Ma/CS 6b
Class 14: The Probabilistic Method

By Adam Sheffer

Basic Probability

- A discrete probability space is a finite set $\Omega$. Each $\omega \in \Omega$ is called an elementary event, and has a certain probability $\Pr[\omega] \in [0, 1]$, such that $\sum_{\omega \in \Omega} \Pr[\omega] = 1$.
- Any subset $A \subseteq \Omega$ is an event, of probability $\Pr[A] = \sum_{\omega \in A} \Pr[\omega]$.
- A union of events corresponds to OR and an intersection of events corresponds to AND.
Independent Events

- Two events \( A, B \subset \Omega \) are independent if 
  \[
  \Pr[A \cap B] = \Pr[a] \cdot \Pr[b].
  \]

- **Example.** We flip two fair coins.
  - Let \( \omega_{i,j} \) be the elementary event that coin A landed on \( i \) and coin B on \( j \), where \( i, j \in \{h, t\} \). Each of the four events has a probability of 0.25.
  - The event where coin A lands on heads is \( a = \{\omega_{h,t}, \omega_{h,h}\} \). For B it is \( b = \{\omega_{t,h}, \omega_{h,h}\} \).
  - The events are independent since 
    \[
    \Pr[a \text{ and } b] = \Pr[\omega_{h,h}] = 0.25 = \Pr[a] \cdot \Pr[b].
    \]

(Discrete) Uniform Distribution

- In a **uniform distribution** we have a set \( \Omega \) of elementary events, each occurring with probability \( \frac{1}{|\Omega|} \).
  - For example, when flipping a fair die, we have a uniform distribution over the six possible results.
Union Bound

- For any two events $A, B$, we have
  $$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B].$$

- This immediately implies that
  $$\Pr[A \cup B] \leq \Pr[A] + \Pr[B],$$
  where equality holds iff $A, B$ are disjoint.

- Union bound. For any finite set of events $A_1, \ldots, A_k$, we have
  $$\Pr[\bigcup_i A_i] \leq \sum_i \Pr[A_i].$$

Recall: Ramsey Numbers

- $R(p, p)$ is the smallest number $n$ such that each blue-red edge coloring of $K_n$ contains a monochromatic $K_p$.

- Theorem. $R(p, p) > 2^{p/2}$.
  - We proved this in a previous class.
  - Now we provide another proof, using probability.
Probabilistic Proof

- For some $n$, we color the edges of $K_n$.
  - Each edge is independently and uniformly colored either red or blue.
  - For any fixed set $S$ of $p$ vertices, the probability that it forms a monochromatic $K_p$ is $2^{1-\binom{p}{2}}$.
  - There are $\binom{n}{p}$ possible sets of $p$ vertices. By the union bound, the probability that there is a monochromatic $K_p$ is at most
    \[ \sum_S 2^{1-\binom{p}{2}} = \binom{n}{p} 2^{1-\binom{p}{2}}. \]

Proof (cont.)

- For some $n$, we color the edges of $K_n$.
  - Each edge is colored blue with probability of 0.5, and otherwise red.
  - The probability for a monochromatic $K_p$ is
    \[ \leq \binom{n}{p} 2^{1-\binom{p}{2}}. \]
  - If $n \leq 2^{p/2}$, this probability is smaller than 1.
  - In this case, the probability that we do not have any monochromatic $K_p$ is positive, so there exists a coloring of $K_m$ with no such $K_p$. 
Non-Constructive Proofs

- We proved that there exists a coloring of $K_n$ with no monochromatic $K_p$, but we have no idea how to find this coloring.
- Such a proof is called non-constructive.
- The probabilistic method often proves the existence of objects with surprising properties, but we still have no idea how they look like.

A Tournament

- We have $n$ people competing in thumb wrestling.
  - Every pair of contestants compete once.
  - How can we decide who the overall winner is?
- We build a directed graph:
  - A vertex for every participant.
  - An edge between every two vertices, directed from the winner to the loser.
  - An orientation of $K_n$ is called a tournament.
The King of the Tournament

The winner can be the vertex with the maximum outdegree (the contestant winning the largest number of matches), but it might not be unique.

A king is a contestant $x$ such that for every other contestant $y$ either $x \rightarrow y$ or there exists $z$ such that $x \rightarrow z \rightarrow y$.

Theorem. Every tournament has a king.

Proof

- $D^+(v)$ – the number of vertices reachable from $v$ by a path of length $\leq 2$.
- Let $v$ be a vertex that maximizes $D^+(v)$.
  - Assume for contradiction that $v$ is not a king.
  - Then there exists $u$ such that $u \rightarrow v$ and there is no path of length two from $v$ to $u$.
  - That is, for every $w$ such that $v \rightarrow w$, we also have $u \rightarrow w$.
  - But this implies that $D^+(u) \geq D^+(v) + 1$, contradicting the maximality of $v$!
The $S_k$ Property

- We say that a tournament $T$ has the $S_k$ property if for every subset $S$ of $k$ participants, there exists a participant that won against everyone in $S$.
  - Formally, this is an orientation of $K_n$, such that for every subset $S$ of $k$ vertices there exists a vertex $v \in V \setminus S$ with an edge from $v$ to every vertex of $S$.

- Example. A tournament with the $S_1$ property.

Tournaments with the $S_k$ Property

- Theorem. If $\binom{n}{k} \left(1 - 2^{-k}\right)^{n-k} < 1$ then there is a tournament on $n$ vertices with the $S_k$ property.
  
  - Proof.
    - For some $n$ satisfying the above, we randomly orient $K_n = (V, E)$, such that the orientation of every $e \in E$ is chosen uniformly.
    - Consider a subset $S \subset V$ of $k$ vertices. The probability that a given vertex $v \in V \setminus S$ does not beat all of $S$ is $1 - 2^{-k}$.
Proof (cont.)

- Consider a subset $S \subset V$ of $k$ vertices. The probability that a specific vertex $v \in V \setminus S$ does not beat all of $S$ is $1 - 2^{-k}$.

- $A_S$ – the event of $S$ not being beat by any vertex of $V \setminus S$.

- We have $\Pr[A_S] = (1 - 2^{-k})^{n-k}$, since we ask for $n - k$ independent events to hold.

- By the union bound, we have
  \[
  \Pr \left[ \bigvee_{S \subseteq V, |S|=k} A_S \right] \leq \sum_{S \subseteq V, |S|=k} \Pr[A_S] = \binom{n}{k} (1 - 2^{-k})^{n-k} < 1.
  \]

Completing the Proof

- $A_S$ – the event of $S$ not being beat by any vertex of $V \setminus S$.

- we have
  \[
  \Pr \left[ \bigvee_{S \subseteq V, |S|=k} A_S \right] < 1.
  \]

- That is, there is a positive probability that every subset $S \subset V$ of size $k$ is beat by some vertex of $V \setminus S$. So such a tournament exists.
Which NBA Player is Related to Mathematics?

Michael Jordan  Shaquille O'Neal  LeBron James

Random Variables

• A random variable is a function from the set of possible events to \( \mathbb{R} \).

• Example. Say that we flip five coins.
  ◦ We can define the random variable \( X \) to be the number of coins that landed on heads.
  ◦ We can define the random variable \( Y \) to be the percentage of heads in the tosses.
  ◦ Notice that \( Y = 20X \).
Indicator Random Variables

• An *indicator random variable* is a random variable $X$ that is either 0 or 1, according to whether some event happens or not.

• **Example.** We toss a fair die.
  ◦ We can define the six indicator variable $X_1, \ldots, X_6$ such that $X_i = 1$ iff the result of the roll is $i$.

Expectation

• The *expectation* of a random variable $X$ is
  \[
  E[X] = \sum_{\omega \in \Omega} X(\omega) \Pr[\omega].
  \]
  ◦ Intuitively, $E[X]$ is the expected value of $X$ in the long-run average value when repeating the experiment $X$ represents.
Expectation Example

- We roll a fair six-sided die.
  - Let $X$ be a random variable that represents the outcome of the roll.

\[
E[X] = \sum_{\omega \in \Omega} X(\omega) \Pr[\omega] = \sum_{i \in \{1, \ldots, 6\}} i \cdot \frac{1}{6} = 3.5
\]

While a prisoner of war during World War II, J. Kerrich conducted an experiment in which he flipped a coin 10,000 times and kept a record of the outcomes. A portion of the results is given in the table below.

<table>
<thead>
<tr>
<th>Number of Tosses</th>
<th>Number of Heads</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>100</td>
<td>44</td>
</tr>
<tr>
<td>500</td>
<td>255</td>
</tr>
<tr>
<td>1,000</td>
<td>502</td>
</tr>
<tr>
<td>5,000</td>
<td>2,533</td>
</tr>
<tr>
<td>10,000</td>
<td>5,067</td>
</tr>
</tbody>
</table>
Linearity of Expectation

- If $X$ is a random variable, then $5X$ is a random variable with a value five times that of $X$

**Lemma.** Let $X_1, X_2, \ldots, X_k$ be a collection set of random variables over the same discrete probability. Let $c_1, \ldots, c_k$ be constants. Then

$$E[c_1X_1 + c_2X_2 + \cdots + c_kX_k] = \sum_{i=1}^{k} c_iE[X_i].$$

Fixed Elements in Permutations

- Let $\sigma$ be a uniformly chosen permutation of $\{1, 2, \ldots, n\}$.
  - For $1 \leq i \leq n$, let $X_i$ be an indicator variable that is 1 if $i$ is fixed by $\sigma$.
  - $E[X_i] = \Pr[\sigma(i) = i] = \frac{(n-1)!}{n!} = \frac{1}{n}$.
  - Let $X$ be the number of fixed elements in $\sigma$.
  - We have $X = X_1 + \cdots + X_n$.
  - By linearity of expectation
    $$E[X] = \sum_i E[X_i] = n \cdot \frac{1}{n} = 1.$$
Hamiltonian Paths

- Given a directed graph $G = (V, E)$, a Hamiltonian path is a path that visits every vertex of $V$ exactly once.
  - Major problem in theoretical computer science: Does there exist a polynomial-time algorithm for finding whether a Hamiltonian path exists in a given graph.

Hamiltonian Paths in Tournaments

- Theorem. There exists a tournament $T$ with $n$ players that contains at least $n! 2^{-(n-1)}$ Hamiltonian paths.
Proof

- We uniformly choose an orientation of the edges of $K_n$ to obtain a tournament $T$.
  - There is a bijection between the possible Hamiltonian paths and the permutations of $\{1, 2, \ldots, n\}$. Every possible path defines a unique permutation, according to the order in which it visits the vertices.
  - For a permutation $\sigma$, let $X_\sigma$ be an indicator variable that is 1 if the path corresponding to $\sigma$ exists in $T$.
  - We have $E[X_\sigma] = \Pr[X_\sigma = 1] = 2^{-n+1}$.

- We uniformly choose an orientation of the edges of $K_n$ to obtain a tournament $T$.
  - For a permutation $\sigma$, let $X_\sigma$ be an indicator variable that is 1 if the path corresponding to $\sigma$ exists $T$.
  - We have $E[X_\sigma] = \Pr[X_\sigma = 1] = 2^{-n+1}$.
  - Let $X$ be a random variable of the number of Hamiltonian paths in $T$. Then $X = \sum_\sigma X_\sigma$.

$$E[X] = E\left[\sum_\sigma X_\sigma\right] = \sum_\sigma E[X_\sigma] = n! \cdot 2^{-n+1}.$$  
  - Since this is the expected number of paths in a uniformly chosen tournament, there must be an orientation with at least as many paths.
The End: Michael Jordan

• Math major in college.
  ◦ In his junior year he switched to cultural geography (whatever that means...).