**CS 761: Randomized Algorithms** Fall 2019

Lecture 6 — October 11, 2019

*Prof. Gautam Kamath* *By: Pablo Millan Arias, Egill Ian Gudmundsson*
*Edited by Vedat Levi Alev*

**Disclaimer:** These notes have not been subject to the usual scrutiny reserved for formal publications.

# 1 Probabilistic Method

The Probabilistic Method is used to determine whether an object or solution (satisfying some given conditions) exists. Finding such a solution using a deterministic method can prove very difficult and complex and often times we can more easily reach a solution with a simpler and suitable probabilistic model instead. We can show that an object or solution satisfying our condition exists by proving that the probability of sampling an object having this property is greater than zero. The probabilistic method can be described as taking the following steps:

1. Construct a probability space over the possible objects and their attributes (e.g. graphs)
2. Show that $\Pr(\texttt{sampled object having the properties}) > 0$

For our first example, we consider the problem of coloring the edges of a graph with two colors so that there are no large cliques with all edges having the same color.

Let us show an example and define the following:

**Definition 1** (Complete Graph)**.** *We define $K_n$ as the complete graph of $n$ vertices where the graph is undirected and every pair of distinct vertices is connected by a unique edge.*

**Theorem 2.** *If $\binom{n}{k}2^{-\binom{k}{2}+1} < 1$ in $K_n$, one can color all the edges in $K_n$ with 2 colors, subject to no subgraph of size $k(K_k)$ being monochromatic.*

*Proof.* First, we construct our probability space of the possible solutions. Consider all the random colorings of the edges ($2^{\binom{n}{2}}$ many) with the colors red and blue and consider some fixed set of $k$ vertices.

We know the probability of getting a monochromatic graph is $2^{-\binom{k}{2}} + 2^{-\binom{k}{2}} = 2^{-\binom{k}{2}+1}$

Where the first term on the left hand side is the chance of getting an all-red graph and the second term the chance of getting an all-blue graph.

Second, we show that the probability of a solution existing is greater than 0. Now, using the union bound, we get:

$$\begin{aligned}\Pr(\cup_i \texttt{clique i is monochromatic }) &\leq \sum \Pr(\texttt{clique is not monochromatic}) \\ &= \binom{n}{k}^i 2^{-\binom{k}{2}+1} \\ &< 1 \\ &\rightsquigarrow \Pr(\cap \texttt{clique is not monochromatic})\end{aligned}$$

□

This works for $k = \Theta(\log(n))$. For the deterministic model, it would be $\Theta(\sqrt{n})$.

**Definition 3** (Erdős-Rényi Graph Model)**.** *The $G_{n,p}$ model, due to Erdős and Rényi, has two parameters, $n$ and $p$. The parameter $n$ is the number of vertices of the graph and $p$ is the edge probability. For each pair of distinct vertices, $v$ and $w$, $p$ is the probability that the edge $(v, w)$ is present. The presence of each edge is statistically independent of all other edges and the graph-valued random variable with these parameters is denoted by $G_{n,p}$.*

Note that if we sample a graph from $G_{n,1/2}$, we can see that it is analogous to the problem above, where an edge being in place would be a red edge and an edge being absent would be a blue edge.

The max clique size of a graph like this would be $O(\log(n))$ and the max independent set would be $O(\log(n))$.

## 2 First Moment Method

The first moment method is a similar and sometimes simpler approach for proving the existence of an object with desired properties. It consists in computing the expected value of a random variable and then concluding that the random variable must take some values that are greater and smaller than the expected value. In other words, if $E[x] = c$ then $\Pr(X \geq C) > 0$. The following lemma is a formal statement of the previous intuition.

**Lemma 4.** *Let $\mathcal{S}$ be a probability space and $X$ a random variable defined on $\mathcal{S}$ such that $\mathbf{E}[X] = \mu$. Then $\Pr(X \geq \mu) > 0$ and $\Pr(X \leq \mu) > 0$*

*Proof.* If $\Pr(X \geq \mu) = 0$, then

$$\mu = \sum_x x \Pr(X = x) = \sum_{x<\mu} x \Pr(X = x) < \sum_{x<\mu} \mu \Pr(X = x) = \mu$$

which is a contradiction. In a similar way, assuming that $\Pr(X \leq \mu) = 0$ , gives:

$$\mu = \sum_x x \Pr(X = x) = \sum_{x>\mu} x \Pr(X = x) > \sum_{x>\mu} \mu \Pr(X = x) = \mu$$

again yielding a contradiction □

A direct application of the first moment method is related to independent sets. An independent set in a graph $G$ is a set of vertices with no edges between them. It is known that finding the largest independent set in a graph is an NP-hard problem. The following theorem shows that the probabilistic method can yield bounds on the size of the largest independent set of a graph. After we have taken the 2 steps detailed below, the remaining vertices form an independent set in the original graph.

**Theorem 5.** *Let $G = (V, E)$ be a graph on $n$ vertices with $m \geq \frac{n}{2}$ edges. Then it has an independent set with at least $\frac{n^2}{4m}$ vertices.*

*Proof.* The average degree of such a graph is $d = \frac{2m}{n} \geq 1$. We perform the following steps:

1. Iterate over all vertices and remove it with probability $1 - \frac{1}{d}$. If the vertex is removed, also remove all adjacent edges.
2. Iterate through all remaining edges. Let each edge be connected to the first vertex $i$ and the second vertex $j$. Remove the edge $(i, j)$ and then remove either $i$ or $j$ (selected randomly).

We define the following random variables:

$X$: The number of vertices left in the graph after step 1 is performed.

$Y$: The number of edges left in the graph after step 1 is performed.

These random variables are of interest because we know the following:

$$\texttt{\# vertices in the end} \geq X - Y$$

We can calculate the expected value of the number of vertices:

$$\begin{aligned} E[X] &= \frac{1}{d} \cdot n \\ &= \frac{n}{d} \end{aligned}$$

And then the expected value of the number of egdes:

$$\begin{aligned} E[Y] &= \left(\frac{1}{d^2}\right) \cdot \left(\frac{nd}{2}\right) \\ &= \frac{n}{2d} \end{aligned}$$

In the expected value of Y, the term $\frac{1}{d^2}$ corresponds to the probability of both vertices of an edge not getting removed (i.e. the edge surviving).

So we calculate the expected value and substitute the value of $d$:

$$
\begin{aligned}
E[X-Y] &= E[X]-E[Y] \\
&= \frac{n}{d} - \frac{n}{2d} \\
&= \frac{n}{2d} \\
&= \frac{n^2}{4m}
\end{aligned}
$$

□

Note that the previous theorem is a weak version of Turán's theorem.

**Definition 6** (Girth of a graph)**.** *The girth of a graph is the length of the shortest cycle within the graph. If there are no cycles within the graph, the girth is considered to be $\infty$.*

The previous definition introduces another interesting application of the probabilistic method. Intuition might make us think that dense graphs will have small girth. However, we will show that there are dense graphs with relatively large girth.

**Theorem 7.** *For any integer $k \geq 3$ and $n$ sufficiently large, there exist graphs with $n$ vertices such that the number of edges is at least $\frac{1}{4}n^{1+1/k}$ and the girth of the graph is at least $k$.*

*Proof.* First we sample a random graph $G \in G_{n,p}$ where $p = n^{\left(\frac{1}{k}\right)-1}$. We define the random variable $X$ as the number of edges in the graph. The expected value is:

$$
E[X] = p\binom{n}{2} = \frac{1}{2}n(n-1)n^{1/k}n^{-1} = \frac{1}{2}\left(1-\frac{1}{n}\right)n^{1+\frac{1}{k}}
$$

In addition, the number of possible cycles of length $i$ is:

$$
\binom{n}{i}\frac{(i-1)!}{2}
$$

Where the term $\binom{n}{i}$ is the different ways in which we can pick vertices for a cycle and the term $\frac{(i-1)!}{2}$ is the number of ways in which we can order $i$ vertices for a cycle, which we divide by 2 since reverse orderings produce the same cycle.

Let $Y$ the random variable that describes the number of cycles of length $\leq k-1$. With the information above, we can calculate $E[Y]$:

$$
\begin{aligned}
E[Y] = \sum_{i=3}^{k-1} p^i \binom{n}{i}\frac{(i-1)!}{2} &= \sum \frac{n!}{(n-i)!\cdot i!}\cdot\frac{(i-1)!}{2}\cdot p^i = \sum \frac{n!}{2i\cdot(n-1)!}\cdot p^i \\
&\leq \sum n^i p^i = \sum n^i n^{i/k-1} = \sum_{i=3}^{k-1} n^{\frac{i}{k}} \\
&\leq kn^{\frac{k-1}{k}} = kn^{1-1/k}
\end{aligned}
$$

Now we modify the original graph $G$, removing one edge from each of the cycles whose length $\leq k-1$. After this is done, the modified graph has a girth of at least $k$. The random variable $X - Y$ will provide an upper bound on the number of remaining edges. In other words, $X - Y \leq$ the number of remaining edges. We can apply linearity of expectation to calculate it's expected value

$$\begin{aligned} E[X-Y] &= \frac{1}{2}\left(1-\frac{1}{n}\right) n^{1+1/k} - kn^{1-1/k} && (1) \\ &= n^{1+1/k}\left(\frac{1}{2} - \frac{1}{2n} - \frac{k}{n^{2/k}}\right) && (2) \\ &\geq \frac{1}{4} n^{1+1/k} && (3) \end{aligned}$$

Which proves that such a graph exists, it is important to note that the inequality in the last step holds for large enough values of $n$. □

The last application of this method has to do with the graph coloring problem, in which the objective is to use the minimum number of colors to color all vertices so that every pair of adjacent vertices receive different colors.

**Definition 8** (Chromatic number)**.** *We define the chromatic number as the least number of colors necessary to color all vertices in a graph so that no 2 neighbors (connected by an edge) have the same color.*

**Theorem 9.** *For all k,l there exist $G = (V, E)$ with a chromatic number larger than l, and girth smaller than k*

*Proof.* Let's consider a random graph $G \in G_{n,p}$ and define $\chi(G)$ as the chromatic number of graph $G$ and $\alpha(G)$ as the size of the maximum independent set of graph $G$. From this, we deduce:

$$\frac{n}{\chi(G)} \leq \alpha(G) \Rightarrow \frac{n}{\alpha(G)} \leq \chi(G)$$

Each color defines an independent set with regards to $\alpha(G)$

$$Pr[\alpha(s) \geq t] \leq \binom{n}{t}(1-p)^{\binom{t}{2}} \leq n^t e^{-p\binom{t}{2}} = \left(n \cdot e^{\frac{p(t-1)}{2}}\right)^t$$

Given that $t$ is at most $\left\lceil 3\frac{\ln(n)}{p} \right\rceil$, we can see that:

$$\left(n \cdot e^{\frac{p(t-1)}{2}}\right)^t \ll \frac{1}{2}$$

$$p = n^{\frac{1}{k}-1} \Rightarrow t \approx \Theta(n^{1-1/k}\log(n)) \Rightarrow \alpha(G) \leq \Theta(n^{1-1/k}\log(n))$$

With a positive probability. We now remove one vertex from each cycle of length $\leq k-1$. These removals do not increase independent set size. According to (3) in Theorem 7, the amount of removed vertices is at most $kn^{1-1/k}$ and the final expression is

$$\chi(G') \geq \frac{|V(G')|}{\alpha(G')} \geq \frac{n - n^{1-1/k}}{3n^{1-1/k}\ln(n)} = \frac{n^{1/k-1}(n - n^{1-1/k})}{3\ln(n)} \geq l$$

For large values of $n$.

□

# 3 Second Moment Method

Using the first moment method we can conclude that given a random variable, there is one outcome with value that is at least $E[x]$ or at most $E[x]$. We can also use Markov's inequality to show that if $x \geq 0$, then $Pr[x \geq 1] < E[x]$. If $x$ is integral valued and $E[x] \ll 1$, then it follows that $Pr[x = 0]$ is high. The question becomes, how do we prove that $Pr[x \geq 1]$ is large?

**Theorem 10.** *If $x$ is integral, then $\Pr[x = 0] \leq \frac{Var[x]}{(E[x])^2}$*

*Proof.* We can see that:

$$\Pr[x = 0] \leq \Pr(|x - E[x]| \geq E[x]) \leq \frac{Var[x]}{(E[x])^2}$$

□

**Remark 11.** *If either of the two following statements holds:*

1. $Var[x] = o(E[x]^2)$
2. $E[x^2] = (1 + o(1))E[x]^2$

*Then $\Pr[x = 0]$ is close to 0. This means that $x \geq 1$ with high probability.*

## 3.1 Thresholds in random graphs

Random graphs undergo some changes in their structure when the edge probability passes some threshold value. One example of this is the appearance of cycles in $G_{n,p}$ when $p$ reaches the value of $1/n$. However, the most important example of this phenomenon is the emergence of a giant component, an isolated sub-graph that contains a finite fraction of the entire graph's vertices.

For small $p$, with $p = \frac{d}{n}$, $d < 1$, each connected component in the graph is small. For $d > 1$, there is a giant component. There is a rapid transition at the threshold $d = 1$. Below the threshold, the probability of a giant component is very small, and above the threshold, the probability is almost one. This is an example of what is called a threshold behavior, as defined below.

**Definition 12** (Threshold Behavior)**.** *A property of a random graph is said to possess a threshold behavior if there exists a function $f(n)$ such that:*

1. *Almost certainly the graph $G_{n,g(n)}$ does not have the property if:*

$$\lim_{n\to\infty} \frac{g(n)}{f(n)} = 0.$$

2. *Almost certainly the graph $G_{n,h(n)}$ has the property if:*

$$\lim_{n\to\infty} \frac{h(n)}{f(n)} = \infty.$$

In the above definition, "almost certainly" means that the probability goes to one as $n$ approaches infinity, and $h, g$ are arbitrary functions describing the probability over the edges.

**Definition 13** (Sharp Threshold Behavior)**.** *A property has a sharp threshold behaviour if there exists a function $f(n)$ such that, for any $\mathcal{E} > 0$, $G_{n,(1-\mathcal{E})f(n)}$ almost certainly does not have the property, but $G_{n,(1+\mathcal{E})f(n)}$ does.*

For example, let's consider the property of having a clique of size 4. We can define the random variable $X$ as the number of 4-cliques in a graph. From this, we get:

$$E[X] = \binom{n}{4} p^6 \approx n^4 p^6.$$

Note that if $p = o(n^{-2/3})$ then $E[X] = o(1)$ and we can use the first moment method to conclude that $Pr[X = 0] \to 0$.

An example of a property with a sharp transition is that of a random graph having diameter less than or equal to two. The diameter of a graph is the maximum length of the shortest path between a pair of vertices.

This property will be studied for the remaining of this section and we will prove the following theorem:

**Theorem 14.** *The property that $G_{n,p}$ has diameter two has a sharp threshold at $p = \sqrt{\frac{2\ln(n)}{n}}$.*

*Proof.* If $G$ has a diameter greater than two, we say that a pair of vertices $i$,$j$ is a "bad pair" if there is no edge between them and no third vertex is connected to both $i$ and $j$.

For every pair $i, j$ with $i < j$, let $X_{ij}$ be an indicator random variable of $i, j$ being a bad pair. Then $E[X_{ij}] = (1-p)(1-p)^{n-2}$ as each of the other $n-2$ vertices is not connected to both $i$ and $j$. Let

$$X = \sum_{i<j} X_{ij}$$

be the number of bad pairs of vertices. Note that a graph has diameter at most two if and only if it has no bad pairs. i.e, $X = 0$. We then get:

$$E[X] = \sum E[X_{ij}] = \binom{n}{2}(1-p)(1-p^2)^{n-2}.$$

Setting $p = \sqrt{c\frac{\ln(n)}{n}}$,

$$\begin{aligned} E[X] &\approx \frac{n^2}{2}\left(1 - \sqrt{\frac{c\ln(n)}{n}}\right)\left(1 - \frac{c\ln(n)}{n}\right)^{n-2} \\ &\approx \frac{n^2}{2}e^{-c\ln(n)} \\ &\approx \frac{n^{2-c}}{2}. \end{aligned}$$

If $c > 2$, then $\lim_{x\to\infty} E[X] \to 0$ and the First Moment Method gives us a diameter of $\leq 2$ with high certainty.

Next, consider the case $c < 2$, where $\lim_{x\to\infty} E[X] \to \infty$. We need no use a second moment argument to prove that almost certainly the graph has a bad pair and subsequently a diameter greater than two.

$$E[X^2] = E\left[\left(\sum_{i<j} X_{ij}\right)^2\right] = E\left[\sum_{i<j} X_{ij} \sum_{k<l} X_{kl}\right] = E\left[\sum_{\substack{i<j\\k<l}} X_{ij}X_{kl}\right] = \sum_{\substack{i<j\\k<l}} E\left[X_{ij}X_{kl}\right].$$

The previous summation can be partitioned into three different cases depending on the distinct number of indices among $i, j, k$ and $l$ that are involved. We now look at three cases:

1. $i \neq j, k \neq l$:

$$\sum_{\substack{i<j\\k<l}} E\left[X_{ij}X_{kl}\right] = \sum_{\substack{i<j\\k<l}} E[X_{ij}]E[X_{kl}] \leq \left(\sum_{i<j} E[X_{ij}]\right)\left(\sum_{k<l} E[X_{kl}]\right) = (E[X])^2.$$

2. $(i, j) = (k, l)$:

$$\sum_{\substack{i<j\\k<l}} E\left[X_{ij}X_{kl}\right] = \sum_{i<j} E[X_{ij}^2] = \sum_{i<j} E[X_{ij}] = E[X].$$

   This is because all $X_{ij}$ are indicator variables for which $X_{ij} = X_{ij}^2$.

3. $(i, j), (i, k)$: Suppose that both pairs $(i, j)$ and $(i, k)$ are bad. Then, for all other vertices either not adjacent to $i$ or adjacent to $i$ but **not** adjacent to $j + k$, the probability would be $(1 - p) + p(1 - p)^2 \approx 1 - 2p^2$. So:

$$E[X_{ij}X_{ik}] = (1 - p)^2(1 - 2p^2)^{n-3} \approx e^{-2p^2 n}.$$

   Since the number of triples of can only be $3\binom{n}{3}$, then:

$$\sum_{\substack{i<j\\k<l}} E\left[X_{ij}X_{kl}\right] < \frac{n^3}{2}e^{-2p^2 n}.$$

Now, recall that $p = \sqrt{c\frac{\ln(n)}{n}}$ and substitute that in the previous expression:

$$\sum_{\substack{i<j\\k<l}} E\left[X_{ij}X_{kl}\right] \leq \frac{n^3}{2}e^{-2c\ln(n)} = \frac{1}{2}n^{3-2c} = o(E[X]^2).$$

Finally, adding all the term together we get:

$$E[X^2] \leq E[X]^2 + E[X] + o(E[X]^2) = (1+o(1))E[X]^2.$$

By the remark of the second moment method (Theorem 10). We conclude that if $c < 2$, $X \leq 1$ almost certainly, meaning that the diameter is $\geq 3$ with a high degree of certainty.

□

# 4 Lovasz Local Lemma (LLL)

Let $E_1, E_2, ..., E_n$ be a set of "bad" events that we would like to avoid. In this case, we would like to reach the goal $\Pr[\bigcap_i E_i] > 0$ where none of these events happen. This would be easy if the events $E_1, E_2, ..., E_n$ were independent and $\sum \Pr[E_i] < 1$ were union bound.

**Definition 15** (Dependency Graph)**.** *A dependency graph $G$ is a graph with vertices $\{E_1, E_2, ..., E_n\}$ where the edges represent a dependency of one object on another.*

$E_i$ is mutually independent of $\{E_j | (i,j) \notin E\}$, that is $\Pr\left(E_i | \cap_{j\in I} E_j\right) = \Pr(E_i) \Leftrightarrow E_i$ is mutually independent of $\{E_j : j \in S\}$. (Note: $\forall I \subseteq S$)

**Theorem 16** (**Lovasz Local Lemma**)**.** *Let $E_1, \ldots, E_n$ be a set of events and suppose the following hold:*

1. $\Pr(E_i) \leq p$
2. *The maximum degree in $G$ is at most $d$*
3. $4dp \leq 1$

*Then*

$$\Pr\left(\bigcap_{i\in[n]} E_i\right) > 0$$

Before going into the details of the proof, we show two applications of the Local Lemma.

**Example 4.1** ($k$**-SAT**)**.** Given a boolean formula with exactly $k$ variables in each clause. We would like to find a truth assignment to the variables $x_1, x_2, \ldots, x_n$ such that every clause $C_1, C_2, \ldots, C_m$ is satisfied. For example, the following assigment:

$$x_1 = \texttt{False},\ x_2 = \texttt{True},\ x_4 = \texttt{False}$$

is a satisfying assignment to the Boolean formula:

$$(x_1 \vee x_2 \vee x_4) \wedge (\bar{x}_1 \vee x_3 \cup \bar{x}_4).$$

This problem is NP-complete in general, but we can prove that if each variable appears in a small number of clauses, then the formula has a satisfying assignment.

**Theorem 17.** *If no variable appears in more than $T = \frac{2^k}{4k}$ clause, then the formula has a satisfying assignment.*

*Proof.* Consider a random assignment were each variable is set to true with probability 1/2 independently. Let $E_i$ be the bad event that $i$-th clause is not satisfied by the random assignment. Since each clause is a disjunction of $k$ variables, this event happens with probability $p = \Pr(E_i) = \frac{1}{2^k}$ . Note that the event $E_i$ is mutually independent with other events that do not share variables with $E_i$. So, the maximum degree $d$ in the dependency graph is at most:

$$kT \leq k\left(\frac{2^k}{4k}\right) = 2^{k-2}.$$

Since $4dp \leq 4(2^{k-2})(2^{-k}) = 1$, there is an outcome with no bad events , hence a satisfying assignment. □

**Example 4.2 (Disjoint Paths).** Given a graph with $k$ pairs $\{(s_1, t_1), (s_2, t_2), \ldots, (s_k, t_k)\}$, we would like to find a path $p_i$ connecting $s_i$ and $t_i$ such that the paths $\{p_1, p_2, \ldots, p_k\}$ are edge-disjoint. This problem is NP-complete, but we can use the local lemma to show that there is always a solution if the possible paths do not share too many edges with each other.

**Theorem 18.** *For each $1 \leq i \leq k$ let $P_i$ be collection of $L$ paths connecting $s_i$ and $t_i$. Suppose each path in $P_i$ does not share edges with more than $C$ paths in $P_j$ for $i \neq j$ and $\frac{8kC}{L} \leq 1$. Then there is a way to choose $p_i \in P_i$ so that the paths $\{p_i, \ldots, p_k\}$ are edge-disjoint.*

*Proof.* Consider a random experiment that for $1 \leq i \leq k$ , we choose a random path $p_i \in P_i$ connecting $s_i$ and $t_i$. Let $E_{ij}$ be the bad event that $p_i$ and $p_j$ are not edge-disjoint. Since a path in $P_i$ share edges with at most $C$ paths in $P_j$ and there are $L$ paths in $P_i$, we have $p = Pr(E_{ij}) \leq \frac{C}{L}$.

Now, since $E_{ij}$ is mutually independent with all the other events, we have that the maximum degree $d$ in the dependency graph is at most $2^k$. As $4dp \leq 4 \times 2k \times \frac{C}{L} \leq 1$ by our assumption, the local lemma implies that there is an outcome of the experiment with no bad events, hence an edge-disjoint path solution. □

Now that we have seen the applications of the Lovazs Local Lemma, lets get into the details of the proof.

*Proof.* We prove that $\Pr\left(\bigcap_{i \in S} \bar{E}_i\right) > 0$ by induction on $|S|$. To prove this, an intermediate step is required. We need to prove that if $|S| < s$, then for all $k \notin S$ we have

$$\Pr\left(E_k | \bigcap_{j \in S} \bar{E}_j\right) \leq 2p$$

We will prove the intermediate step also using induction. Note that the base $|S| = 1$ can be seen from the fact that $\Pr\left(\bar{E}_i\right) = 1 - \Pr\left(E_i\right) = 1 - p > 0$. For $|S| > 1$:

For the inductive step we assume that $S = \{1, 2, \ldots, l\}$, then:

$$\Pr\left(\bigcap_{i=1}^{l} \bar{E}_i\right) = \prod_{i=1}^{l} \Pr\left(\bar{E}_i | \bigcap_{j=1}^{i-1} \bar{E}_j\right)$$

$$= \prod_{i=1}^{l}\left(1 - \Pr\left(E_i | \bigcap_{j=1}^{i-1} \bar{E}_j\right)\right)$$

At this point we can use the induction hypothesis to see that:

$$\geq \prod_{i=1}^{s}(1 - 2p) > 0$$

Next, let $S_1 = \{j \in S | (k, j) \in E\}$ and $S_2 = S - S_1$. Note that if $|S| = |S_2|$, the $E_k$ is mutually independent with all the other events in $S$ and $\Pr\left(E_k | \bigcap_{j \in S} \bar{E}_j\right) = \Pr\left(E_k\right) \leq p$

It is then safe to continue with the case $|S_2| < s$. Let $F_S = \bigcap_{i \in S} \overline{E_i}$, $F_{S_1} = \bigcap_{i \in S_1} \overline{E_i}$ and $F_{S_2} = \bigcap_{i \in S_2} \overline{E_i}$. Then, applying the definition of conditional probability:

$$\begin{aligned}
\Pr\left(E_k | F_S\right) &= \frac{\Pr\left(E_k \cap F_S\right)}{\Pr\left(F_S\right)} \\
&= \frac{\Pr\left(E_k \cap F_{S_1} \cap F_{S_2}\right)}{\Pr\left(F_S\right)} \\
&= \frac{\Pr\left(E_k \cap F_{s_1} | F_{s_2}\right) \Pr\left(F_{s_2}\right)}{\Pr\left(F_{s_1} | F_{s_2}\right) \Pr\left(F_{s_2}\right)} \\
&= \frac{\Pr\left(E_k \cap F_{s_1} | F_{s_2}\right)}{\Pr\left(F_{s_1} | F_{s_2}\right)}
\end{aligned}$$

We can bound the numerator of the previous expression using the fact that the probability of an intersection is always less that the probability of any of the events in addition with the fact that $E_k$ is independent of the events in $S_2$

$$\Pr\left(E_k \cap F_{S_1} | F_{S_2}\right) \leq \Pr\left(E_k | F_{S_2}\right) = \Pr\left(E_k\right) \leq p$$

For the denominator we can establish a lower bound using the fact that $|S_1| \leq d$:

$$
\begin{aligned}
\Pr\left(F_{S_1}|F_{S_2}\right) &= \Pr\left(\bigcap_{i\in S_1}\bar{E}_i|\bigcap_{j\in S_2}\bar{E}_j\right)\\
&\geq 1-\sum_{i\in S_1}\Pr\left(E_i|\bigcap_{j\in S_2}\bar{E}_j\right)\\
&\geq 1-\sum_{i\in S_1}2p\\
&\geq 1-2pd\\
&\geq \frac{1}{2}
\end{aligned}
$$

Using the upper bound for the numerator and the lower bound for the denominator, we prove the induction:

$$
\begin{aligned}
\Pr\left(E_k|F_S\right) &= \frac{\Pr\left(E_k\cap F_{S_1}|F_{S_2}\right)}{\Pr\left(F_{S_1}|F_{S_2}\right)}\\
&\leq \frac{p}{1/2}=2p
\end{aligned}
$$

(Note that $2p \geq 1-d(2p)$. Using the fact that $4dp \leq 1$, we get $2p \geq 1-d(2p) \leq \frac{1}{2}$)

Then the theorem follows from:

$$
\begin{aligned}
\Pr\left(\bigcap_{i=1}^{n}\bar{E}_i\right) &= \prod_{i=1}^{n}\Pr\left(\bar{E}_i|\bigcap_{j=1}^{i-1}\bar{E}_j\right)\\
&= \prod_{i=1}^{n}\left(1-\Pr\left(E_i|\bigcap_{j=1}^{i-1}\bar{E}_j\right)\right)\\
&\geq \prod_{i=1}^{n}(1-2p)>0
\end{aligned}
$$

□

## References

[HK14] John Hopcroft and Ravindran Kannan. *Foundations of Data Science*. 2014.

[MU17] Michael Mitzenmacher and Eli Upfal. *Probability and Computing*. Cambridge University press, second edition, 2017.

[Nog00] Joel H. Spencer Noga Alon. *The Probabilistic Method*. Wiley, 2000.