Randomized Algorithms I

- Probability
- Contention Resolution
- Minimum Cut

Philip Bille

Probability

- Probability spaces.
  - Set of possible outcomes $\Omega$.
  - Each element $i \in \Omega$ has probability $p(i) \geq 0$ and $\sum_{i \in \Omega} p(i) = 1$.

  - Event $E$ is a subset of $\Omega$ and probability of $E$ is $\Pr(E) = \sum_{i \in E} p(i)$.
  - The complementary event $\bar{E}$ is $\Omega - E$ and $\Pr(\bar{E}) = 1 - \Pr(E)$.

- Example. Flip two fair coins.
  - $\Omega = \{HH, HT, TH, TT\}$.
  - $p(i) = 1/4$ for each outcome $i$.
  - Event $E$ = "the coins are the same".
  - $\Pr(E) = 1/2$.

- Conditional probability.
  - What is the probability that event $E$ occurs given that event $F$ occurred?
  - The conditional probability of $E$ given $F$:
    \[
    \Pr(E \mid F) = \frac{\Pr(E \cap F)}{\Pr(F)}
    \]

- Example.
  - $\Pr(E \mid F) = \frac{\Pr(E \cap F)}{\Pr(F)} = \frac{2/8}{5/8} = \frac{2}{5}$.
Probability

- Independence.
  - Events E and F are independent if information about E does not affect outcome of F and vice versa.

\[
\Pr(E \mid F) = \Pr(E) \quad \Pr(F \mid E) = \Pr(F)
\]

- Same as \(\Pr(E \cap F) = \Pr(E) \cdot \Pr(F)\)

Randomized Algorithms I

- Probability
- Contention Resolution
- Minimum Cut

Probability

- Union bound.
  - What is the probability that any of event \(E_1, \ldots, E_k\) will happen, i.e., what is \(\Pr(E_1 \cup E_2 \cup \cdots \cup E_k)\)?

\[
\Pr(E_1 \cup \cdots \cup E_k) = \Pr(E_1) + \cdots + \Pr(E_k).
\]

- If events are disjoint, \(\Pr(E_1 \cup \cdots \cup E_k) = \Pr(E_1) + \cdots + \Pr(E_k)\).
- If events overlap, \(\Pr(E_1 \cup \cdots \cup E_k) < \Pr(E_1) + \cdots + \Pr(E_k)\).
- In both cases, the union bound holds:

\[
\Pr(E_1 \cup \cdots \cup E_k) \leq \Pr(E_1) + \cdots + \Pr(E_k)
\]

Contention Resolution

- Contention resolution. Consider \(n\) processes \(P_1, \ldots, P_n\) trying to access a shared database:
  - If two or more processes access database at the same time, all processes are locked out.
  - Processes cannot communicate.
  - Goal. Come up with a protocol to ensure all processes will access database.
  - Challenge. Need symmetry breaking paradigm.
Contention Resolution

- Applications.
  - Distributed communication and interference.
  - Illustrates simplicity and power of randomized algorithms.

- Protocol. Each process accesses the database at time $t$ with probability $p = 1/n$.

```
\text{database}\quad P_1\quad P_2\quad \ldots\quad P_n
```

Contention Resolution

- Analysis. How do we analyze the protocol?

```
\text{P}_i \quad \text{request for access}
```

- Success for a single process in a single round.
  - $S_{i,t}$ event that $P_i$ successfully accesses database at time $t$.
  - $\Pr(S_{i,t}) = p(1-p)^{n-1} = \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1} \geq \frac{1}{en}$

- Probability that process $i$ requests access.
- Probability that no other process requests access.
- $\left(1 - \frac{1}{n}\right)^{n-1}$ converges to $1/e$ from above.
Contestion Resolution

- Failure for a single process in rounds $1, \ldots, t$.
  - $F_{ix}$ = event that $P_i$ fails to access database in any of rounds $1, \ldots, t$.

$$\Pr(F_{ix}) = \Pr\left(\bigcap_{i=1}^{t} S_{ix}\right) = \prod_{i=1}^{t} \Pr(S_{ix}) = \left(1 - \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1}\right)^t \leq \left(1 - \frac{1}{en}\right)^t$$

- $t = \lceil en \rceil \Rightarrow \Pr(F_{ix}) \leq \left(1 - \frac{1}{en}\right)^{\lceil en \rceil} \leq \left(1 - \frac{1}{en}\right)^{en} \leq \frac{1}{e}$

- $t = \lceil en \rceil (\lceil \ln n \rceil) \Rightarrow \Pr(F_{ix}) \leq \left(\frac{1}{e}\right)^{\lceil \ln n \rceil} = \frac{1}{en^\epsilon} \Rightarrow \Pr(F_{ix}) \leq \frac{1}{en^\epsilon}$

- Conclusion. After $\lceil en \rceil (\lceil \ln n \rceil)$ rounds all processes have accessed database with probability at least $1 - 1/n$.

Randomized Algorithms I

- Probability
- Contention Resolution
- Minimum Cut
**Graphs.** Consider undirected, connected graph $G = (V,E)$.

**Cuts.**
- A cut $(A,B)$ is a partition of $V$ into two non-empty disjoint sets $A$ and $B$.
- The size of a cut $(A,B)$ is the number of edges crossing the cut.
- A minimum cut is a cut of minimum size.

**Minimum Cut**

**Applications.**
- Network fault tolerance.
- Image segmentation.
- Parallel computation.
- Social network analysis.
- ...

**Which solutions do we know?**

**Contraction algorithm.**
- Pick edge $e = (u,v)$ uniformly at random.
- **Contract** $e$.
  - Replace $e$ by single vertex $w$.
  - Preserve edges, updating endpoints of $u$ and $v$ to $w$.
  - Preserve parallel edges, but remove self-loops.
- Repeat until two vertices $a$ and $b$ left.
- Return cut (all vertices contracted into $a$, all vertices contracted into $b$).
Minimum Cut

- **Round 1.**
  - What is the probability that we contract an edge from \( F \) in round 1?
  - Each vertex has \( \deg \geq |F| \) (otherwise smaller cut exists) \( \Rightarrow \sum_{v \in V} \deg(v) \geq |F| \cdot n \).
  - \( \sum_{v \in V} \deg(v) = 2m \Rightarrow m = \frac{\sum_{v \in V} \deg(v)}{2} \geq \frac{|F| \cdot n}{2} \).
  - Probability we contract edge from \( F \) is \( \frac{|F|}{m} \leq \frac{|F|}{|F|/2} = \frac{2}{n} \).

- **Round \( j+1 \).**
  - What is the probability that we contract an edge in round \( j+1 \) from \( F \), given that no edge from \( F \) was contracted in rounds \( 1, \ldots, j \)?
  - \( G' \) is graph after \( j \) rounds with \( n-j \) nodes and no edges from \( F \) was contracted in rounds \( 1, \ldots, j \).
  - Every cut in \( G' \) is a cut in \( G \Rightarrow \) at least \( |F| \) edges incident to every node in \( G' \)
  - \( \Rightarrow |G'| \) contains at least \( \frac{|F| (n-j)}{2} \) edges \( \Rightarrow \) probability is \( \leq \frac{|F|}{m} = \frac{2}{n-j} \).
Minimum Cut

- **Success after all rounds.**
  - $E_j$ = event that an edge from $F$ is not contracted in round $j$.
  - The probability that we return the correct minimum cut is $Pr(E_{0,2} \cap \cdots \cap E_1)$.
  - We know:
    - $Pr(E_1) \geq 1 - \frac{2}{n}$
    - $Pr(E_{j+1} | E_1 \cap \cdots \cap E_j) \geq 1 - \frac{2}{n-j}$
    - Conditional probability definition + algebra $\Rightarrow Pr(E_1 \cap \cdots \cap E_{j+1}) \geq \frac{2}{n^2}$.

Minimum Cut

- **Conclusion.**
  - We return the correct minimum cut with probability $\geq \frac{2}{n^2}$ in polynomial time.

  - **Probability amplification.**
    - Correct solution only with very small probability
    - Run contraction algorithm many times and return smallest cut.
    - With $n^2$ in $n$ runs with independent random choices the probability of failure to find minimum cut is
      $\leq \left(1 - \frac{2}{n^2}\right)^n \leq \left(\frac{1}{e}\right)^{2 \log n} = \frac{1}{n^2}$.

  - **Time.**
    - $\Theta(n^2 \log n)$ iterations that take $\Omega(m)$ time each.
    - More techniques and tricks $\Rightarrow m \log^{O(1)} n$ time solution. [Karger 2000]

Minimum Cut

- **Monte Carlo algorithm.**
  - Randomized algorithm.
  - Guarantee on running time, likely to find correct answer.

  - **Las Vegas algorithm.**
    - Randomized algorithm.
    - Guaranteed to find the correct answer, likely to be fast.

Randomized Algorithms I

- **Probability**
- **Contention Resolution**
- **Minimum Cut**