概率方法

徐子昊

2018 年 6 月 18 日

目录

1 引入

本扩展材料由图论组合课程“概率方法”一节的课程内容整理而来。在这里感谢陈晓敏老师的精彩讲解！

1.1 定义

所谓“概率方法”，其本质就是将一个证明存在性的问题转化为证明该事件出现的概率大于 0。听起来似乎非常抽象？那就用下面这个例子给大家一个直观的印象。

1.2 开胃菜

问题： 在一个图 G 中，有 n 个点和 m 条边。问是否存在一种分割方法将点集 V 分割成两个集合 A，B，使得 A，B 这两个点集之间的边数大于等于 $m/2$？

证明： （采用概率方法）：我们可以对这一点集做一个随机染色，将其染为红（R）蓝（B）两色（即将点集 V 映射到集合 R, B）。概率空间 ω 即为所有可能的染色方案所组成的集合，$|\omega| = 2^n$。
对每条边 e，定义随机变量 X_e：

$$X_e = \begin{cases} 1 & \text{若 e 的两端不同色} \\ 0 & \text{若 e 的两端同色} \end{cases}$$

则 $E[X_e] = P(\text{e 的两端不同色}) = \frac{1}{2}$
（其中“e 的两端颜色不同”可以看作是一种好的性质，或者说是我们证明所需要的一种性质）
定义随机变量 $X = \sum\limits_{e} X_e$，X 的实际意义即为: 在这样的一个随机染色后确定的图中，两端颜色不同的边的数量。（即“好”边的数量）
则 $E[X] = \sum\limits_{e} E[X_e] = \frac{m}{2}$。由于 X 的期望为 $\frac{m}{2}$，所以可以肯定，至少存在一种染色方案 C 使得整个图中两端颜色不同的边数大于等于 $\frac{m}{2}$。我们就可以这样构造点集 A 和 B：采用上述染色方案 C，A 是所有红色点构成的集合，而 B 是所有蓝色点构成的集合，从而命题得证。

评注 概率方法通常都会出现某种“好的性质”，比如上题中“好的性质”就是“一条边的两个端点的颜色不同”。这种“好的性质”和我们的证明目标密切相关。

在看过了这道“开胃菜”之后，让我们来看下对于概率方法更复杂应用。

2 其他例子

2.1 例一：Sum Free 问题

不知道大家对于概率论第一节课上的一道“Sum Free”问题还有没有印象？想不起来也没有关系，下面会复述这道题的题面。这道题也可以用概率方法来解决，其基本想法是一致的。

问题： 如果 $\forall\ i,j,k \le |A|, \nexists\ \ a_i, a_j, a_k\ \ s.t.\ \ a_i + a_j = a_k$, 我们称 A 为 Sum-Free 集。尝试证明：任意一个由非 0 整数构成的集合 $A = \{a_1, \cdots, a_n\}$ 包含一个大小大于 $\frac{n}{3}$ 的 Sum Free 子集。

问题解答：

引理 集合 $\{1, 2, \cdots, 3k+1\}$ 的子集 $\{k+1, k+2, \cdots, 2k+1\}$ 为 Sum-Free。
关于这条引理的证明，可参考李怿铠同学的第一堂课笔记。

本问题证明： 取一大数 p，满足 $p = 3k+2$ 并且 $p \nmid b_1 \cdots b_n$。随机取定一个 $d \in [p-1]$，定义 $G = [k+1, 2k+1]$。对 $\forall a \in A$，定义随机变量 X_a 为：

$$X_a = \begin{cases} 1 & ad \mod p \in G \\ 0 & ad \mod p \notin G \end{cases}$$

由于 $\{ad \mid \text{a 为 A 中元素且} d \text{取遍}[p-1]\ \}$ 构成模 p 的完全剩余类，所以可得 $P(ad \mod p \in G) = \frac{k+1}{3k+1} > \frac{1}{3}$。从而有 $E[X_a] > \frac{1}{3}$。
定义随机变量 $X = \sum\limits_{a \in A} X_a$，则 X 的实际含义是：对于一个随机选定的 d，集合 A 中满足 $ad \mod p \in G$ 的元素个数。从而有 $E[X] = \sum\limits_{a \in A} E[X_a] > \frac{n}{3}$。既然个数的期望都大于 $\frac{n}{3}$ 了，则我们必定能找到一个 d，使得集合 A 中满足 $ad \mod p \in G$ 的元素个数大于 $\frac{n}{3}$。将上述元素组成集合 B，则该集合满足“Sum Free”性质且个数大于 $\frac{n}{3}$，命题得证。（关于为什么 B 是“Sum Free”的，可以采用反证法并且用到上述引理，在这里就不叙述了）

2.2 例二：Ramsey 问题

人和人之间的交往总是能给图论组合学者以新问题的灵感，而 Ramsey 问题正是由此而来。Ramsey 问题可以表述为：
寻找最小的自然数 $N = r(y, b)$, 使得 N 个人中必定有 y 个人互相认识或 b 个人互不相识。（$y \ge 3, b \ge 3$，且假设人和人之间相识、不相识两种关系必居其一，且不出现 A 认识 B 而 B 不认识 A 的情况）

将上述问题抽象化，可以表述为：寻找最小的自然数 N，使得对一个含有 N 个点的完全图 G，对其边进行黄蓝二染色，无论怎么染色，图 G 都至少含有一个全部由黄色边构成的含 y 个点的完全图，或者一个全部由蓝色边构成的含 b 个点的完全图。
但今天我们不是直接去解 Ramsey 问题，而是要去证明某一类 Ramsey 问题的下界，即：

$$r(k,k) > \sqrt{2}^k, \quad k \in N$$

问题解答 将本问题翻译一下，就是：对一个含有 $\sqrt{2}^k$ 个点的完全图的边进行黄蓝染色，存在一种染色方案，既不含有 k 个点的蓝色完全图，也不含有 k 个点的黄色完全图。

证明 采用概率方法：
将整张图上的边随机染色，每条边染成某种颜色的概率均为 $\frac{1}{2}$。设 $N = \sqrt{2}^k$。设 T 是从该图的点集中选取的一个 k 元子集。T 中所有点之间的连边和这些点构成了子图 G_T。定义随机变量 X_T：

$$X_T = \begin{cases} 1 & G_T \text{ 中边同色} \\ 0 & G_T \text{ 中边颜色不完全相同} \end{cases}$$

由于每条边都是随机染色，所以 $E[X_T] = 2 * 2^{-\binom{k}{2}} + 0 = 2^{1-\binom{k}{2}}$。
定义随机变量 $X = \sum\limits_T X_T$（枚举所有可能的 T），则 X 的实际含义是：在一种确定的随机染色方案下，该图包含的满足点数要求的全同色边子图的个数。于是 $E[x] = \sum\limits_T E[X_T] = \binom{N}{k} 2^{1-\binom{k}{2}}$。由于 $N = \sqrt{2}^k$，所以：

$$\binom{N}{k} 2^{1-\binom{k}{2}} < \frac{N^k}{k!} 2^{1-\binom{k}{2}} = \frac{2^{\frac{k}{2}+1}}{k!} < 1$$

由于期望都小于 1，且 X 为整数，所以至少存在一个染色方案，使得 $X = 0$。命题得证。

2.3 例三：Ramsey 问题的继续探索

一个更有意思的问题，是使用概率方法探索 $r(k,4)$ 这一类问题的下界。先给出结果：

$$r(k,4) \approx k^{1.5-\epsilon}, \quad \epsilon \in Constant, \text{当 k 足够大}$$

在这里，为了取得这一结果，我们需要一些技巧。在上述几题中，无论是对点集随机二染色，还是对边集随机二染色，我们染色概率的分布律都是均匀的。但在这到题中，

对于图中两种边的需要是不一样的。于是，我们便可以在随机二染色的分布律上做文章。

证明　设这张完全图有 N 个点。依旧是对这张图的边进行随机二染色，每一条边染黄色的概率为 p，染蓝色的边的概率为 $1-p$。
设 T 是从该图的点集中选取的一个 k 元子集。T 中所有点之间的连边和这些点构成了子图 G_T。定义随机变量 X_T：

$$X_T=\begin{cases}1 & G_T\text{ 中边全是黄色}\\0 & G_T\text{ 中边不全是黄色}\end{cases}$$

定义随机变量 Y_T：

$$Y_T=\begin{cases}1 & G_T\text{ 中边全是蓝色}\\0 & G_T\text{ 中边不全是蓝色}\end{cases}$$

类似上一问题，我们可以得到 $E[X_T]=p^{\binom{k}{2}}$，$E[Y_T]=(1-p)^6$。
定义随机变量 $Z=\sum\limits_T X_T+\sum\limits_T Y_T$，则 Z 的实际依旧含义是：在一种确定的随机染色方案下，该图包含的满足点数要求的全同色边子图的个数。于是：

$$\begin{aligned}E[Z]&=\sum_T E[X_T]+\sum_T E[Y_T]\\&=\binom{N}{k}p^{\binom{k}{2}}+\binom{N}{4}(1-p)^6\end{aligned}\tag{1}$$

为了能够将上式凑成常数，并设法使上式小于 1，我们不妨做以下尝试：

$$1-p\approx cN^{-\frac{2}{3}},\quad c>0\quad\text{（因为我们想让 }\binom{N}{4}(1-p)^6\text{ 项近乎常数）}$$

于是对前一项 $\binom{N}{k}p^{\binom{k}{2}}$，有：

$$\begin{aligned}\binom{N}{k}p^{\binom{k}{2}}&\approx\frac{N^k}{k!}(1-cN^{-\frac{2}{3}})^{\binom{k}{2}}\\&\approx\frac{N^k}{k!}e^{-0.5cN^{-\frac{2}{3}}(k^2-k)}\\&=\frac{N^k}{k!}e^{-0.5cN^{-\frac{2}{3}}k^2}e^{0.5cN^{-\frac{2}{3}}k}\end{aligned}\tag{2}$$

为保证其近似为常数，则应该使其唯一有可能大于 1 的一项趋于常数，即：

$$cN^{-\frac{2}{3}}k\approx C$$

从而有：

$$N \approx k^{1.5-\epsilon}$$

命题得证。

上述方法可能在细节上需要推敲，希望同学们能一起把它完善。

如有错误请与我联系，非常感谢！

邮箱：shsjxzh@sjtu.edu.cn