马尔可夫链

参考答案

1. (6 分) 考虑 $\mathbb{Z}$ 上的随机游走. 马尔可夫核是

$$P(x+1|x) = p, \quad P(x-1|x) = 1-p$$

其中 $p \in (0,1)$ 是参数. 请计算这个马尔可夫链返回初始点的概率.

$$\Pr[\exists i > 0 \text{ such that } X_i = X_0].$$

解 不失一般性，可以假定 $X_0 = 0$，然后考虑返回 0 点的概率

$$\Pr_{X_0=0}[\exists i > 0 \text{ such that } X_i = 0].$$

定义 $v(j)$ 为如果马尔可夫链从状态 j 出发，经过 0 点的概率

$$v(j) := \Pr_{X_0=j}[\exists i \geq 0 \text{ such that } X_i = 0].$$

这里在 Pr 的下标规定 X_0 的（退化）分布. 不失一般性，计算我们关心的概率时可以假定 $X_0 = 0$，

$$\begin{aligned}\Pr[\exists i > 0, X_i = X_0] = \Pr_{X_0=0}[\exists i > 0, X_i = 0] &= p \Pr_{X_0=0}[\exists i > 0, X_i = 0 | X_1 = 1] \\ &+ (1-p) \Pr_{X_0=0}[\exists i > 0, X_i = 0 | X_1 = -1] = p\, v(1) + (1-p) v(-1).\end{aligned}$$

因此我们关心 v 的值. 显然 $v(0) = 1$. 对 $j \neq 0$，

$$\begin{aligned}v(j) = \Pr_{X_0=j}[\exists i \geq 0, X_i = 0] &= p \Pr_{X_0=j}[\exists i \geq 0, X_i = 0 | X_1 = j+1] \\ &+ (1-p) \Pr_{X_0=j}[\exists i \geq 0, X_i = 0 | X_1 = j-1] = p\, v(j+1) + (1-p) v(j-1).\end{aligned}$$

对于 $j \geq 0$，数列 $v(0), v(1), v(2), \ldots$ 有简单的线性递推公式，它们一定满足

$$v(j) = \begin{cases} \alpha j + \beta, & \text{if } p = \frac{1}{2} \\ \alpha (\frac{1-p}{p})^j + \beta, & \text{if } p \neq \frac{1}{2} \text{ 其中 } \frac{1-p}{p} \text{ 是特征方程 } 1 = px + (1-p)/x \text{ 除了 } 1 \text{ 之外的另一个根} \end{cases}$$

其中 α, β 是待定常数. 这时分三种情况考虑，

- 如果 $p = \frac{1}{2}$：因为 $v(0) = 1$ 且 $\forall j, v(j) \in [0,1]$，只能是 $v(j) = 1$.
- 如果 $p < \frac{1}{2}$：这时 $\frac{1-p}{p} > 1$. 因为 $v(0) = 1$ 且 $\forall j, v(j) \in [0,1]$，只能是 $v(j) = 1$.

- 如果 $p > \frac{1}{2}$: 这时 $\frac{1-p}{p} < 1$. 根据 $v(0) = 1$ 且 $\forall j, v(j) \in [0,1]$ 只能推出 $v(j) = \alpha(\frac{1-p}{p})^j + (1-\alpha)$ 其中 $\alpha \in [0,1]$. 因此我们要额外证明 $\lim_{j\to+\infty} v(j) = 0$，这样可以说明 $v(j) = (\frac{1-p}{p})^j$.

$$v(j) = \Pr_{X_0=j}[\exists n \geq 0, X_n = 0] \leq \sum_{n\geq 0} \Pr_{X_0=j}[X_n = 0] \leq \sum_{n\geq j} \Pr_{X_0=j}[X_n \leq j]$$
$$\leq \sum_{n\geq j} \exp\Big(-n \cdot D\big(\frac{1}{2} \,\|\, p\big)\Big) = e^{-\Theta(j)}.$$

其中一步使用了 Chernoff bound。

综合三种情况，$p\,v(1) = p\min(1, \frac{1-p}{p}) = \min(p, 1-p)$. 对称地，$(1-p)\,v(-1) = \min(p, 1-p)$. 所以

$$\Pr[\exists i > 0, X_i = X_0] = p\,v(1) + (1-p)v(-1) = 2\min(p, 1-p).$$

另一种解法：由第 2 题中的引理 1，我们只需要计算 $\mathbb{E}[N] = \sum_{n\geq 1} \Pr[X_{2n} = X_0]$（$N$ 和 E 的定义也请参照那里）. $\Pr[X_{2n} = X_0] = \binom{2n}{n}(p(1-p))^n$. 由生成函数 $\sum_{n\geq 0}\binom{2n}{n}x^n = \frac{1}{\sqrt{1-4x}}$ 代入 $x = p(1-p)$，可以得知 $\mathbb{E}[N] = \frac{1}{\sqrt{1-4p(1-p)}} - 1 = \frac{1}{|1-2p|} - 1$，从而 $\Pr[E] = 1 - \frac{1}{1+\mathbb{E}[N]} = 1 - |1-2p|$.

2. (6 分) 考虑 $\mathbb{Z}^d$ 上的随机游走. 马尔可夫核是

$$P(y_1, \ldots, y_d | x_1, \ldots, x_d) = \begin{cases} 1/3^d, & \text{if } \forall i, |y_i - x_i| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

这个马尔可夫核在各个维度上独立，便于分析. 证明

$$\Pr[\exists i > 0 \text{ such that } X_i = X_0] = \begin{cases} 1, & \text{if } d = 2 \\ 1 - \Omega(1), & \text{if } d > 2 \end{cases}$$

提示：考虑

$$\mathbb{E}[\text{number of } i > 0 \text{ such that } X_i = X_0].$$

解 定义随机变量 N 和事件 E 为

$$N := \text{number of } i > 0 \text{ such that } X_i = X_0,$$
$$E := \{\exists i > 0 \text{ such that } X_i = X_0\}.$$

引理 1. $\Pr[E] = 1$ 当且仅当 $\mathbb{E}[N] < \infty$. 当 $\Pr[E] < 1$ 时，有 $\mathbb{E}[N] = \frac{\Pr[E]}{1-\Pr[E]}$，或言 $\Pr[E] = \frac{\mathbb{E}[N]}{1+\mathbb{E}[N]}$.

证明. 由于 $N \geq 0$, $\mathbb{E}[N]$ 一定存在（可能为 ∞）. 如果 E 未发生，那么 $N = 0$，因此 $\mathbb{E}[N] = \mathbb{E}[N|E]\Pr[E]$. 现在 condition on E 发生，我们记 i_0 是最小的 $i > 0$ such that $X_i = X_0$，此时 $N = 1 + (\text{number of } i > i_0 \text{ such that } X_i = X_0)$. 由 Markov 链的无记忆性，$\mathbb{E}[\text{number of } i > i_0 \text{ such that } X_i = X_0 | E] = \mathbb{E}[N]$. 由于 $N \geq 0$ 以及期望线性性，我们有 $\mathbb{E}[N] = \Pr[E](1 + \mathbb{E}[N])$. 如果 $\Pr[E] = 1$, $\mathbb{E}[N]$ 只能为 ∞，否则 $\mathbb{E}[N] = \frac{\Pr[E]}{1-\Pr[E]}$ 是有限数. □

由期望线性性，我们有 $\mathbb{E}[N] = \sum_{n\geq 1} \Pr[X_n = X_0]$. 设 p_n 表示 $d=1$ 时的随机游走满足 $X_n = X_0$ 的概率，由于每一维独立，有 $\Pr[X_n = X_0] = p_n^d$. 在剩下的答案中，我们要证明 $p_n = \Theta(n^{-1/2})$. 这说明，当 $d \leq 2$ 时，$\mathbb{E}[N] = \infty$；当 $d \geq 3$ 时，$\mathbb{E}[N] < \infty$. 结合引理 1，便得到题目要求的结论。

估计 p_n 有多种办法，下面我们展示一种完全初等的办法. 不失一般性，可以假定 $X_0 = 0$。注意到给定的马尔可夫核可以直观地理解为如下两步过程：首先抛一个概率为 2/3 的硬币，根据硬币结果决定是否留在当前状态。如果不留在当前状态，那么以 50-50 的概率 $+1$ 或 -1. 具体来说，定义随机变量 $Z_i = \mathbb{1}[X_i \neq X_{i-1}]$. 那么 $Z_i \sim \mathrm{Bern}(2/3)$。condition on $Z_1 + \cdots + Z_n = m$，X_n 等于 m 个独立随机的 ± 1 的和，因此 $\Pr[X_n = 0 | Z_1 + \cdots + Z_n = m] = \binom{m}{m/2}/2^m$（如果 m 是偶数）。

$$\begin{aligned}\Pr[X_n = 0] &= \sum_m \Pr[X_n = 0 | Z_1 + \cdots + Z_n = m] \Pr[Z_1 + \cdots + Z_n = m] \\ &= \sum_{\text{even } m} \frac{\binom{m}{m/2}}{2^m} \Pr[Z_1 + \cdots + Z_n = m] = \sum_{\text{even } m > n/2} \frac{\binom{m}{m/2}}{2^m} \Pr[Z_1 + \cdots + Z_n = m] + 2^{-\Omega(n)}. \quad (*)\end{aligned}$$

最后一步利用 Chernoff bound，$m \leq n/2$ 的可能性可以忽略。

引理 2. 对 $k \geq 1$，我们有

$$\binom{2k}{k} \Big/ 2^{2k} = \Theta\left(\frac{1}{\sqrt{k}}\right).$$

证明. 记 $a_k = \binom{2k}{k} 2^{-2k}$，展开得 $a_k = (1 - \frac{1}{2k}) a_{k-1}$.

$$\ln a_k = \sum_{i=1}^{k} \log\Big(1 - \frac{1}{2k}\Big) \leq -\sum_{i=1}^{k} \frac{1}{2k} \leq -\frac{1}{2}\log k + \text{常数}.$$

因而 $a_k = O(k^{-1/2})$. 下界类似，只需使用 $\ln(1-x) \geq -x - x^2$ 对 $x \in [0, 1/2]$ 成立.

下界也可以用二阶矩估计。注意到 $a_k = \mathrm{Binom}(2k, \frac{1}{2})(k)$. 二项分布 $\mathrm{Binom}(2k, \frac{1}{2})$ 的期望是 k，方差是 $k/2$. 根据 Chebyshev 不等式，以至少 1/2 的概率落在 $(k - \sqrt{k}, k + \sqrt{k})$ 内。我们又知道二项分布中，期望的概率最高。所以 $a_k = \mathrm{Binom}(2k, \frac{1}{2})(k) \geq \frac{1/2}{2\sqrt{k}} = \Omega(k^{-1/2})$. □

引理 2 实际在说，存在常数 $C > c > 0$，使得对任意 $k > 1$，

$$\binom{2k}{k} \Big/ 2^{2k} \in \left[\frac{c}{\sqrt{k}}, \frac{C}{\sqrt{k}}\right].$$

回到 (*)，我们可以得到

$$\begin{aligned}\Pr[X_n = 0] &\leq \max_{\text{even } m \in (n/2, n]} \frac{\binom{m}{m/2}}{2^m} + 2^{-\Omega(n)} \leq \frac{C}{\sqrt{n/2}} + 2^{-\Omega(n)} = O\Big(\frac{1}{\sqrt{n}}\Big), \\ \Pr[X_n = 0] &\geq \max_{\text{even } m \in [0, n]} \frac{\binom{m}{m/2}}{2^m} \cdot \Pr[Z_1 + \cdots + Z_n \text{ is even}] \geq \frac{c}{\sqrt{n}} \cdot \frac{1}{3} = \Omega\Big(\frac{1}{\sqrt{n}}\Big).\end{aligned}$$

3. (6 分) 有 n 种不同卡片. 可以从一个抽卡机中，每次独立地获得一张随机卡片.

(1) 期望需要抽多少次卡，才能收集到每种卡片至少一张.

(2) 需要抽多少次卡，才能以至少 $1-1/n$ 的概率收集到每种卡片至少一张. （给出一个尽量紧的上界即可. 可以有常数倍的放松.）

解

(1) 用随机变量 $X_1, X_2, \ldots$ 依次表示我们抽到的卡片的种类，$\tau_k := \min\{m : |\{X_1, \ldots, X_m\}| = k\}$ 表示第一次收集到 k 种不同卡片的时刻. 题目只需求解 $\mathbb{E}[\tau_n]$. 那么注意到

$$\tau_n = \sum_{k=1}^{n} (\tau_k - \tau_{k-1}).$$

（这里额外定义 $\tau_0 = 0$.）并且 $\tau_k - \tau_{k-1}$ 满足几何分布，其期望为 $(1 - \frac{k-1}{n})^{-1}$，从而我们知道

$$\mathbb{E}[\tau_n] = \sum_{k=1}^{n} \Big(1 - \frac{k-1}{n}\Big)^{-1} = n \sum_{m=1}^{n} \frac{1}{m} \sim n \ln n.$$

(2) 假设我们一共抽了 t 张卡片，对于 $\forall i \in [n]$，那么有

$$\Pr[\text{没有抽到第 i 种卡片}] = \Big(\frac{n-1}{n}\Big)^t.$$

由 union bound 可知，当 $t = \alpha n \ln n$ 时，我们有

$$\Pr[\text{至少有一种卡片没有抽到}] \le \sum_{i=1}^{n} \Pr[\text{没有抽到第 i 种卡片}] = n\Big(\frac{n-1}{n}\Big)^{\alpha n \ln n} \le n^{1-\alpha}.$$

也即是说，当抽卡次数不少于 $2n \ln n$ 时，至少 $1 - 1/n$ 的概率收集到每种卡片至少一张.

4. (10 分) 简单图 G 中有 n 个点, 最大度数记为 Δ. 用 $C > 5\Delta$ 种颜色对 G 随机点染色, 要求任意一对相邻点的染色不同. 为了均匀采样一个随机染色, 我们使用 MCMC 方法. 马尔可夫核是:

- 假设当前染色为 $f : V \to C$.
- 随机选取一个点 $v \in C$, 随机选取一个颜色 $c \in C$. (TODO 明年改成随机选取一个邻居没有的颜色)
- 如果 v 的邻居的颜色都不是 c, 就将 v 的染色修改为 c; 否则保持染色不变.

请估算混合时间 $\tau(\varepsilon)$, 给出一个尽量好的上届.

$$\tau(\varepsilon) = \text{smallest } t \text{ s.t. } d(t) \le \varepsilon$$
$$d(t) = \max_x \Delta_{\mathrm{TV}}(P^t(x, \cdot), \pi)$$

注. 如果想用 *coupling* 分析 $C > 2\Delta$ 的情形, 建议用 $S_t \subseteq V$ 表示 t 时刻 *coupling* 中两个染色一致的点集. 考虑被 S_t 切的边 (一个端点在 S_t 中, 另一个端点在 S_t 外) 有怎样的影响.

解 使用 coupling 分析, 定义马尔可夫链 $\{(X_t, Y_t)\}_{t\geq 0}$. 马尔可夫核是:

- 当前染色为 X_{t-1}, Y_{t-1}.
- 随机选取一个点 $v \in V$. 在 v 以外的部分, 令 X_t 与 X_{t-1} 相同, 令 Y_t 与 Y_{t-1} 相同.

  用 $\mathcal{X} = \{X_{t-1}(u) \mid \{u, v\} \in E\}$ 和 $\mathcal{Y} = \{Y_{t-1}(u) \mid \{u, v\} \in E\}$ 分别表示 v 的邻居的颜色集合.
- 如果 $X_{t-1}(v) \neq Y_{t-1}(v)$, 随机选取一个颜色 $c \in C$. 如果 $c \notin \mathcal{X}$, 就令 $X_t(v) = c$; 否则保持染色不变. 如果 $c \notin \mathcal{Y}$, 就令 $Y_t(v) = c$; 否则保持染色不变.

  这样的话, $X_t(v) = Y_t(v)$ 的概率至少是 $\dfrac{C - |\mathcal{X} \cup \mathcal{Y}|}{C}$.
- 如果 $X_{t-1}(v) = Y_{t-1}(v)$. 从一个特定的分布采样 (c, c'). 边缘分布是均匀分布. 如果 $c \notin \mathcal{X}$, 就令 $X_t(v) = c$; 否则保持染色不变. 如果 $c' \notin \mathcal{X}$, 就令 $Y_t(v) = c'$; 否则保持染色不变.

  为了让 $X_t(v) = Y_t(v)$ 的概率尽量大, 需要适当地设置 (c, c') 的分布. 不难做到,

  $$\Pr[c = c' \wedge c \notin \mathcal{X} \cup \mathcal{Y}] = \frac{C - |\mathcal{X} \cup \mathcal{Y}|}{C}, \quad \Pr[c \in \mathcal{X} \wedge c' \in \mathcal{Y}] = \frac{\min(|\mathcal{X}|, |\mathcal{Y}|)}{C}.$$

  这样的话, $X_t(v) = Y_t(v)$ 的概率是 $\dfrac{C - |\mathcal{X} \cup \mathcal{Y}| + \min(|\mathcal{X}|, |\mathcal{Y}|)}{C} = \dfrac{C - \min(|\mathcal{X} \setminus \mathcal{Y}|, |\mathcal{Y} \setminus \mathcal{X}|)}{C}$.

根据提示, 定义 $S_t = \{v | X_t(v) = Y_t(v)\}$. 定义 $N_t = |S_t|$.

考虑 X_{t-1}, Y_{t-1} 到 X_t, Y_t 的马尔可夫核的采样过程.

- 如果 $X_{t-1}(v) \neq Y_{t-1}(v)$ (即 $v \notin S_{t-1}$), 只要采样得到的颜色 c 不在 $\mathcal{X} \cup \mathcal{Y}$ 中, 那么 v 在两边的颜色都会被更新到 c. 用 $E(v, S)$ 表示 v 和 S 之间的边数. 那么

  $$|\mathcal{X} \cup \mathcal{Y}| \leq E(v, S_{t-1}) + 2E(v, V \setminus S_{t-1}) \leq 2\Delta - E(v, S_{t-1}).$$

  于是 $X_t(v) = Y_t(v)$ 的概率至少是 $\frac{C - 2\Delta + E(v, S_{t-1})}{C}$.
- 如果 $X_{t-1}(v) = Y_{t-1}(v)$ (即 $v \in S_{t-1}$), $X_t(v) \neq Y_t(v)$ 的概率至多是

  $$\frac{\min(|\mathcal{X} \setminus \mathcal{Y}|, |\mathcal{Y} \setminus \mathcal{X}|)}{C} \leq \frac{E(v, V \setminus S_{t-1})}{C}.$$

观察上面的概率. 当 $v \notin S_{t-1}$ 时, v 与 S_{t-1} 的连边 "有益"; 当 $v \in S_{t-1}$ 时, v 与 $V \setminus S_{t-1}$ 的连边 "有害". 因为

$$\sum_{v \notin S_{t-1}} E(v, S_{t-1}) = E(V \setminus S_{t-1}, S_{t-1}) = \sum_{v \in S_{t-1}} E(v, V \setminus S_{t-1}),$$

这两种作用恰好可以相互抵消.

具体来说, 给定 X_{t-1}, Y_{t-1} 时,

$$\begin{aligned}
\mathbb{E}\Big[N_t - N_{t-1} \;\Big|\; X_{t-1}, Y_{t-1}\Big] &\geq \sum_{v \notin S_{t-1}} \frac{1}{n} \frac{C - 2\Delta + E(v, S_{t-1})}{C} - \sum_{v \in S_t} \frac{1}{n} \frac{E(v, V \setminus S_{t-1})}{C} \\
&= \sum_{v \notin S_{t-1}} \frac{1}{n} \frac{C - 2\Delta}{C} \\
&= \frac{n - N_{t-1}}{n} \frac{C - 2\Delta}{C}.
\end{aligned}$$

对两边同时求期望, 可以得到

$$\mathbb{E}[N_t] - \mathbb{E}[N_{t-1}] = \frac{n - \mathbb{E}[N_{t-1}]}{n} \frac{C - 2\Delta}{C}.$$

于是解得

$$\mathbb{E}[N_t] = n - \Big(1 - \frac{C - 2\Delta}{nC}\Big)^t (n - \mathbb{E}[N_0]).$$

当 $t \geq \frac{\log(\varepsilon/n)}{\log(1-\frac{C-2\Delta}{nC})} \geq \frac{nC}{C-2\Delta}\log(n/\varepsilon)$ 时, $\mathbb{E}[N_t] \geq n - \varepsilon$, 根据 Markov bound,

$$\Pr[X_t \neq Y_t] = \Pr[N_t \neq n] = \Pr[N_t \leq n-1] \leq \varepsilon.$$

上述分析不依赖于 X_0, Y_0 的分布. 只要令 X_0 是任意染色, Y_0 服从稳态分布 (均匀分布), 就得到

$$\tau(\varepsilon) \leq \frac{nC}{C - 2\Delta}\log(n/\varepsilon).$$

5. (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：

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

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 $v_{i+1},\dots,v_n$ 皆无连边. 如果 B 也同时成立，那么刚采样完 $v_1,\dots,v_{i-1}$ 时，总共有不超过 $n-k$ 条边，此时 $V\setminus\{v_1,\dots,v_{i-1}\}$ 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 $v_1,\dots,v_{i-1}$ 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

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

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

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 $\Pr[\neg B]\le e^{-n^{\Omega(1)}}$.

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

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

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. $m\ge3n\ln n$，从而结论成立.