Mathematical Tools in Computer Science, Winter 2006	
Lecture 10	
Lecturer: Nati Linial	*Scribe: Robby Lampert*

This lecture is about the probabilistic method. We demonstrate an interesting feature of the probabilistic method - using probabilistic arguments it derive theorems that do not have relation to probability theory. We start with basic definitions from probability, and proceed with examples of the probabilistic method.

1 Basic Definitions

- A *discrete probability space* is a (finite or countable) set Ω equiped with a non-negative function $\Pr : \Omega \to \mathbb{R}_+$, that is sum up to unity, i.e., $\sum_{\omega\in\Omega} \Pr[\omega] = 1$. The set Ω is called *sample space* and the function $\Pr[\cdot]$ is called *probability function.*
- An *event* is a subset $A \subseteq \Omega$ and its *probability* is set to be $\Pr[A] = \sum_{\omega\in A} \Pr[\omega]$.
- For two events $A, B \subseteq \Omega$, we define the *conditional probability* of A given B to be $\Pr[A|B] \stackrel{def}{=} \Pr[A \cap B] / \Pr[B]$.
- Two events $A, B \subseteq \Omega$ are called *independent* if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$, or equivalently $\Pr[A|B] = \Pr[A]$ or $\Pr[B|A] = \Pr[B]$. Intuitively two events are independent if they do not effect each other, i.e., the occurrence of one event does not change the probability of the occurrence of the other.
- A *discrete random variable* X is a real function $X : \Omega \to \mathbb{R}$ with finite range $\{r_1, ..., r_n\}$. Given a set $A \subseteq \{r_1, ..., r_n\}$ we denote by $\Pr[X \in A]$ the probability of the set $\{\omega : X(\omega) \in A\}$.
- Two discrete random variables X, Y are called *independent* if for every r_i, r_j in their range, we have $\Pr[X = r_i \wedge Y = r_j] = \Pr[X = r_i] \cdot \Pr[Y = r_j]$.
- The *expectation* of a random variable X is the weighted average of its values, i.e., $\mathbb{E}[X] = \sum_{\omega\in\Omega} \Pr[\omega] \cdot X(\omega)$. The expectation of random variable $\mathbb{E}[x]$ is also called its mean and denoted by μ. It is easy to see that the expectation is linear, i.e., for any two constants $\alpha, \beta \in \mathbb{R}$, and two random variables X, Y, the expectation of the random variable $\alpha X + \beta Y$ is $\alpha\mathbb{E}[X] + \beta\mathbb{E}[Y]$.
- The **variance** of a random variable X is defined to be $\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$. The variance quantifies the scattering of X's values around their average, by means of the l_2 norm.

The union bound is a direct consequence of the probability function definition, and is stated without a proof:

Proposition 1 (union bound) *For a family of events $\{A_i\}_{i\in I}$ in a space Ω holds $Pr[\cup_{i\in I} A_i] \leq \sum_{i\in I} Pr[A_i]$*

we also state two basic inequalities which imply that a random variable is not very far from its expectation:

Theorem 2 Markov inequality
For a non-negative random variable $X \geq 0$ and a constant $c \geq 1$ holds

$$Pr[X > c \cdot \mathbb{E}[X]] \leq \frac{1}{c}$$

.

Theorem 3 Chevishev inequality
For a random variable X with a mean $\mathbb{E}[X] = \mu$ and a variance $Var(X) = \sigma^2$, and for any constant c, we have

$$Pr[|X - \mu| > (c \cdot \sigma)] \leq 1/c^2$$

.

2 Stationary cycles in permutations

A *permutation* is a bijection from a finite set onto itself. We define $S_n = \{\pi : [n] \to [n]\}$ the set of permutations over $[n] = \{1, 2, \ldots, n\}$. A *fixed point* of a permutation π is an i such that $\pi(i) = i$. We use probabilistic arguments to show that a permutation π has one fixed point on the average.

Proposition 4 *Let $X : S_n \to \mathbb{N}$ be a function that counts the number of fixed points in a permutation π, i.e., $X(\pi) = |\{i : \pi(i) = i\}|$. Assume that π is sampled from S_n according to the uniform distribution, i.e., $Pr[\pi] = 1/n!$ for every $\pi \in S_n$, then $\mathbb{E}[X] = 1$.*

Proof To calculate $\mathbb{E}[X]$ we use the linearity of expectation by representing X as a sum of random variables. For $1 \leq i \leq n$ define the random variable X_i as

$$X_i(\pi) = \begin{cases} 1 & \pi(i) = i \\ 0 & \text{otherwize} \end{cases}$$

This kind of random variable is called a *characteristic random variable*, since it characterizes a certain event, in the sence that its value is 1 exactly when the event occures, and 0 otherwise. Here, X_i characterizes the event in which i is a fixed point of π. Since $X(\pi) = \sum_{i=1}^{n} X_i(\pi)$ its expectation reduces to $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i]$. Now, for all $1 \leq i \leq n$,

$$\begin{aligned} \mathbb{E}[X_i] &= \tfrac{1}{n!} \cdot |\{\pi \in S_n : \pi(i) = i\}| && \text{(because } X_i \text{ characterizes a fixed point at } i) \\ &= \tfrac{1}{n!} \cdot |\{S_{n-1}\}| && \text{(the number of permutations over } [n]\backslash\{i\}) \\ &= \tfrac{(n-1)!}{n!} = \tfrac{1}{n} . \end{aligned}$$

Thus $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = 1$. ■

A *cycle* of length 2 in a permutation π is a couple $1 \leq i \neq j \leq n$ such that $\pi(i) = j$ and $\pi(j) = i$. Similarly, a cycle of length k is a k-tuple $i_1, \ldots, i_k \in [n]$ such that $\pi(i_j) = i_{j+1}$ for all $1 \leq j \leq k-1$ and $\pi(i_k) = i_1$. There are $nchoosek(k-1)!$ k-cycles in $[n]$ since every set of k elements contains $(k-1)!$ cycles, or equiavlently in every set of k elements there are k different tuples which represents the same cycle.

Proposition 5 *Let $X : S_n \to \mathbb{N}$ be a function that counts the number of k-cycles in a permutation. Assume that π is sampled from S_n according to the uniform distribution, then $\mathbb{E}[X] = 1/k$.*

Proof Let $\mathcal{S}$ be the family of all possible k-cycles. Define a random variable $X : S_n \to \mathbb{N}$ that counts the number of k-cycles in a permutation π. For every $(i_1, ..., i_k) \in \mathcal{S}$ let $X_{i_1,...,i_k} : S_n \to \{0, 1\}$ characterize the event $\pi(i_j) = i_{j+1}$ for $j = 1, ..., k-1$ and $\pi(i_k) = i_1$. For every $\pi \in S_n$, the random variable $X(\pi)$ is represented as $\sum_{i_1,..,i_k \in \mathcal{S}} X_{i_1,...,i_k}(\pi)$. Since $X_{i_1,...,i_k}$ is a characteristic random variable, its mean is the probability for a random permutation to have the k-cycle $i_1, ..., i_k$. Therefore,

$$\mathbb{E}\left[X_{i_1,...,i_k}\right] = \frac{|S_{n-k}|}{|S_n|} = \frac{(n-k)!}{n!} = \frac{(n-k)!}{n!},$$

and

$$\mathbb{E}\left[X\right] = \sum_{i_1,...,i_k \in \mathcal{S}} \frac{(n-k)!}{n!} = \binom{n}{k}(k-1)!\frac{(n-k)!}{n!} = \frac{1}{k}$$

■

3 Ramsey numbers and the probabilistic method

Another application for the probabilistic method presents a lower bound for Ramsey numbers.

Theorem 6 (Ramsey Theorem) *For every $k, l \in \mathbb{N}$ there exists a number $R = R(k, l) \in \mathbb{N}$ such that if we color the edges of K_R (the clique on R vertices) with blue and red, it will necessarily contain a blue K_k or a red K_l.*

An upper bound can be shown to be $R(k, l) \leq \binom{k+l-2}{k-1}$ or for $k = l$: $R(k, k) \leq 2^{2k} = 4^k$. Its meaning is that for any graph with n vertices there exists a clique or an anti-clique of size $\geq \frac{1}{2}\log_2 n$.

We show that there are graphs on n vertices without a clique or an anti-clique of size $> 2\log_2 n$, therefore proving the lower bound $R(k, k) \geq 2^{k/2}$. The key idea is to consider the set of graphs on n vertices as a probability space Ω of size $2^{\binom{n}{2}}$, where every edge can be present or not. An important family of such probability spaces is $G(n, p)$ - where each edge is present with probability p. If we set $e(G)$ to be the number of edges in a graph G then the probability we randomly select the graph G is $\Pr[G] = p^{e(G)} \cdot (1-p)^{\binom{n}{2} - e(G)}$. Alternatively, we can describe a sampling of a graph from $G(n, p)$ as a process in which for every $1 \leq i < j \leq n$, independently, we put the edge (i, j) in the graph with probability

p. Note that in $G(n, \frac{1}{2})$ all the graphs have the same probability (we have the uniform distribution on the set of graphs of size n).

To prove the desired lower bound we define two random variables: $X(G)$=the number of k-cliques in G, and $Y(G)$=the number of k-anti-cliques in G. We show that for the choise of $k = 2\log n$ we get $\mathbb{E}[X] = o(1)$ and $\mathbb{E}[Y] = o(1)$, which imply $\mathbb{E}[X+Y] = o(1)$. By proposition **??**, this implies that $\Pr[X+Y=0] > 0$, i.e., there are graphs in the probability space for which $X+Y=0$, that is, they have neither a clique nor an anti-clique of size k. We represent X as a sum of characteristic random variables X_S, for all $S \subseteq [n]$ of size k, where for each such S,

$$X_S = \begin{cases} 1 & S \text{ is a clique} \\ 0 & \text{otherwise} \end{cases}$$

Then,

$$\begin{aligned} \mathbb{E}[X] &= \sum_S \mathbb{E}[X_S] = \sum_S \Pr[S \text{ is a clique}] = \sum_S \frac{1}{2}^{\binom{k}{2}} = \binom{n}{k} 2^{-\frac{k(k-1)}{2}} \\ &\leq \left(\frac{ne}{k}\right)^k \left(2^{-\frac{k-1}{2}}\right)^k = \left(\frac{ne}{k2^{\frac{k-1}{2}}}\right)^k \end{aligned}$$

where the inequality is derived using Stirling approximation to $\binom{n}{k}$. If we take $k > 2\log n$ then we get $\mathbb{E}[X] < 1$.

Proposition 7 *Let Z be a random variable with range $0, 1, 2, \ldots, N$ If $\mathbb{E}[Z] < \varepsilon$ then $Pr[Z=0] > 1-\varepsilon$. In particular, if $\mathbb{E}[Z] < 1$ then $Pr[Z=0] > 0$.*

Proof $\mathbb{E}[X] = \sum_{i=1}^{N} \Pr[X \geq i] < \varepsilon$. Since $\sum_{i=0}^{N} \Pr[X=i] = 1$ the claim follows. ■