# Probablistic methods as a proof technique

Garvit Juniwal 08005008
Pritish Kamath 08005004
Nishant Totla 08005028

March 7, 2012

## 1 Introduction

The *probabilistic method* is a nonconstructive method, primarily used in combinatorics, was pioneered by Paul Erdös, for proving the existence of a prescribed kinds of mathematical objects.

Suppose we want to show that an object of a specific class with a certain property exists. Often it might be difficult to show the existence by explicitly constructing such an object. In the light of this difficulty, the main idea of *probabilistic method* is that if one randomly chooses objects from a specified class, such that the probability of the result having the said property is more than zero, then an object with the said property exists! Note that although the proof uses probability, the final conclusion is determined for certain.

In this article, we demonstrate the use of probabilistic methods using two problems as described in sections 2 and 3.

## 2 Girth vs. Chromatic Number

### 2.1 Preliminaries

We review some graph theoretic definitions,

**Definition 1** (Girth)**.** *For a graph $G$, the girth of $G$ is defined as the size of the smallest cycle in $G$.*

**Definition 2** (Chromatic Number)**.** *For a graph $G$, the chromatic number $\chi(G)$ is the minimum number of colours needed to colour the vertices of $G$, such that adjacent vertices get different colours.*

### 2.2 Motivation

An apriori look at the two definitions would suggest that there might be a trade-off between girth and chromatic number of a graph. That is, if a graph has a high chromatic number, then it would suggest that the graph should be highly connected and hence the graph must have a *small* cycle. Conversely, if the girth of the graph is high, then it suggests that graph is loosely spread out and should have a low chromatic number. So, a natural question to ask is whether there exist graphs with high girth as well as high chromatic number?

| | Low girth | High girth |
|---|---|---|
| Low $\chi$ | Small Graphs | Large cycles |
| High $\chi$ | Cliques | ? |

It is not easy to construct graphs which have high girth and high chromatic number simultaneously. However, using probabilistic methods we show that infact such graphs do exist!

### 2.3 The main result

We prove the following theorem,

**Theorem 1** (Girth versus Chromatic Number)**.** *For every $g$, $k$ there exists a graph of* girth *more than $g$ and* chromatic number *more than $k$.*

*Proof.* Consider a random graph $G_{n,p}$ with $n$ vertices and probability of edge between any two vertices as $p$. First, we prove the following claim:

**Claim 1.** *If a graph $G$ with $n$ vertices does not have any independent set of size $\frac{n}{\chi}$ i.e. all the independent sets are smaller, then it cannot be colored with $\chi$ colors.*

*Proof.* We prove the contrapositive. Let it is possible to color the graph with $\chi$ colors. Then there exists a color such that atleast $\frac{n}{\chi}$ are of that color by the pigeon hole principle. This set of vertices would clearly form an independent set of size at least $\frac{n}{\chi}$. $\square$

So, to show that chromatic number is more than $k$, it suffices to show that the graph does not have any independent set of size $\frac{n}{k}$. Our proof proceeds in two parts. First, we show that for some choice for $n$ and $p$ more than half of the probability density of the graphs does not have any independent set of size $\frac{n}{2k}$ (it was sufficient to show for $\frac{n}{k}$ but we show this stronger result for some technical reason which will be clear later). Second, we show that for the same choice of $n$ and $p$, more than half probability density of graphs have girth more than $g$. This means there exists a graph satisfying both the properties.

$$\text{Pr}(G_{n,p} \text{ has an independent set of size } \tfrac{n}{2k}) \leq \binom{n}{\frac{n}{2k}}(1-p)^{\binom{\frac{n}{2k}}{2}} < 2^n \times e^{-p\frac{n^2}{8k^2}}$$

Therefore if we take $p = n^{\epsilon-1}$, for some positive $\epsilon$, we have $p\frac{n^2}{8k^2} = -\frac{n^{\epsilon+1}}{8k^2}$, and therefore, $2^n \times e^{-p\frac{n^2}{8k^2}}$ = exp((ln 2)$n$ - $(\frac{1}{8k^2})n^{1+\epsilon}$) tends to zero as $n$ tends to infinity. Therefore for $n$ sufficiently large, the probability is less than $\frac{1}{2}$.

Now, we use linearity of expectation to compute the expected number of cycles of length less than or equal to $g$ in $G_{n,p}$. We denote this number by $X$.

$$E(X) = \sum_{i=3}^{g} \frac{n!}{2i(n-i)!}p^i < \sum_{i=3}^{g}(np)^i < g(np)^g = gn^{\epsilon g} \tag{1}$$

Note that we have assumed $np > 1$ in the above bound and we see that this is the case with our choice of $n$ and $p$ later. Now if we choose $\epsilon$ so that $\epsilon g < 1$, we have E$(X) = o(n)$. Thus for sufficiently large $n$, the expected number of cycles of length atmost $g$ is less than $\frac{n}{4}$, and hence by Markov's inequality,

$$Pr(X \geq \frac{n}{2}) < \frac{1}{2} \tag{2}$$

Therefore if we choose n sufficiently large, and $p = n^{\epsilon-1}$, where $0 < \epsilon < \frac{1}{g}$ (Note that we are fixing our choice of $n$ and $p$ here. This choice also works for the previous bound), the probability that $G_{n,p}$ has an independent set of size $\frac{n}{2k}$, or the number of cycles of length atmost $g$ in it is atleast $\frac{n}{2}$, is less than 1. Thus, there is a graph $G$ without any independent set of size $\frac{n}{2k}$ and with less than $\frac{n}{2}$ cycles of length at most $g$. Now we delete one vertex from each cycle of length atmost $g$ in $G$, and call the resulting graph $G'$. $G'$ now has at least $\frac{n}{2}$ vertices and no independent set of size $\frac{\frac{n}{2}}{k}$. By the previous claim, the chromatic number of this graph is more than $k$. $\square$

# 3 Discrepancy of Hypergraphs

## 3.1 Preliminaries

Hypergraphs are a natural generalization of graphs that we normally study.

**Definition 3** (Hypergraphs)**.** *$G = (V, E)$ is a hypergraph, where,*

- *$V$ is a set of vertices*
- *$E \subseteq \mathcal{P}(V)$; edges are subsets of $V$*

*$G$ is said to be $n$-uniform if $|e| = n$ for every $e \in E$.*

The notion of colouring of graphs can be extended to hypergraphs as follows,

**Definition 4** (Colouring of hypergraphs)**.** *Consider hypergraph $G = (V, E)$. Let $\mathcal{C} = \{c_1, c_2, \cdots, c_m\}$ be a set of colours. We call $\Psi : V \to \mathcal{C}$ a colouring of $G$.*
*$\Psi$ is said to be* valid*, if for every $e \in E$, there exists $v_1, v_2 \in e$ such that $\Psi(v_1) \neq \Psi(v_2)$*

Suppose a certain $n$-uniform hypergraph has a valid 2-colouring. All we know from this statement is that for every edge $e \in E$, there exist two vertices in $e$ which are of different colour. It could be that $n-1$ out of $n$ vertices in $e$ have same colour and the remaining one has a different colour; or it could be that roughly $n/2$ vertices in $e$ are of one colour and the remaining are of the other colour. The notion of valid colouring does not distinguish between these two colourings. So it is interesting to have a quantitative version of 2-colouring, which tells more about the proportion in which colours are present in the edges. The notion of *discrepancy* is one such quantitative characterization of 2-colouring.

**Definition 5** (Discrepancy). *Suppose $\Psi$ is a 2-colouring of a $n$-uniform hypergraph $G = (V, E)$. The discrepancy of an edge $e \in E$ in this colouring is,*

$$disc_{\Psi}(e) = |\#1(e) - \#2(e)|$$

*The discrepancy of the colouring is,*

$$disc_{\Psi}(G) = \max_{e \in E} disc_{\Psi}(e)$$

## 3.2 The main result

Before stating the theorem, we state a version of the Chernoff bound that we would use in it's proof. This version is infact weaker than the one we studied in class, but suffices for the theorem we intend to prove.

**Lemma 1** (Chernoff Bound). *For a binomial random variable $BIN(n,p)$, with parameters $n$ and $p$, and any $0 \leq t \leq np$,*

$$\mathbf{Pr}(|BIN(n,p) - np| > t) < 2e^{-t^2/3np}$$

We now state the main theorem,

**Theorem 2.** *Every n-uniform hypergraph H with n edges has a 2-colouring with discrepancy at most $\sqrt{8nln(n)}$*

*Proof.* Consider a random 2-colouring of H. We shall show that with positive probability, the discrepancy of such a random colouring is at most $\sqrt{8nln(n)}$. Equivalently, we show that the probablility that the discrepancy is greater than $\sqrt{8nln(n)}$ is strictly less than 1. Notice that the discrepancy of a 2-colouring is defined as the maximum discrepancy taken over all edges. Thus, to stay below $\sqrt{8nln(n)}$, each individual edge must have discrepancy no more than $\sqrt{8nln(n)}$. The proof goes in two steps:

1. We first show that the probability that a fixed edge $e$ has discrepancy $> \sqrt{8nln(n)}$ is strictly less than $\frac{1}{n}$.

2. We then use union bound and show that the probability that every edge has discrepancy $< \sqrt{8nln(n)}$ is strictly less than 1.

Consider a fixed edge $e$ of $H$. $e$ has $n$ vertices. Let $X$ be a random variable that denotes the number of vertices of $e$ which get the colour 1 in the random colouring. Clearly $X = \text{BIN}(n, \frac{1}{2})$. Using Chernoff Bound with $t = \sqrt{2nln(n)}$ gives

$$\mathbf{Pr}(|X - \frac{n}{2}| > t) < 2e^{-\frac{t^2}{3n/2}} = 2e^{-\frac{4}{3}ln(n)} = 2n^{-\frac{4}{3}} < \frac{1}{n}$$

Note that the discrepancy of $e$ is given by $|X - (n - X)| = |2X - n|$. So discrepancy $> \sqrt{8nln(n)}$ is equivalent to $|2X - n| > \sqrt{8nln(n)} \Leftrightarrow |X - \frac{n}{2}| > \sqrt{2nln(n)}$, and by the above inequality, this probability is $< \frac{1}{n}$. Using the union bound then gives

$$\mathbf{Pr}(\exists e \in H \cdot discrepancy(e) > \sqrt{8n.ln(n)}) \leq \sum_{e \in H} \mathbf{Pr}(discrepancy(e) > \sqrt{8n.ln(n)}) < n \cdot \frac{1}{n} = 1$$

This completes the proof. $\square$

# 4 Conclusion

So as can be seen from the preceding proofs, computing statistical data about the problem at hand was a much easier task than combinatorially proving the existence of a certain object. Additionally, the statistical data had certain implications for individual object data as well, and this is the property we exploited.

But also, it is a question worth asking, whether a certain problem that can be solved using probabilistic methods can also be solved using more traditional (combinatorial) techniques? We note that in most cases, it might indeed be possible to obtain such a solution. As an example, consider the following toy scenario:

Consider a set of objects $S = \{x_1, x_2, \cdots, x_n\}$ and a function $f : S \to \mathbb{R}$, where $f$ has a certain form (known). Now we want to show that there exists some $x_i \in S$ such that $f(x_i) \leq M$, for some given constant $M$. Instead of analytically finding such an $x_i$, if we could find the expected value of $f(X)$ where $X$ is the random variable that takes values from $S$ (according to some probability distribution on $S$), and show that $E[f(X)] \leq M$, then we can claim that there indeed exists such an $x_i$. Note that expectation is essentially a weighted average, but computing expectation itself might be relatively easier, because we could use various simplifying ideas such as linearity of expectation etc. But even analytically, one could (in theory) compute the average and demonstrate the same idea. The crux is that probabilistic methods are easier to handle.