图论中的概率方法

郭睿涵

July. 01, 2020

Summary 因为这学期学习了概率论和图论的知识，我在课外时间读了alon和spencer的概率方法教材的前九章内容，总结了一下概率方法的本质，并选出了若干道比较有意义的习题。

概率方法是什么 可以用这样一句话来简单概括概率方法：尝试去证明一个结构具有某个特定的性质，那么就证明这个性质存在的概率是大于0的。

例子 关于Ramsey数的估计
如果$\binom{n}{k}\times 2^{1-\binom{k}{2}} < 1$，那么$R(k,k) > n$。同时我们得到$k \geq 3$时，有$R(k,k) > \lfloor 2^{\frac{k}{2}} \rfloor$。

证明 考虑一个边随机染色的大小为n的完全图K_n，每条边被二染色，要么染成红色，要么染成蓝色，二者概率相等。对任意一个大小为k的点集R，设A_R为K_n的R导出子图为同色的事件。显然$P(A_R) = 2^{1-\binom{k}{2}}$。又因为图中有$\binom{n}{k}$个可能的R集合，所以有这么多种可能的同色诱导子图，有一个大小为k的诱导子图的概率为

$$\binom{n}{k} \times 2^{1-\binom{k}{2}}$$

如果让前面的式子概率小于1，那么就存在一种二染色，使得K_n中没有K_k同色诱导子图。我们取$n = \lfloor 2^{\frac{k}{2}} \rfloor$，就得到

$$\binom{n}{k} \times 2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}} n^k}{k! 2^{\frac{k^2}{2}}} < 1$$

由此我们得到$R(k,k) > \lfloor 2^{\frac{k}{2}} \rfloor$

上面的例子就是概率方法的核心思想体现。

历史 早在1943年，Erdos就认识到了这个方法的强大，并将其应用到了众多的问题中。这个例子出现在 P.Erdos 在1947年的论文中，（P. Erdos 1947, Some remarks on the theory of graphs）。

我们为什么要学习概率方法 因为在我们讨论的大部分问题中，概率问题都是一个有限的空间。这个性质导致了概率方法在离散数学领域里面的众多应用。而这正是概率方法在组合与图论中大放异彩的原因。而且据我了解，到目前为止，很多概率方法的内容没有其他方法可以代替的，比如 Lovasz Local Lemma。

同样的，算法层面上来讲，概率方法也是一个有用的工具。比如我们前面给出的例题，我们知道了K_n里面有一个大小为K_{2log_2n}的同色导出子图。那么，我们是不是可以尝试去找出来它呢？据我自己的了解，这个问题应该只有n阶乘复杂度的搜索算法可以做到。那么这也就说明，这个问题不是可以通过计算机的简单枚举来证明的。但是仔细思考上面的证明，我们可以看到当$n = \lfloor 2^{\frac{k}{2}} \rfloor$的时候，

$$\binom{n}{k} \times 2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!}(\frac{n}{2^{\frac{k}{2}}})^k \leq \frac{2^{1+\frac{k}{2}}}{k!} \ll 1$$

上面的式子说明了什么呢？如果我们对K_n随机染色的话，非常可能就找到了一个同色的K_{2log_2n}。随机产生一个图K_n，那么存在K_{2log_2n}的概率比$\frac{2^{\frac{log_2n}{2}}}{(log_2n)!}$还要小。这么小的概率，在数据量比较大的时候，我们就可以在算法里面忽略它了！

下面是我在自学过程中看到的一些有代表性的例子

Erdos-Ko-Rado定理 对一系列两两相交的集合构成的集合族$\mathcal{F}$，$\forall A, B \in \mathcal{F}, A \cap B \neq$。假定$n \geq 2k$，$\mathcal{F}$中的每个元素都是含有k个元素的集合。假设全集大小为n。 Erdos-Ko-Rado 定理说$|\mathcal{F}| \leq \binom{n-1}{k-1}$。这个构造可以通过选取所有包含同一个元素的集合来得到。下面是Katona在1972年给出的一个简短的证明。

Lemma 对$0 \leq s \leq n-1$,令$A_s = s, s+1, \ldots, s+k-1$，其中这个加法是模n意义下的加法。那么$\mathcal{F}$中最多有k个这样的集合$A_s$。这个的证明是非常显然的，否则就无法保证更多的$A_i$两两相交了。

Proof 设σ为一个$\{0, 1, \ldots, n = 1\}$的排列，$i \in \{0, 1, \ldots, n = 1\}$，是被随机选取的。令$A = \{\sigma(i), \sigma(i+1), \ldots, \sigma(i+k-1)\}$，这里的加法也是对n取模的加法。考虑$A \in \mathcal{F}$的概率，我们显然得到$P(A \in \mathcal{F}) \leq \frac{k}{n}$。但是A是随机的从所有大小为k的集合里面取出来的，因此我们有

$$\frac{|\mathcal{F}|}{\binom{n}{k}} = P(A \in \mathcal{F}) \leq \frac{k}{n}$$

这也就得出

$$|\mathcal{F}| \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$$

Problem 1. Erdos 1965
令$A = a_1, \ldots, a_n$是n个非零整数。证明存在A的一个子集B,使得$|B| > \frac{n}{3}$而且B中不存在a,b,c，满足$a + b = c$

Proof 选择一个大质数p，形如$p = 3k + 2$。考虑一个n行p-1列的矩阵。我们对(i, x)这个位置做一个标记，当且仅当$a_i x$在$k+1$和$2k+1$之间。
每行的数字都是不同的，因此，每行都至少有$\frac{1}{3}$的格子被做了标记。因此，这里存在一列，至少有$\frac{1}{3}$被选中。也就是说，存在一个数x满足

$$B = a_i | k+1 \leq a_i x \leq 2k+1$$

这个集合有多于$\frac{n}{3}$个元素。我们声明不存在a,b,c满足$a + b = c$，否则一定存在$ax + ay = az (\% p)$，这和前面的题设矛盾了。

Problem 2. 令$G = (V, E)$为一个有n个点，m条边的图，我们总是可以将其中的点分成两个集合，是的两个集合之间的边至少有$\frac{m}{2}$条。

Proof 对每条点随机染成红蓝两种颜色，所以这里有一个大小为2^n的概率空间。考虑对每一条边，定义一个单独的随机变量X_e。如果e这条边连接的两个端点的颜色相同，那么$X_e = 0$，否则$X_e = 1$。再定义随机变量

$$X = \sigma_{e \in E} X_e$$

作为红色集合和蓝色集合之间的期望边数。因为$E(X_e)=\frac{1}{2}$，所以$E(X)=\frac{m}{2}$。因此至少存在一种情况满足$X\geq\frac{m}{2}$。

Problem 3. 在一个学校有 1000 个俱乐部，每个俱乐部有 11 个人，俱乐部之间的成员可能会重复。证明，可以将所有学生分成两组，使得每个俱乐部都同时有来自两组的学生。

Solution 使用概率方法。单独考虑每一个人的分配方式，用X_i表示第i组人来自几个集合。$X_i=0$代表第i组人来自2个集合；$X_i=1$代表第i组人来自1个集合。$P_{X_i=1}=\frac{1}{2^{10}}$。考虑求期望，

$$E(\sum_{i=1}^{1000}X_i)=\frac{1000}{2^{10}}<1$$

因此存在一种分组，使得每个X_i都是0，也就是说说每个俱乐部都同时又来自两组的学生。

Problem 4. 一个竞赛图是一个每两个点之间恰好有一条边的有向图。证明：对每个 n，存在一个竞赛图包含至少 $n!/2^{n-1}$ 条哈密尔顿路。

Solution 哈密尔顿路定义：经过每个节点恰好一次
首先一个完全图里面，经过每个节点恰好一次的路径有$n!$条，而且长度都是$n-1$。所以题目的要求就转化为大小为n的竞赛图里面至少包含这么多哈密尔顿路经。考虑无向图中的每条路径，它构成哈密尔顿路的概率为$\frac{1}{2^{n-1}}$，因为每条边都有两种选择，全部情况中只有一种构成可行路径。考虑大小为n的竞赛图中，哈密尔顿路径数的期望

$$E(P)=\sum_{all\ edge}\frac{1}{2^{n-1}}=\frac{n!}{2^{n-1}}$$

根据鸽笼原理，至少有一个大小为n的竞赛图有这么多条哈密尔顿路，题目证毕。

Problem 5. $G=(V,E)$是一个二分图，$|V|=n$，$k>\log_2 n$ 是一个整数。在 G 的每个顶点 v 上有一个候选的颜色集合 $C(v)$满足$|C(v)\geq k|$。证明：可以给每个点 v 选一个 $C(v)$ 中的颜色，使得 G中任意相邻两点不同色。

Solution 不妨先强化条件为完全二分图，这样可以发现如果拓展出来的完全二分图满足，那么原二分图一定满足。

考虑每一个颜色分成两类，分到第一组或者第二组的概率都是$\frac{1}{2}$。题目等价为把颜色分成两类，让左边的每个点的染色集合里至少有一个第一类里面的颜色，右边的每个点的染色集合里至少有一个第二类里面的颜色。

对左边的每个点，考虑它的染色集合$C(v)$里面不包含第一组颜色的概率大于等于$\frac{1}{2^k}$。求期望得到

$$E(X) = \sum_{i=1}^{n} X_i P_i \leq \frac{n}{2^k} < 1$$

因此对左边而言，一定存在一种可行的分组。类似的，对右边也一样。

Problem 6. 图$G = ([1000], E)$中有至多 27000 个三角形。证明：G 中有一个 70 点的不含三角形的诱导子图。

Solution 类似上课思路。考虑随机变量X_i表示第i个三角形的存在情况，$X_i = 1$表示存在，反之表示不存在。考虑从里面选90个点的情况：

$$E(\sum_{i=1}^{27000} X_i) = 27000 \times \frac{\binom{1000-3}{90-3}}{\binom{1000}{90}} = \frac{27000 * 90 * 89 * 88}{1000 * 999 * 998} < 20$$

也就是说，里面至多有19个三角形。对每个三角形删除一个节点，我们就得到了一个至少71个点的不包含三角形的图。再任意删除几个点，就得到了70个点的没有三角形的图。

Problem 7. $G = (V, E)$是一个 n 阶的、20 正则的简单图。证明：存在$W \subseteq V$，使得 $V \setminus W$中任意一个顶点在W 中至少有一个邻居，$|W| \leq n/5$.

Solution （方法是学来的，理解并记住这个方法）

这样定义随机变量，对于一个点，有p的概率选择到集合A中。那么如果对图里的一个点，它和它的邻居都没有被选入A，那么就把它选入集合B中。

$$\begin{aligned}
P(x \in A) &= p \\
P(x \in B) &= (1-p)^{21} \\
P(x \in A \cup B) &= P(x \in A) + P(x \in B) = p + (1-p)^{21} \\
E(A \cup B) &= n(p + (1-p)^{21})
\end{aligned}$$

让$p = 1 - (\frac{1}{100})^{\frac{1}{21}}$（一百分之一可以换成其他东西），那么$E(A \cup B) < \frac{n}{5}$，让$W$取集合$A \cup B$即可满足。