概率方法

参考答案

1. (10 分) 随机图 $G(n,p)$ 是连通的概率是多少?
   (1) 证明存在常数 $\alpha > 0$, 使得对所有充分大的 n, $G(n, \alpha \ln n/n)$ 是连通图的概率不高于 $1/n$.
   (2) 证明存在常数 $\alpha > 0$, 使得对所有充分大的 n, $G(n, \alpha \ln n/n)$ 是连通图的概率不低于 $1 - 1/n$.
   提示: $G(n,p)$ 可以如此采样: 先从二项分布 $\mathrm{Binom}(\binom{n}{2}, p)$ 中采样 m, 再从 $G(n,m)$ 中采样. $G(n,m)$ 可以如此采样: 从没有边的图开始, 每次随机添加一条新边, 重复 m 次.

解

(1) 注意到任意两条边其是否存在的概率是互相独立的, 因此我们可以按照任意顺序进行边的采样. 考虑如下的采样策略实现从 $G(n,p)$ 中采样:

- 首先任意选定一个顶点记作 v_1, 并开始采样 v_1 与其它点是否连边. 具体来说, 用 $X_{u,v} \in \{0,1\}$ 表示 u, v 之间连边的示性函数. 对于每个 $u \neq v_1$, 我们独立地从 $\mathrm{Bern}(p)$ 中采样 $X_{v_1,u}$.
- 采样完 v_1 的邻边之后, 我们任取一个剩下点中当前度数最小的, 记为 v_2, 并采样 v_2 与其它点是否连边. 具体来说, 对于每个 $u \notin \{v_1, v_2\}$, 我们独立地从 $\mathrm{Bern}(p)$ 中采样 $X_{v_2,u}$.
- 更一般地, 当采样完 $v_1, \ldots, v_{i-1}$ 的邻边之后, 我们任取一个剩下点中当前度数最小的, 记为 v_i, 并采样 v_i 与其它点是否连边. 具体来说, 对于每个 $u \notin \{v_1, \ldots, v_{i-1}\}$, 我们独立地从 $\mathrm{Bern}(p)$ 中采样 $X_{v_i,u}$.
- 重复这样的过程直到每对顶点之间是否连边都被采样过.

定义事件 $A_1, \ldots, A_{n-1}$. 其中

$$A_i := \{\forall j > i, X_{v_i, v_{i+1}} = 0\} = \{v_i \text{ 与 } v_{i+1}, \ldots, v_n \text{ 皆无连边}\}.$$

令 $k(n) = n^{1-\delta}$, 其中 $\delta > 0$ 是可以任意小的常数. 定义如下两个事件:

$$\begin{aligned} A &:= A_1 \vee \cdots \vee A_k = \{\exists i \le k, \text{点 } v_i \text{ 与 } v_{i+1}, \ldots, v_n \text{ 皆无连边}\}, \\ B &:= \{v_1, \ldots, v_{k-1} \text{ 一共连出了至多 } n-k \text{ 条边}\}. \end{aligned}$$

注意到 $A \wedge B \implies$ 图中有孤立点 $\implies$ 不连通. 因为 A 推出存在点 v_i ($i \le k$) 与 $v_{i+1}, \ldots, v_n$ 皆无连边. 如果 B 也同时成立, 那么刚采样完 $v_1, \ldots, v_{i-1}$ 时, 总共有不超过 $n-k$ 条边, 此时 $V \setminus \{v_1, \ldots, v_{i-1}\}$ 至少有一个点度数为零. 特别地, 此时 v_i 的度数为零, v_i 与 $v_1, \ldots, v_{i-1}$ 也皆无连边.

由 union bound 知, $\Pr[\text{连通}] \le \Pr[\neg A \vee \neg B] \le \Pr[\neg A] + \Pr[\neg B]$. 我们分别估计 $\neg A$ 和 $\neg B$ 的概率.

估计 $\Pr[\neg A]$ 时, 利用 $A_1, \dots, A_k$ 之间的独立性

$$\Pr[A_i] = (1-p)^{n-i} = \left(1 - \frac{\alpha \ln n}{n}\right)^{n-i} \geq e^{-(1-o(1))\frac{\alpha \ln n}{n}\cdot(n-k)} = n^{-\alpha(1-o(1))},$$
$$\Pr[\neg A] = \Pr[\neg A_1 \wedge \cdots \wedge \neg A_k] = \prod_{i \leq k} \Pr[\neg A_i] \leq (1 - n^{-\alpha(1-o(1))})^k \leq e^{-n^{1-\delta-\alpha-o(1)}}.$$

只要 $\alpha < 1 - \delta$, 就有 $\Pr[\neg A] \leq e^{-n^{\Omega(1)}}$.

注意到 $v_1, \dots, v_{k-1}$ 一共连出的边数服从二项分布 $\mathrm{Binom}((k-1)(n-k/2), p)$, 其期望是

$$(k-1)(n-k/2)p = O(n^{1-\delta} \ln n).$$

依据 multiplicative Chernoff bound, 这部分边数超过 $n-k$ 的概率 $\Pr[\neg B] \leq e^{-n^{\Omega(1)}}$.

Remark: v_1, v_2 之间连边的概率为 p^{n-1}, 远远低于 $p = \alpha \ln n/n$. 造成这个现象的原因, 是因为 v_2 实际上是一个随机变量, v_2 指代哪个点取决于哪些点与 v_1 连边.

(2) 使用提示中的等价的采样策略. 我们考虑在此过程中, 图上的连通分量数目. E.g., 初始时有 n 个孤立点, 故连通分量数目为 n; 采样第一条边之后连通分量为 $n-1$.

假设某个时刻图中的连通分量数目为 $k+1$, 那么采样下一条边之后连通分量数目变为 k 的概率至少为

$$1 - \frac{\binom{n-k}{2}}{\binom{n}{2}} \geq \frac{k}{n}.$$

和第一题进行比较. 从连通分量为 $k+1$ 到连通分量为 k 的次数（在某个 coupling 下）小于第一题中从已集 $n-k-1$ 种卡片到已集 $n-k-1$ 种卡片之间抽的卡片数 $\tau_{n-k} - \tau_{n-k-1}$. 因此连通需要的边数（在某个 coupling 下）小于集齐 n 中卡片需要的抽卡次数.

这样由第一题的结论可以知道当边数 $m = 3n \ln n$ 时, G 连通的概率不低于 $1 - O(1/n^2)$. 只需选取 $\alpha = 4$, 这样 $\mathbb{E}[m] = 4n \ln n$, 由 multiplicative Chernoff bound 知 w.h.p. $m \geq 3n \ln n$, 从而结论成立.

2. (8 分) 用 $G(2n, p, q)$ 表示如下的随机图的分布: 点集为 $V = \{1, 2, \dots, 2n\}$. 选取随机的 $S^* \subseteq V$ 且 $|S^*| = n$. 这样点集被划分为了两个大小相同的块 S 和 $V \setminus S$. 对于任意两个点 (u, v), 如果 u, v 在同一个块中（$u, v \in S^*$ 或者 $u, v \notin S^*$）, 那么 u, v 之间以概率 p 有边相连; 如果 u, v 在不同的块中, 那么 u, v 之间以概率 q 有边相连.

考虑高度稀疏的场景. 令 $p = \alpha/n$, $q = \beta/n$, 其中 $\alpha, \beta \in \mathbb{R}$ 是常数. 请问是否可以依据图本身的信息对 S^* 进行一个非平凡的估计. 具体来说, 是否存在一个算法 $\mathcal{M}$ 和常数 $c > 0$ 使得

$$\Pr\left[|\hat{S}| = n \wedge \frac{|\hat{S} \cap S^*|}{n} \notin (1/2 - c, 1/2 + c) \,\middle|\, \begin{matrix} G \leftarrow G(2n, \alpha/n, \beta/n) \\ \hat{S} \leftarrow \mathcal{M}(G) \end{matrix} \right] = 1 - O(1/n).$$

答案显然依赖于 (α, β) 的取值. 请找到一个尽量大的 (α, β) 的范围使得可以对 S^* 进行上述的非平凡估计.

提示：当 $(\alpha - \beta)^2 < \alpha + \beta$ 时, 不存在任何算法可以非平凡地估计 S^* [Mossel-Neeman-Sly 2012].

解 对任意 $S \subseteq V$, 用 $e(S)$ 表示被被 S 割的边的数目.

$$e(S) := |\{(u,v) \in E | u \in S, v \notin S\}|.$$

考虑一个非常直观的算法: 给定输入 G 之后, 算法输出一个平衡的割 $\hat{S} \subseteq V$, 使得 $e(\hat{S})$ 最大（如果 $\beta > \alpha$）或最小（如果 $\beta < \alpha$）.

我们先考虑 $\beta < \alpha$ 的情况. 只需要证明, 割边最小的平衡割大概率不是一个平凡估计.

- 最非平凡的割, 也就是 ground truth S^*, 期望切割边数 $\mathbb{E}[e(S^*)] = \beta n$. 大概率 $e(S^*)$ 集中在 βn 附近. 具体来说, 对于任意常数 $\delta > 0$, 根据 multiplicative Chernoff bound,

$$\Pr[e(S^*) \geq (\beta + \delta)n] \leq 2^{-\Omega(n)}.$$

- 我们选取一个常数 $c > 0$. 对于任意平衡割 T, 如果满足 $\frac{|T \cap S|}{n} \in (1/2 - c, 1/2 + c)$, 我们称 T 是平凡的. 对任意平凡的 T, 其切割的边数的期望不低于

$$\Big(\frac{\alpha+\beta}{2} - 2c^2(\alpha - \beta)\Big)n = \Big(\frac{\alpha+\beta}{2} - \delta\Big)n.$$

  其中常数 $\delta = 2c^2(\alpha - \beta)$. 通过选取合适的 c, 我们可以让常数 δ 的值任意小.

  根据 multiplicative Chernoff bound,

$$\Pr[e(T) \leq (\beta + \delta)n] \leq e^{-\Big(\frac{\frac{\alpha-\beta}{2} - 2\delta}{\frac{\alpha+\beta}{2} - \delta}\Big)^2 (\frac{\alpha+\beta}{2} - \delta)n/3}.$$

  只要常数 α, β 满足 $(\alpha - \beta)^2/(\alpha + \beta) > 12 \ln 2$, 就可以找到足够小的常数 $\delta > 0$ 使得上面的概率不超过 $2^{-(2+\delta)n}$.

  再使用 union bound, 以 $1 - 2^{-\Omega(n)}$ 的概率, 所有平凡平衡割 T 都满足 $e(T) > (\beta + \delta)n > e(S^*)$.

对称地, 对于 $\alpha > \beta$ 的情况, 只要常数 α, β 满足 $(\alpha - \beta)^2/(\alpha + \beta) > 8 \ln 2$, 以 $1 - 2^{-\Omega(n)}$ 的概率, 所有平凡平衡割 T 都满足 $e(T) < (\beta - \delta)n < e(S^*)$.

3. (8 分) 有若干集合 $A_1, \ldots, A_n, B_1, \ldots, B_n$, 满足 $\forall i, j \;\; A_i \cap B_j = \varnothing \iff i = j$. 证明

$$\sum_i \frac{1}{\binom{|A_i| + |B_i|}{|A_i|}} \leq 1.$$

解 选取一个 $\bigcup_i (A_i \cup B_i)$ 上随机的序. 定义事件 E_i 为 "A_i 中每一个元素小于 B_i 中每一个元素". 对于每一个 i, 有 $\Pr[E_i] = 1/\binom{|A_i| + |B_i|}{|A_i|}$.

用反证法. 如果

$$\sum_i \Pr[E_i] = \sum_i \frac{1}{\binom{|A_i| + |B_i|}{|A_i|}} > 1,$$

那么存在 $i \neq j$ 使得 E_i, E_j 有可能同时发生. 考虑使 E_i, E_j 同时发生的一个序. 在这个序下, $\max(A_i) < \min(B_i), \max(A_j) < \min(B_j)$. 因此必有 $\max(A_i) < \min(B_j)$ 或者 $\max(A_j) < \min(B_i)$.

这两种可能分别都会推出矛盾.

4. (8 分) 如果集合 $S \subseteq \{0,1\}^n$ 满足

$$\forall \text{distinct } a,b,c \in S,\ \Delta(a,b)+\Delta(b,c) > \Delta(a,c),$$

我们称 S 是不共线的. 这里 Δ 表示汉明距离（Hamming distance）.

(1) 证明, 对所有足够大的 n, 存在不共线的 $S \subseteq \{0,1\}^n$ 满足 $|S| \ge 1.01^n$.

(2) 证明, 对所有足够大的 n, 任何不共线的 $S \subseteq \{0,1\}^n$ 都满足 $|S| \le 1.99^n$.

解 对于任何三个点 $a,b,c \in \{0,1\}^n$, 如果 $\Delta(a,b)+\Delta(b,c) = \Delta(a,c)$, 我们称 a,b,c 共线；否则我们称 a,b,c 不共线.

(1) 如果独立随机选取 $a,b,c \in \{0,1\}^n$, 那么

$$\Pr[a,b,c\text{共线}] = \Pr\Big[\forall i \in [n], (a_i,b_i,c_i) \notin \{(0,1,0),(1,0,1)\}\} = \varnothing\Big] = (3/4)^n.$$

独立随机选取 $x_1,x_2,\dots,x_m \in \{0,1\}^n$, 其中 $m = 2^{\alpha n}$. 令 $S = \{x_1,\dots,x_m\}$. 那么

$$\Pr\big[|S| = m\big] = \Pr\big[\forall \text{distinct } i,j,\ x_i \ne x_j\big] \le \frac{1}{2}2^{2\alpha n}\cdot 2^{-n}.$$

$$\Pr[S\text{中存在共线三元组}] \le \frac{1}{2}2^{3\alpha n}\cdot(3/n)^n.$$

只要 $3\alpha < \log_2(4/3)$, 就有非零概率 $|S| = 2^{\alpha n}$ 且 S 不共线.

(2) 对任何不共线的 S, 构造一个二部图 $G = (L,R,E)$. 左右点集分别为 $L = \{0,1\}^{\lfloor n/2\rfloor}$, $R = \{0,1\}^{\lceil n/2\rceil}$. 边集与 S 一一对应: $(u,v)\in E$ 当且仅当 $u|v \in S$. 这里 $|$ 表示拼接.

图中一定没有长度为 3 的路径. 反证, 假如有长度为 3 的路径 $(u,v),(u',v),(u',v')$, 那么

$$\Delta(u|v,u'|v)+\Delta(u'|v,u'|v') = \Delta(u,u')+\Delta(v,v') = \Delta(u|v,u'|v),$$

说明 $u|v,u'|v,u'|v' \in S$ 这三点共线. 与 S 不共线矛盾.

换言之, 图中任何两个度数大于 1 的点不相邻. 再换言之, 每条边都有至少一个端点的度数等于 1. 所以存在一个从度数为 1 的顶点集到 E 的满射.

$$|S| = |E| \le |L|+|R| = 2^{\lfloor n/2\rfloor}+2^{\lceil n/2\rceil}.$$

5. (5 分) 假设 n 足够大. 证明存在不依赖 n 的常数 $\alpha > 0$, 使得

采样有 αn 条边的随机图 $G \sim G(n,\alpha n)$, 并采样两个不同的随机点 u,v. 那么 u,v 在 G 上联通的概率不超过 $1/n$.

解 不妨先采样 (u,v). 考虑 G 中, 从 u 到 v 的简单 (无重复点) 的路径个数的期望.

$$\begin{aligned}\mathbb{E}[\#\text{paths from } u \text{ to } v] &= \sum_{\ell=1}^{n} \mathbb{E}[\#\text{length-}\ell \text{ paths from } u \text{ to } v] \\ &= \sum_{\ell=1}^{n} \sum_{w_1,\dots,w_{\ell-1}} \Pr[uw_1w_2\dots w_{\ell-1}v \text{ is a path}] \\ &\le \sum_{\ell=1}^{n} (n-2)^{\ell-1}\left(\frac{\alpha n}{\binom{n}{2}}\right)^{\ell} \\ &\le \frac{1}{n}\sum_{\ell=1}^{n}(2\alpha)^{\ell} \\ &= \frac{1}{n}\frac{1}{(2\alpha)^{-1}-1}.\end{aligned}$$

根据 Markov bound,

$$\Pr[u,v \text{ 联通}] = \Pr[\#\text{paths from } u \text{ to } v \ge 1] \le \frac{1}{n}\frac{1}{(2\alpha)^{-1}-1}.$$

选取 $\alpha = 1/4$ 便可以满足题目要求.

另一种解法 无放回采样 K_n 中的边 $e_1, e_2, e_3, \dots$. 记 $G_{a:b} = (V, \{e_a, \dots, e_b\})$, 那么 $G_{a:b} \sim G(n, b-a+1)$.

用 p_m 表示 e_{m+1} 的两个端点在 $G_{1:m}$ 中联通的概率. 这个概率随着 m 的增加单调递增, 因为它也可以看作 e_1 的两个端点在 $G_{2:m+1}$ 中联通的概率.

用 c_m 表示 $G_{1:m}$ 中环的数目的期望.

$$\begin{aligned}c_m &= \sum_{k=3}^{n} \frac{n^{\underline{k}}}{2k} \Pr[v_1,\dots,v_k \text{ 按此顺序构成一个环}] \\ &\le \sum_{k=3}^{n} \frac{n^{\underline{k}}}{2k}\left(\frac{m}{\binom{n}{2}}\right)^{k} \\ &\le \sum_{k=3}^{n} \frac{1}{6}\left(\frac{2m}{n}\right)^{k}.\end{aligned}$$

注意到, 如果 e_{m+1} 的两个端点在 $G_{1:m}$ 中联通, 那么 $G_{1:m+1}$ 至少比 $G_{1:m}$ 多含有一个环. 因此

$$c_{m+1} \ge c_m + p_m, \qquad c_{2m} \ge c_m + \sum_{i=m}^{2m-1} p_i \ge m p_m, \qquad p_m \le \frac{c_{2m}}{m}.$$

可以选择 $\alpha = 1/4$

$$p_{\alpha n} \le \frac{c_{2\alpha n}}{\alpha n} \le \frac{\frac{1}{6}(2\alpha)^3}{\alpha n(1-2\alpha)} = \frac{1}{6n}.$$

所求概率为,

$$\begin{aligned}\Pr[u,v \text{ 在 } G_{\alpha n} \text{ 中联通}] &\le \Pr\Big[(u,v) \in G_{\alpha n}\Big] + \Pr\Big[u,v \text{ 在 } G_{\alpha n} \text{ 中联通} \;\Big|\; (u,v) \notin G_{\alpha n}\Big] \\ &= \frac{\alpha n}{\binom{n}{2}} + p_{\alpha n} \le \frac{1}{n}.\end{aligned}$$