CS 683: Advanced Design and Analysis of Algorithms 01/25/2008

Lecture Notes 3

Professor: John Hopcroft *Scribes: Anand Bhaskar & Kathy King*

1 Recap of last class

Threshold of a property: For a particular property, if there exists a function $p(n)$ such that for any function $p_1(n)$ where $\lim_{n\to\infty} \frac{p_1(n)}{p(n)} = 0$, then $G(n, p_1(n))$ does not have the property and for any function $p_2(n)$ where $\lim_{n\to\infty} \frac{p_2(n)}{p(n)} = \infty$, then $G(n, p_2(n))$ has the property, then $p(n)$ is called a **threshold** for that property.

Sharp threshold of a property: For a particular property, if there exists a function $p(n)$ such that $G(n, cp(n))$ for any $c < 1$ does not have the property and for $G(n, cp(n))$ for any $c > 1$ has the property, then $p(n)$ is called a **sharp threshold** for the property.

Note: An interesting project would be to determine the conditions under which a threshold is a sharp threshold.

2 Review of inequalities

2.1 Markov's Inequality

If X is a non-negative random variable and $a > 0$, then $P(X \geq a) \leq \frac{E(X)}{a}$.

Proof:

$$\begin{aligned} E(X) &= \int_0^\infty xp(x)dx \\ &= \int_0^a xp(x)dx + \int_a^\infty xp(x)dx \\ &\geq \int_a^\infty xp(x)dx \\ &\geq \int_a^\infty ap(x)dx \\ &= aP(X \geq a) \end{aligned}$$

Hence, $P(X \geq a) \leq \frac{E(X)}{a}$. $\square$

2.2 Chebyshev's Inequality

Let X be a random variable with mean $\mu = E(X)$ and finite variance $\sigma^2 = Var(X)$. Then $P(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2}$.

Proof:
$P(|X-\mu| \geq a\sigma) = P((X-\mu)^2 \geq a^2\sigma^2)$
Using Markov's inequality for the non-negative random variable $(X-\mu)^2$, we get
$P((X-\mu)^2 \geq a^2\sigma^2) \leq \dfrac{E((X-\mu)^2)}{a^2\sigma^2} = \dfrac{\sigma^2}{a^2\sigma^2} = \dfrac{1}{a^2}$

Hence, $P(|X-\mu| \geq a\sigma) \leq \frac{1}{a^2}$. □

3 Method of second moments

The method of second moments is used to show that when the expected value of a non-negative random variable X is large compared to its variance, the random variable takes the value zero with probability approaching zero. That is, $\Pr(X=0) \to 0$.

Let X be a non-negative random variable with variance σ^2.
Clearly, $P(X=0) = P\left[E(X) - X = E(X)\right] \leq P\left[|X - E(X)| \geq E(X)\right]$.

Using Chebshev's inequality on $P(|X-E(X)| \geq E(X))$, we get
$P(|X-E(X)| \geq E(X)) \leq \frac{1}{\left(\frac{E(X)}{\sigma}\right)^2} = \frac{\sigma^2}{E^2(X)}$.
Hence, $P(X=0) \leq \frac{\sigma^2}{E^2(X)}$.
If $\frac{\sigma^2}{E^2(X)} \to 0$, then $P(X=0) \to 0$.

We can rewrite $\sigma^2 = Var(X) = E(X^2) - E^2(X)$.
Hence, $\frac{\sigma^2}{E^2(X)} = \frac{E(X^2)}{E^2(X)} - 1$.
To show that $\frac{\sigma^2}{E^2(X)} \to 0$ is equivalent to showing that $\frac{E(X^2)}{E^2(X)} - 1 \to 0$, which in turn is equivalent to showing that $\frac{E(X^2)}{E^2(X)} \to 1^+$ (it has to approach 1 from the right because $E(X^2) \geq E^2(X)$ since $Var(X) \geq 0$).

4 Application of the Method of second moments: Sharp threshold for the disappearance of isolated vertices

Consider the property that a graph does not have any isolated vertices. We will show that this property has a sharp threshold of $p(n) = \dfrac{\ln n}{n}$ in the random graph model $G(n, p(n))$.

Let X be the random variable which counts the number of isolated vertices of a graph generated by $G(n, p(n))$.
Define the indicator variable $X_i = \begin{cases} 1 & \text{vertex } i \text{ is isolated} \\ 0 & \text{vertex } i \text{ is not isolated} \end{cases}$

Hence, $X = \displaystyle\sum_{i=1}^{n} X_i$. So we have

$$
\begin{aligned}
E(X) &= E\left(\sum_{i=1}^{n} X_i\right) \\
&= \sum_{i=1}^{n} E(X_i) \\
&= nE(X_1) \\
E(X) &= n(1-p(n))^{n-1} \quad \text{(since vertex 1 shouldn't be connected to any of the other } n-1 \text{ vertices)}
\end{aligned}
$$

Setting $p(n) = \dfrac{c\ln n}{n}$, we get

$$
\begin{aligned}
E(X) &= n\left(1-\tfrac{c\ln n}{n}\right)^{n-1} \\
\lim_{n\to\infty} E(X) &= \lim_{n\to\infty} n\left(1-\frac{c\ln n}{n}\right)^{n-1} = \lim_{n\to\infty} n\left(1-\frac{c\ln n}{n}\right)^{n} \\
&= \lim_{n\to\infty} ne^{-c\ln n} \\
\lim_{n\to\infty} E(X) &= \lim_{n\to\infty} n^{1-c}
\end{aligned}
$$

If $c > 1$, then $\lim_{n\to\infty} E(X) = 0$. This says that as $n \to \infty$, the expected number of isolated vertices in a random graph generated by $G(n, \frac{c\ln n}{n})$ is 0. Using this, we can argue that as $n \to \infty$, the probability that a random graph generated by $G(n, \frac{c\ln n}{n})$ has isolated vertices tends to 0. Suppose this were not true, ie. the probability that a random graph generated by $G(n, \frac{c\ln n}{n})$ has isolated vertices tends to some positive constant q. In that case, since a graph with isolated vertices has at least one isolated vertex, the expected number of isolated vertices is at least $q\cdot 1 = q > 0$ which contradicts the fact that $\lim_{n\to\infty} E(X) = 0$.

If $c < 1$, then $\lim_{n\to\infty} E(X) = \infty$. We cannot simply assume that any random graph generated by $G(n, \frac{c\ln n}{n})$ has many isolated vertices as $n \to \infty$, because it might be that the isolated vertices are concentrated in a small subset of these graphs. Hence, we must use a second moment argument. We want to show that $\lim_{n\to\infty} \dfrac{E(X^2)}{E^2(X)} = 1$. We have already calculated $E(X)$, so it remains to calculate $E(X^2)$:

$$
\begin{aligned}
E(X^2) &= E\left(\sum_{i=1}^{n} X_i^2 + \sum_{i\neq j} X_iX_j\right) \\
&= E\left(\sum X_i\right) + n(n-1)E(X_1X_2) \text{ (since } X_i \text{ is either 0 or 1, then } X_i^2 = X_i) \\
&= E(X) + n(n-1)\Pr(X_1 = 1 \text{ and } X_2 = 1) \\
&= E(X) + n(n-1)\left((1-p(n))^{n-1}(1-p(n))^{n} - 2\right) \\
&= E(X) + n(n-1)(1-p(n))^{2n-3}.
\end{aligned}
$$

So, recalling that $E(X) = n(1-p(n))^{n-1}$, we have

$$
\frac{E(X^2)}{E^2(X)} = \frac{E(X) + n(n-1)(1-p)^{2n-3}}{E^2(X)}
$$

$$= \frac{1}{n(1-p(n))^{n-1}} + \frac{n(n-1)(1-p(n))^{2n-3}}{n^2(1-p(n))^{2n-2}},$$

and since $p(n) = \dfrac{c \ln n}{n}$ with $c < 1$ we get

$$\begin{aligned}
\lim_{n\to\infty} \frac{E(X^2)}{E^2(X)} &= \lim_{n\to\infty} \left(0 + \frac{1}{1-p(n)}\right) \\
&= \lim_{n\to\infty} \frac{1}{1-\frac{c\ln n}{n}} \\
&= \lim_{n\to\infty} \frac{n}{n - c\ln n} \\
&= 1.
\end{aligned}$$

So by the method of second moments we have $\Pr(X = 0) \to 0$ as $n \to \infty$. The number of graphs generated by $G(n, \frac{c\ln n}{n})$ with $c < 1$ with no isolated vertices tends to zero as n grows. Equivalently, as $n \to \infty$, the probability that a graph has one or more isolated vertices approaches 1. Be careful, though: it is *not* the case that *all* random graphs generated by $G(n, \frac{c\ln n}{n})$ will have isolated vertices, but there is not a constant fraction of these graphs that have no isolated vertices. Hence, the set of graphs with no isolated vertices is vanishingly small as n gets large.

In conclusion, $p(n) = \frac{\ln n}{n}$ is a sharp threshold for the property of a graph not containing any isolated vertices.