, \ldots, \varepsilon_j, 1]\right\} \end{aligned}$$

- Thus, if one for any $(\varepsilon_1^*, \ldots, \varepsilon_j^*)$ could choose the $\varepsilon_{j+1}^* \in \{0,1\}$ maximising the conditional expectation as above, the desired ω^* could be constructed as:

$$k \leq E[X] \leq E[X \mid \varepsilon_1^*] \leq E[X \mid \varepsilon_1^*, \varepsilon_2^*] \leq \cdots$$

$$\leq E[X \mid \varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_n^*] = X(\omega^*).$$

- Typically, $X = X_1 + \cdots + X_m$, where the X_i are indicators for some subevents A_i, and

$A \sim$ "at least k of $A_1, \ldots, A_m$ occur".

Then $E[X \mid \vec{\varepsilon}] = \sum_{i=1}^{m} \Pr(A_i \mid \vec{\varepsilon})$, and it is (sometimes) easy to determine which of the sums

$$\sum_i \Pr(A_i \mid \vec{\varepsilon}, 0) \quad , \quad \sum_i \Pr(A_i \mid \vec{\varepsilon}, 1)$$

is bigger.

- Example 1. Splitting graphs.

Recall Thm 2.2 (p. 12): In any graph $G = (V, E)$ with $V = [n]$ and m edges, the vertices can be partitioned into $V = T \cup (V \setminus T)$, so that the number of edges crossing the cut $(T, V \setminus T)$ is at least $m/2$.

The proof considers a random cut T, and cross-edge indicator variables

$$X_e = \begin{cases} 1, & \text{if } e = \{i,j\} \text{ crosses } T \\ 0, & \text{otherwise.} \end{cases}$$

Then $E[X_e] = \Pr(i \in T, j \notin T \text{ or } i \notin T, j \in T)$

$$= \Pr(i \in T) \cdot \Pr(j \notin T) + \Pr(i \notin T) \cdot \Pr(j \in T)$$

$$= 2 \cdot \tfrac{1}{4} = \tfrac{1}{2},$$

and for $X = \sum_{e \in E} X_e$,

$$E[X] = \sum_e E[X_e] = m \cdot \tfrac{1}{2}$$

- Now a cut T^* satisfying $X(T^*) \geq \frac{m}{2}$ can be constructed sequentially as follows:
  - denote $T^*_j = T^* \cap \{1, \dots, j\}$; thus $T^*_0 = \emptyset$, $T^*_n = T^*$
  - to obtain T^*_j from T^*_{j-1}, one needs to maximise between:

$$\sum_e \Pr(X_e = 1 \mid T^*_{j-1}), \quad \sum_e \Pr(X_e = 1 \mid T^*_{j-1} \cup \{j\})$$

  - note, however, that the sums differ only at the edges e that contain j as one of their endpoints, and even then only if the other endpoint is some $i < j$.
  - thus, the optimal choice is to put vertex j on the opposite side of the cut from the majority of its neighbours among vertices $\{1, \dots, j-1\}$.
  - this iterative process eventually guarantees $X(T^*) \geq \frac{m}{2}$.

• Example 2 Shader-colourings

Consider a special case (for simplicity) of Thm 2.3 (p.13): For any n, there is a two-colouring of the edges of K_n that contains at most

$$\binom{n}{3} \cdot 2^{-2} = \frac{1}{4}\binom{n}{3}$$

monochromatic triangles.

The proof considers a random red/blue-colouring ω of the edges of K_n, and associates to each triangle K in K_n an indicator $X_K \sim$ "K is monochromatic under ω". Then for $X = \sum_K X_K$,

$$E[X] = \sum_K E[X_K] = \binom{n}{3} \cdot 2 \cdot 2^{-3} = \frac{1}{4}\binom{n}{3}.$$

• To derandomise this proof (note that we are minimising this time), order the edges of K_n as $e_1, \dots, e_m$, and consider a partial colouring ω_j^* of the edges $e_1, \dots, e_j$.

Then for each triangle K in K_n,

$$\Pr(X_K = 1 \mid \omega_j^*) = \begin{cases} 0, & \text{if } \omega_j^* \text{ colours two edges in } K \text{ by different colours,} \\ 2 \cdot 2^{-3}, & \text{if no edges in } K \text{ are yet coloured by } \omega_j^*, \\ 2^{r-3}, & \text{if } \omega_j^* \text{ colours } r = 1,2,3 \text{ edges in } K \text{ by the same colour} \end{cases}$$

• Thus, to determine the choice of colour for edge e_j given partial colouring ω_{j-1}^*, one needs to consider all the (at most $n-2$) triangles in which edge e_j participates, and see which choice minimises the sum

$$\sum_{K \ni e_j} \Pr(X_K = 1 \mid \omega_{j-1}^*, \omega(e_j) = \text{red/blue}).$$

10.2 Small Sample Spaces

- Consider again the sample space $\Omega = \{0,1\}^n$, with uniform probability distribution, and an event A such that $Pr_\Omega(A) > 0$ [or $Pr_\Omega(A) \geq \frac{1}{2}$].

- Another approach to derandomising the search for a sample point satisfying $\omega \in A$ is to look for a very small sample space $\Omega' \subseteq \Omega$ with the property $Pr_{\Omega'}(A) = Pr_\Omega(A)$ [or $Pr_{\Omega'}(A) \approx Pr_\Omega(A)$].

  Then one may exhaustively search through <u>all</u> $\omega' \in \Omega'$.

- Such sample space reductions exist under some general conditions about the nature of the event A. (And it is a very interesting question how far this technique can be extended.)

- Let us say that binary random variables $Y_1, \ldots, Y_n$ are k-<u>wise independent</u> if for any $\{Y'_1, \ldots, Y'_k\} \subseteq \{Y_1, \ldots, Y_n\}$ and any $(\varepsilon_1, \ldots, \varepsilon_k) \in \{0,1\}^k$:

$$Pr(Y'_1 = \varepsilon_1, \ldots, Y'_k = \varepsilon_k) = \prod_{i=1}^{k} Pr(Y'_i = \varepsilon_i).$$

- Note that e.g. in the proof of Thm 2.2 (cf. p. 79), it was not necessary that the choices of which vertices to put in the random cut T [i.e. events "$i \in T$"] were fully independent, but just <u>pairwise</u> independent: for any <u>two</u> $i, j \in V$:

$$Pr(i \in T,\ j \notin T) = Pr(i \in T) \cdot Pr(j \notin T) = \tfrac{1}{4} \quad \text{and}$$
$$Pr(i \notin T,\ j \in T) = Pr(i \notin T) \cdot Pr(j \in T) = \tfrac{1}{4}.$$

- Thus, the expectations $E[X_e]$ etc would evaluate to the same even if the choices of "$i \in T$" were determined by a sequence of just pairwise independent binary variables $Y_1, \ldots, Y_n$.

- Now how to generate a big set of pairwise independent r.v.'s from a small sample space ("randomness amplification")

- Suppose for simplicity that $n = 2^d$ for some d, and identify $1, \ldots, n$ with the elements of the finite field $\mathbb{F}_n$ (which can in another representation be seen as d-bit vectors).

  Define for each $i \in \mathbb{F}_n \approx [n]$ a random variable

  $$Z_i = Z_i(a,b) = a \cdot i + b,$$

  where a, b are two elements of $\mathbb{F}_n$ chosen randomly and independently.

- Claim. The variables $Z_1, \ldots, Z_n$ are pairwise independent.

  Proof For all $i, j \in \mathbb{F}_n$, $x, y \in \mathbb{F}_n$:

  $$\Pr_{a,b}(Z_i = x, Z_j = y) = \Pr(a \cdot i + b = x, \; a \cdot j + b = y)$$
  $$= \Pr\left(a = \frac{x-y}{i-j}, \; b = \frac{yi - xj}{i-j}\right) = \frac{1}{n^2}$$
  $$= \Pr(Z_i = x) \cdot \Pr(Z_j = y). \quad \square$$

- Thus, a set of n pairwise independent $\log n$-bit vector variables are generated from a sample space of size n^2.

  Almost by definition, the first bits of the vector-valued Z_i are also pairwise independent, and can be used as the decision variables Y_i.

- To summarise: Thm 2.2 shows that when one considers the space of all random cuts $T \in \Omega = \mathcal{P}([n])$, then $\Pr_\Omega(X \geq \frac{m}{2}) > 0$ and hence there exists $T^* \in \Omega$ s.th. $X(T^*) \geq \frac{m}{2}$. The technique of pairwise independence shows that in fact it suffices to consider cuts generated as above by random pairs of elements $a, b \in \mathbb{F}_n$, of which there are only n^2.