Randomized Algorithms

- Algorithm can use randomness.
- Output correct with high probability.
- Often simple, efficient, only known ones.

Examples
- Randomized quicksort
- Hashing
- Primality testing
- Distributed computing

Basics of Discrete Probability

Def A discrete prob. space is a pair $(\Omega, \Pr)$

where
- Ω is a finite set (of "outcomes")
- $\Pr : \Omega \to [0,1]$ s.t.

$$\sum_{\omega \in \Omega} \Pr[\omega] = 1.$$

Note
- $\Pr[\omega]$ is the prob. of outcome ω.
- Typically assume $\Pr[\omega] > 0 \ \forall \omega \in \Omega$.

Examples

- $\Omega = \{\text{Heads}, \text{Tails}\}$. $\quad Pr[\text{Heads}] = \frac{1}{2}$
  $Pr[\text{Tails}] = \frac{1}{2}$.

- Biased coin toss.

  $\Omega = \{H, T\}$. $\quad Pr[H] = \frac{2}{3}, \; Pr[T] = \frac{1}{3}$.

- k Independent tosses of unbiased coin.

  $\Omega = \{H, T\}^k$. $\quad Pr[\omega] = \frac{1}{2^k} \;\; \forall \omega \in \Omega$.

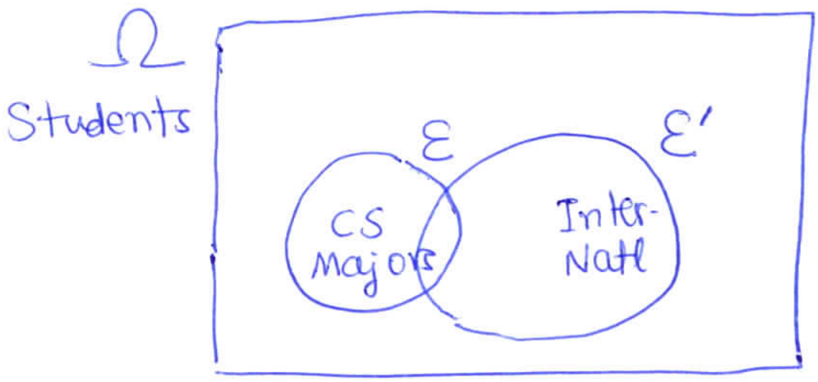
- Roll of a dice.

  $\Omega = \{1, 2, \dots, 6\}$. $\quad Pr[\omega] = \frac{1}{6} \;\; \forall \omega \in \Omega$.

Def Uniform prob. space.

$$Pr[\omega] = \frac{1}{|\Omega|} \quad \forall \omega \in \Omega.$$

Def An event $\mathcal{E}$ in a prob space $(\Omega, \Pr)$ is any subset $\mathcal{E} \subseteq \Omega$.



Def $$\Pr[\mathcal{E}] = \sum_{\omega \in \Omega} \Pr[\omega].$$

- Understanding events & their probabilities.
- Tools: Union bound, independence, random vars, expectation etc.

Example $\Omega = \{H, T\}^{10}$.

$\mathcal{E}$ = Event that #heads = 6.

$$\Pr[\mathcal{E}] = \frac{\binom{10}{6}}{2^{10}}.$$

Fact $\Pr[\bar{\mathcal{E}}] = 1 - \Pr[\mathcal{E}]$.

Union bound

Fact If $\mathcal{E}_1, \mathcal{E}_2$ are events in $(\Omega, \Pr)$

then $\Pr[\mathcal{E}_1 \cup \mathcal{E}_2] \leq \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2]$.

Proof

$$\Pr[\mathcal{E}_1 \cup \mathcal{E}_2] = \sum_{\omega \in \mathcal{E}_1 \cup \mathcal{E}_2} \Pr[\omega]$$

$$\leq \sum_{\omega \in \mathcal{E}_1} \Pr[\omega] + \sum_{\omega \in \mathcal{E}_2} \Pr[\omega]$$

$$= \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2].$$

$\blacksquare$

Note Equality above if $\mathcal{E}_1 \cap \mathcal{E}_2 = \phi$.

Fact $\Pr[\mathcal{E}_1 \cup \mathcal{E}_2 \cup \dots \cup \mathcal{E}_n] \leq \sum_{i=1}^{n} \Pr[\mathcal{E}_i]$.

$\blacksquare$

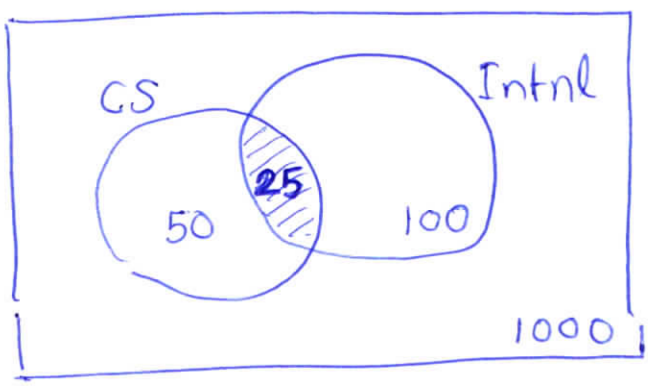
Def Events A, B in (Ω, Pr) are independent if

- $Pr[A \cap B] = Pr[A] \cdot Pr[B]$.

eq. $Pr[A \mid B] \overset{def}{=} \dfrac{Pr[A \cap B]}{Pr[B]} = Pr[A]$.

Note $Pr[A \mid B]$ is "conditional probability of event A given event B".

① Students



$Pr[\text{Intnl}] = \frac{1}{10}$.

$Pr[\text{Intnl} \mid CS] = \frac{1}{2}$.

So these are dependent events.

② $\Omega = \{H,T\} \times \{H,T\}$. $\qquad Pr[\omega] = \frac{1}{4} \ \forall \omega \in \Omega$.

A = First toss is H. $\qquad Pr[A] = \frac{1}{2}$.

B = Second toss is H. $\qquad Pr[B] = \frac{1}{2}$.

$Pr[A \cap B] = \frac{1}{4}$. $\qquad A \cap B = \{(H,H)\}$.

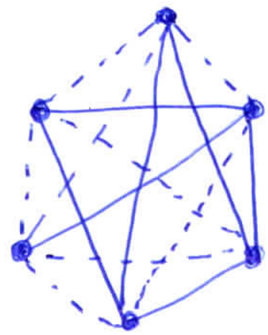
$\Omega = \{(H,H), (H,T), (T,H), (T,T)\}$

$$Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]} = \frac{1/4}{1/2} = \frac{1}{2} = Pr[A].$$

So these are independent events.

Ramsey Numbers

G_n = Complete graph on n vertices.



Fact Any red-blue coloring of edges of G_6 contains a monochromatic triangle.

Theorem For any integer $k \geq 3$, there is an integer n s.t. any red-blue coloring of edges of G_n contains a monochromatic copy of G_k.

Let $R(k)$ be minimum such n.

Known

$R(3) = 6$.

$R(4) = 18$.

$R(5)$ = Unknown.

$$R(k) \leq \binom{2k-2}{k-1}.$$

Theorem $R(k) > \lfloor 2^{k/2} \rfloor$.

I.e. there exists a red-blue coloring of edges of G_n, $n = \lfloor 2^{k/2} \rfloor$, s.t. there is no monochromatic copy of G_k.

Proof - "Probabilistic method"!

- To show that an object with property P exists, construct the object randomly and show that it has the property P with positive probability.

Proof Let $n = \lfloor 2^{k/2} \rfloor$. Color edges of G_n red/blue uniformly & independently.

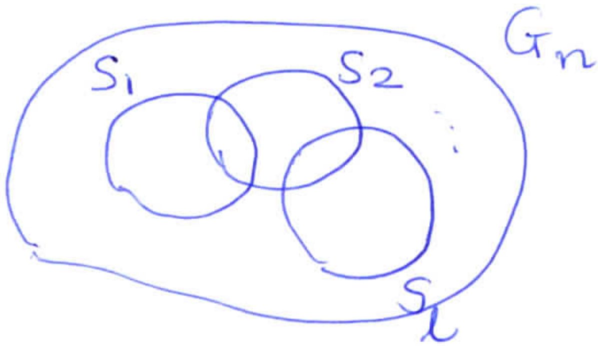
Let $\mathcal{E}$ = Event that there is no monochromatic copy of G_k.

We'll show that $Pr[\mathcal{E}] > 0$.

More convenient to work with

$B = \overline{\mathcal{E}}$ and show $\Pr[B] < 1$.

- B = Event that there exists a monochr copy of G_k.



$|S_i| = k \ \forall i.$

Let $S_1, S_2, \ldots, S_\ell$ be all k-subsets. $\ell = \binom{n}{k}$.

Let A_i = Event that S_i is monochromatic.

$\therefore \ B = \bigcup_{i=1}^{\ell} A_i.$ $\qquad \Pr[A_i] = \dfrac{2}{2^{\binom{k}{2}}}$,

$\therefore \ \Pr[B] \leq \sum_{i=1}^{\ell} \Pr[A_i]$ union bound

$$= \ell \cdot \frac{2}{2^{\binom{k}{2}}}$$

$$\leq \binom{n}{k} \cdot 2 \cdot 2^{-\binom{k}{2}}$$

$$\Pr[B] \leq \frac{n^k}{k!} \cdot 2 \cdot 2^{-k(k-1)/2}$$

$$\leq \frac{2 \cdot 2^{k/2}}{k!} \cdot n^k \cdot 2^{-k^2/2}$$

$$< n^k \cdot 2^{-k^2/2} \qquad k \geq 3$$

$$< 1 \qquad \text{provided} \quad n = \lfloor 2^{k/2} \rfloor.$$

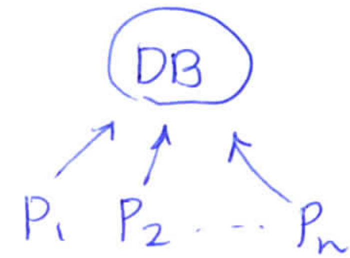
$\blacksquare$

<u>Note</u>

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

- $k!$ is $2^{k \log k - \Theta(k)}$.

Contention Resolution



- n processes $P_1, P_2, \ldots, P_n$ want to access a database.
- Time steps $t = 1, 2, 3, \ldots$
- At each time step, each process decides whether to send in a request.
- If exactly one process sends a request, that process gets the access. O/W none does.

Problem
- Design a protocol s.t. each process gets access "in reasonable time".
- No collaboration allowed among processes.

Protocol Let $p = \frac{1}{n}$.

- At each time step t, each process sends a request w.p. p independently.

Let $A_{i,t}$ = Event that P_i sends req. at time step t.

$S_{i,t}$ = Event that P_i "wins" at step t (i.e. it is the only one requesting).

$$\Pr[S_{i,t}] = \Pr[A_{i,t} \bigcap_{j \neq i} \overline{A}_{j,t}]$$

$$= \Pr[A_{i,t}] \cdot \prod_{j \neq i} \Pr[\overline{A}_{j,t}]$$

$$= p \cdot (1-p)^{n-1}$$

$$= p\left(1 - \frac{1}{n}\right)^{n-1}$$

$n = 2$ $\quad$ $1/2$

$n = 3$ $\quad$ $4/9$

$\downarrow$ $\quad$ $\downarrow$

$1/e$

$$= \beta \cdot p \quad \text{for some } \beta \in [1/e, 1/2].$$

$\therefore \Pr[S_{i,t}] = \Theta\left(\frac{1}{n}\right)$.

$e \approx 2.71$

rounds $= T$.

Let F_i = Event that P_i succeeds at least once in rounds $1, 2, \ldots, T$.

$$\Pr[\bar{F}_i] = (1 - \Pr[S_{i,t}])^T$$

$$\leqslant \left(1 - \frac{1}{e} \cdot \frac{1}{n}\right)^T$$

If $T = \lceil en \rceil$

$$\leqslant \frac{1}{e}.$$

$$\left(1 - \frac{1}{x}\right)^x \underset{x \to \infty}{\uparrow} \frac{1}{e}$$

Suppose $T = \lceil 2en \ln n \rceil$. Then

$$\Pr[\bar{F}_i] \leqslant \left(1 - \frac{1}{en}\right)^{en \cdot 2 \ln n}$$

$$\leqslant (1/e)^{2 \ln n}$$

$$= \frac{1}{n^2}.$$

Let $F = \bigcap_{i=1}^{n} F_i$ = Event that each process succeeds at least once.

$$\therefore \quad \overline{F} = \bigcup_{i=1}^{n} \overline{F}_i .$$

$$\therefore \quad \Pr[\overline{F}] \leq \sum_{i=1}^{n} \Pr[\overline{F}_i]$$

$$\leq n \cdot \frac{1}{n^2}$$

$$= \frac{1}{n}.$$

$\therefore \quad \Pr[F] \geq 1 - \frac{1}{n}$ provided

$$T = \lceil 2en \ln n \rceil .$$

This $\ln n$ factor is "useful" to make this and similar arguments work