期末试题

试卷共 11 页, 共 13 题, 满分 30 分.

判断/选择题: 无需写出证明.

1. (1 分) 对任意 $n \in \mathbb{N}$, $\binom{2n}{n}$ 都是偶数.

   解 是

   $\binom{2n}{n} = \binom{2n-1}{n} + \binom{2n-1}{n-1} = 2\binom{2n-1}{n}$.

   可以将 $\binom{[2n]}{n}$ 中元素两两组对. S 与 $[2n] \setminus S$ 组对.

2. (1 分) 两个 (不独立的) 随机变量 X, Y 满足 $0 \le X \le Y$ 恒成立. 那么 $\mathrm{Var}[X] \le \mathrm{Var}[Y]$.

   解 否

   例如 Y 是常数且是 X 的上界.

3. (1 分) 如果一个平面图是简单图, 那么它的对偶 (dual graph) 也是简单图.

   解 否

   最小反例是 K_2.

4. (1 分) 对任意 $\delta > 0$, 存在 $\varepsilon > 0$, 使得对于任何有限空间中的分布 P, Q,

   $$\Delta_{\mathrm{TV}}(P, Q) \le \varepsilon \implies D_{\mathrm{KL}}(P \| Q) \le \delta.$$

   解 否

   $D_{\mathrm{KL}}(\mathrm{Bern}(\varepsilon) \| \mathrm{Bern}(0)) = +\infty$.

填空题: 无需写出证明.

5. (2 分) $S_n = \{\mathbf{v} \in \{0,1,2\}^n \mid \sum v_i = n\}$. 估算 $|S_n|$, 要求误差不超过 $n^{O(1)}$.

   解 3^n

   3^n 显然是 $|S_n|$ 的上界.

   从 $\{0,1,2\}^n$ 中随机采样, 和的分布是单峰. 说明 $|S_n| \ge 3^n/(2n+1)$.

   下界也可以用 Sanov bound 说明.

6. (2 分) 有 n 张互异的扑克牌, 摞成一堆. 用以下方式洗牌: 每次随机抽取一张扑克牌, 再放回牌堆顶部. 请估算洗牌时间.

   具体来说, 这个过程可以视为一个马尔可夫链, 其稳态分布是均匀分布 (洗匀的牌堆). 请估计混合时间 $t_{\mathrm{mix}}(1/100)$. 允许有常数误差.

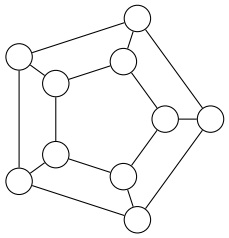
   解 $\Theta(n \log n)$

   洗牌过程中, 可以把牌堆分为两部分: 底部尚未被抽取过的牌, 和顶部被放回的牌. 顶部是洗好的. 当底部还剩不超过 1 张时, 整个牌堆都是均匀的. 当底部还剩 2 张时, 离均匀分布的 TVD 是 1/2. 所以问题转化为何时底部不超过 1 张.

   这等价于 coupon collection 问题.

   记首次将底部抽到还剩 1 张的时间为 T. 那么 T 是 $\mathrm{Geom}(\frac{n}{n}), \mathrm{Geom}(\frac{n-1}{n}), \ldots, \mathrm{Geom}(\frac{2}{n})$ 的独立和. 期望约为 $n \log n - n$, 方差为 $\Theta(n^2)$. 上界用期望和 Markov bound, 时间 $100n \log n$ 后以至少 99% 的概率洗匀. 下界用二阶矩方法, 在时间 $n \log n - n/\varepsilon$ 时底部剩至少两张牌的概率不低于 $1 - O(\varepsilon^2)$.

7. (2 分) 考虑一个 10 个珠子构成的双层项链, 用两种颜色对珠子染色, 有多少种不同的染色方案.



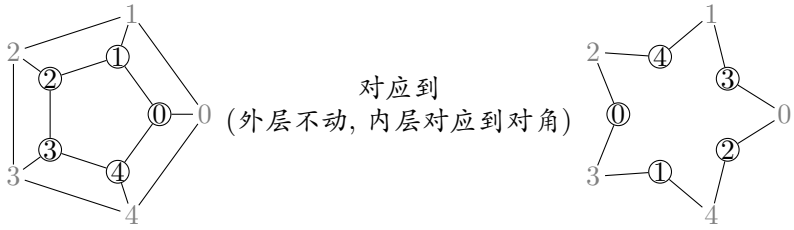
   只要存在图同构都视为相同染色. 包括旋转、内外层交换、镜像.

   解 染色是 $\mathbb{Z}_2 \times \mathbb{Z}_5 \to \{0,1\}$. 考虑 $\mathbb{Z}_2 \times D_5$ 对染色的群作用.

   根据 Pólya 计数, 不同染色数为 $\dfrac{2^{10} + 2^5 + 4 \cdot 2^2 + 4 \cdot 2 + 5 \cdot 2^6 + 5 \cdot 2^5}{20} = 78.$

群元素	描述	不动点个数
$(0, e)$		2^{10}
$(1, e)$	交换内外	2^5
$(0, r^i)$ for $i \in \mathbb{Z}_5^*$	旋转	2^2
$(1, r^i)$ for $i \in \mathbb{Z}_5^*$	旋转 + 交换内外	2^1
$(0, sr^i)$ for $i \in \mathbb{Z}_5$	镜像	2^6
$(1, sr^i)$ for $i \in \mathbb{Z}_5$	镜像 + 交换内外	2^5

   另一种解法: 注意到问题等价于对 10 个珠子构成的单层项链染色:



   根据 Pólya 计数, 不同染色数为 $\dfrac{2^{10} + 2^5 + 4 \cdot 2^2 + 4 \cdot 2 + 5 \cdot 2^6 + 5 \cdot 2^5}{20} = 78.$

群元素	描述	不动点个数
e		2^{10}
r^5		2^5
r^i for $\gcd(i,10)=1$		2^2
r^i for $\gcd(i,10)=2$		2^1
sr^i for even i	过两点的对称轴镜像	2^6
sr^i for odd i	过两边的对称轴镜像	2^5

解答题: 请选择 4 道作答.

8. (5 分) 用 C 种颜色对满二叉树中的点进行染色. 两个染色方案等价, 当且仅当其中一种方案可以经过一系列左右子树的交换转化为另一种方案.

(1) 高度为 2 的满二叉树 (共有 3 个点) 有多少种不同的染色方案?

(2) 高度为 3 的满二叉树 (共有 7 个点) 有多少种不同的染色方案?

(3) 高度为 4 的满二叉树 (共有 15 个点) 有多少种不同的染色方案?

不强制要求把结果完全展开, 但请表述为清晰便于验证的形式.

解 用 C_d 表示高度为 d 的满二叉树的染色方案数.

注意到, 可以把 C_d 种染色方案都视为一种颜色. 这样用 C 种颜色染色 $d+1$ 层满二叉树, 就等价于对 2 层满二叉树染色, 其中根节点可选 C 种颜色, 叶子节点可选 C_d 种颜色. 因此

$$C_{d+1} = C(\underbrace{C_d}_{\text{两叶子同色}} + \underbrace{\binom{C_d}{2}}_{\text{两叶子异色}}) = C\frac{C_d^2+C_d}{2}.$$

那么从 $C_0 = 1$ 开始归纳.

$$C_1 = C, \quad C_2 = \frac{C^3+C^2}{2}, \quad C_3 = \frac{C^7+2\cdot C^6+C^5+2\cdot C^4+2\cdot C^3}{8},$$

$$\begin{aligned} C_4 &= \frac{C(\frac{C^7+2\cdot C^6+C^5+2*C^4+2*C^3}{8})(\frac{C^7+2\cdot C^6+C^5+2\cdot C^4+2\cdot C^3}{8}+1)}{2} \\ &= \frac{C^4(C+1)(C^3+C^2+2)(C^7+2C^6+C^5+2C^4+2C^3+8)}{128} \\ &= \frac{C^4(C^{11}+4C^{10}+6C^9+8C^8+13C^7+12C^6+8C^5+16C^4+20C^3+8C^2+16C+16)}{128}. \end{aligned}$$

另一种 (暴力) 解法:

(1) $\frac{C^3+C^2}{2}$.

(2) $\frac{C^7+2\cdot C^6+C^5+2\cdot C^4+2\cdot C^3}{8}$

一共有 8 种变换方式, 对应的不动点个数如下表

群元素	个数	不动点个数
不动	1	C^7
第一层左右子树无交换, 第二层 1 个节点左右子树有交换	2	C^6
第一层左右子树无交换, 第二层 2 个节点左右子树有交换	1	C^5
第一层左右子树有交换, 第二层左右子树同时交换/不交换	2	C^4
第一层左右子树有交换, 第二层左右子树一交换一不交换	2	C^3

(3) 一共有 2^7 种变换方式.

- 如果第一层左右子树无交换, 这些变换构成一个变换子群, 子群的大小为 2^6. 由前面一问, 在这些变换下等价的染色方案数为

$$C \cdot \left(\frac{C^7 + 2 \cdot C^6 + C^5 + 2 \cdot C^4 + 2 \cdot C^3}{8}\right)^2$$

  这些变换对应的不动点总数为 $C \cdot \left(\frac{C^7+2\cdot C^6+C^5+2\cdot C^4+2\cdot C^3}{8}\right)^2 \cdot 2^6$.

- 如果第一层的左右子树有交换, 对应另外 2^6 个变换.

群元素	个数	不动点个数
第二层左右子树同时交换/不交换, 第三层两对子树中 2 对同时交换/不交换	8	C^8
第二层左右子树同时交换/不交换, 第三层两对子树中 1 对同时交换/不交换	16	C^7
第二层左右子树同时交换/不交换, 第三层两对子树中 0 对同时交换/不交换	8	C^6
第二层左右子树一交换一不交换, 第三层 4 个节点中 0/2/4 个交换	16	C^5
第二层左右子树一交换一不交换, 第三层 4 个节点中 1/3 个交换	16	C^4

综上述, 不同的染色方案数总数是

$$\frac{C \cdot \left(\frac{C^7+2\cdot C^6+C^5+2\cdot C^4+2\cdot C^3}{8}\right)^2 \cdot 2^6 + 8C^8 + 16C^7 + 8C^6 + 16C^5 + 16C^4}{128}.$$

评分标准: 第一问 1 分; 第二三问各 2 分

9. (5 分) 每个置换 $f \in \mathrm{Sym}(n)$ 都可以拆成若干不交的轮换 (cycle). 不动点就是一个长度为 1 的轮换.

(1) 求 $\mathrm{Sym}(2n)$ 中只包含偶数长度轮换的置换个数 a_{2n}, 并证明.

(2) 求 $\mathrm{Sym}(2n)$ 中只包含奇数长度轮换的置换个数 b_{2n}, 并证明.

举例来说, (1)(2)(34) $\in \mathrm{Sym}(4)$ 包含了奇长度轮换 (1), 所以不在第一问的计数范围; 同时包含了偶长度轮换 (34), 所以也不在第二问的计数范围.

解 用 $c_0 = 0$, $c_n = (n-1)!$ 表示 $\mathrm{Sym}(n)$ 中的轮换置换的个数. 其对应 EGF $\tilde{C}(x) = \ln(\frac{1}{1-x})$.

(1) 用 c_n^{even} 表示 $\mathrm{Sym}(n)$ 中的轮换置换的个数并限制 n 为偶数, 并定义对应的 EGF 为 $\tilde{C}^{\text{even}}$. 那

么

$$c_n^{\text{even}} = \begin{cases} c_n, & \text{if } n \text{ even} \\ 0, & \text{otherwise} \end{cases} \qquad \tilde{C}^{\text{even}}(x) = \frac{1}{2}\Big(\tilde{C}(x) + \tilde{C}(-x)\Big) = \frac{1}{2}\ln(\frac{1}{1-x^2}).$$

用 $\tilde{A}$ 表示 $\{a_n\}$ 的 EGF. 那么 $\tilde{A}(x) = \exp(\tilde{C}^{\text{even}}(x)) = \sqrt{\frac{1}{1-x^2}}$. 使用广义二项式定理

$$\begin{aligned} \tilde{A}(x) &= (1-x^2)^{-1/2} \\ &= \sum_{n=0}^{\infty} \frac{(-\frac{1}{2})^{\underline{n}}}{n!}(-x^2)^n \\ &= \sum_{n=0}^{\infty} \frac{(-\frac{1}{2})(-\frac{3}{2})\dots(-\frac{2n-1}{2})}{n!}(-1)^n x^{2n} \\ &= \sum_{n=0}^{\infty} \frac{1\cdot 3\dots(2n-1)}{n!2^n} x^{2n} \\ &= \sum_{n=0}^{\infty} \frac{(2n)!}{n!n!4^n} x^{2n}. \end{aligned}$$

所以

$$a_{2n} = \frac{(2n)!(2n)!}{n!n!4^n} = \left(\frac{(2n)!}{n!2^n}\right)^2 = ((2n-1)!!)^2.$$

另一种解法: 考虑以下方式选取一个符合要求的置换: 依次进行 $2n$ 轮, 每轮确定 f 在一个位置的输出.

- 初始化 $s = v = 0$ 为起点.
- 第一轮: 设置 $f(v) = a_1$. 因为 f 只含偶长度轮换, a_1 只能在 $[2n] \setminus \{s\}$ 中选择. 更新 $v \leftarrow a_1$.
- 第二轮: 设置 $f(v) = a_2$. 注意到 a_2 的选择空间是 $[2n] \setminus \{a_1\}$. 更新 $v \leftarrow a_2$.
  一种特殊情况是 $a_2 = s$. 这时已经构成了一个长为 2 的轮换. 这时更新 s 为 $\setminus\{a_1, a_2\}$ 中的任意值 (不妨选其中的最小值), 并令 $v \leftarrow s$.
- 一般地, 在奇数 i 轮, 设置 $f(v) = a_i$, 其中 a_i 的选择空间是 $[2n] \setminus \{s, a_1, \dots, a_{i-1}\}$, 有 $2n - i$ 种选择. 更新 $v \leftarrow a_i$.
- 一般地, 在偶数 i 轮, 设置 $f(v) = a_i$, 其中 a_i 的选择空间是 $[2n] \setminus \{a_1, \dots, a_{i-1}\}$, 有 $2n - i + 1$ 种选择. 更新 $v \leftarrow a_i$.
  如果 $a_i = s$, 令 $s \leftarrow \min([2n] \setminus \{a_1, \dots, a_i\})$ 并更新 $v \leftarrow s$.

因此符合要求的置换的个数为

$$(2n-1)(2n-1)(2n-3)(2n-3)\dots 3\cdot 3\cdot 1\cdot 1 = ((2n-1)!!)^2.$$

(2) 类似地, 用 c_n^{odd} 表示 $\text{Sym}(n)$ 中的轮换置换的个数并限制 n 为偶数, 并定义对应的 EGF 为 $\tilde{C}^{\text{odd}}$. 那么

$$c_n^{\text{odd}} = \begin{cases} c_n, & \text{if } n \text{ odd} \\ 0, & \text{otherwise} \end{cases} \qquad \tilde{C}^{\text{odd}}(x) = \frac{1}{2}\Big(\tilde{C}(x) - \tilde{C}(-x)\Big) = \frac{1}{2}\ln(\frac{1+x}{1-x}).$$

用 b_n 表示 $\mathrm{Sym}(n)$ 中只包含奇数长度轮换的置换个数, 其对应 EGF 为 $\tilde{B}(x) = \exp(\tilde{C}^{\mathrm{odd}}(x)) = \sqrt{\frac{1+x}{1-x}}$. 用 b_n^{even} 表示限制 n 为偶数, 对应 EGF 为 $\tilde{B}^{\mathrm{even}}$. 那么

$$b_n^{\mathrm{even}} = \begin{cases} b_n, & \text{if } n \text{ even} \\ 0, & \text{otherwise} \end{cases} \qquad \tilde{B}^{\mathrm{even}}(x) = \frac{1}{2}\Big(\tilde{B}(x) + \tilde{B}(-x)\Big) = \frac{1}{2}\sqrt{\frac{1}{1-x^2}} = \tilde{A}(x).$$

所以 $b_{2n} = a_{2n} = (\frac{(2n)!}{n!2^n})^2 = ((2n-1)!!)^2$.

评分标准: 做出任意一问得 3 分; 其中结果 1 分证明 2 分

10. (5 分) 有 n 对情侣, 随机将他们分为 n 组, 每组两人. 有多少对情侣被分到同组中?

    (1) 当 n 充分大时, 所有情侣都没有分入同组的概率是多少. 证明你的结果.

    (2) 用 $P_{n,k}$ 表示其中恰好有 k 对情侣分到同组的概率. 对于任意常数 k, 请计算 $\lim_{n\to\infty} P_{n,k}$ 并证明你的结果.

解

(1) 用 $[n]$ 表示这 n 对情侣. 用 $C \subseteq [n]$ 表示被分配到同组的情侣. 我们关心 $\Pr[|C| = k]$.

对于任意 $S \in \binom{[n]}{t}$ 易知,

$$\Pr[C \supseteq S] = \frac{1}{2n-1}\frac{1}{2n-1}\cdots\frac{1}{2n-2t+1} = \frac{(2n-2t-1)!!}{(2n-1)!!}$$

由容斥原理,

$$\Pr[C \neq \varnothing] = \sum_{t=1}^{n}\sum_{S\in\binom{[n]}{t}} (-1)^{t-1}\Pr[C \supseteq S] = \sum_{t=1}^{n}(-1)^{t-1}\binom{n}{t}\frac{(2n-2t-1)!!}{(2n-1)!!}$$

定义求和的第 t 项为 $W_{n,t}$. 可以估算为

$$A_{n,t} = \binom{n}{t}\frac{(2n-2t-1)!!}{(2n-1)!!} = \frac{1}{t!}\frac{n}{2n-1}\frac{n-1}{2n-1}\cdots\frac{n-t+1}{2n-2t+1} \approx \frac{1}{t!}2^{-t}.$$

形式化来说, 当 t 是常数时 (可以放松到 $t = o(n)$ 时), $\lim_{n\to\infty} A_{n,t} = \frac{1}{t!}2^{-t}$. 如果这里极限和求和能交换顺序, 那么

$$\begin{aligned}
\lim_{n\to\infty}\Pr[C \neq \varnothing] &= \lim_{n\to\infty}\sum_{t=1}^{n}\binom{n}{t}(-1)^{t-1}\frac{(2n-2t-1)!!}{(2n-1)!!} \\
&\overset{(*)}{=} \sum_{t=1}^{\infty}(-1)^{t-1}\frac{1}{t!}2^{-t} \\
&= 1 - \sum_{t=0}^{\infty}(-1)^{t}\frac{1}{t!}2^{-t} \\
&= 1 - e^{-1/2}
\end{aligned}$$

接下来, 我们说明 (*) 确实成立. 我们使用一种迂回的办法, 先证明对于任意 $m \le n$,

$$\sum_{t=1}^{m}\sum_{S\in\binom{[n]}{t}}(-1)^{t-1}\Pr[C \supseteq S] \begin{cases} \le \Pr[C \neq \varnothing], & \text{if } m \text{ even} \\ \ge \Pr[C \neq \varnothing], & \text{if } m \text{ odd} \end{cases}$$

证明如下

$$\begin{aligned}
&\sum_{t=1}^{m}\sum_{S\in\binom{[n]}{t}}(-1)^{t-1}\Pr[C\supseteq S]\\
=&\sum_{t=1}^{m}\sum_{S\in\binom{[n]}{t}}(-1)^{t-1}\sum_{T\supseteq S}\Pr[C=T]\\
=&\sum_{T\neq\varnothing}\Pr[C=T]\sum_{t=1}^{m}\sum_{S\in\binom{T}{t}}(-1)^{t-1} && \sum_{t=0}^{m}\sum_{S\in\binom{T}{t}}(-1)^{t}\\
=&\sum_{T\neq\varnothing}\Pr[C=T]\left(1-\sum_{t=0}^{m}\sum_{S\in\binom{T}{t}}(-1)^{t}\right) && =\sum_{t=0}^{m}\binom{|T|}{t}(-1)^{t}\\
=&\Pr[C\neq\varnothing]-\sum_{T\neq\varnothing}\Pr[C=T]\sum_{t=0}^{m}\sum_{S\in\binom{T}{t}}(-1)^{t} && =(-1)^{m}\binom{|T|-1}{t}
\end{aligned}$$

因此, 对于任何偶常数 m

$$\lim_{n\to\infty}\Pr[C\neq\varnothing]\geq\lim_{n\to\infty}\sum_{t=1}^{m}\binom{n}{t}(-1)^{t-1}\frac{(2n-2t-1)!!}{(2n-1)!!}=\sum_{t=1}^{m}(-1)^{t-1}\frac{1}{t!}2^{-t}.$$

同时对于任何奇常数 m, 上式的不等号方向反转. 因为右侧的交错级数收敛到 $1-e^{-1/2}$, 说明左边等于 $1-e^{-1/2}$.

(2) 总有 $\Pr[C=S]=\Pr[C\subseteq S]\Pr[C=S\mid C\subseteq S]$. 注意到, 最后的条件概率其实就是将 $n-|S|$ 对情侣随机分组后没有情侣同组的概率.

因此, 对与任意常数 k.

$$\begin{aligned}
\lim_{n\to\infty}\Pr[|C|=k]&=\lim_{n\to\infty}\sum_{S\in\binom{[n]}{k}}\Pr[C\subseteq S]\Pr[C=S\mid C\subseteq S]\\
&=\lim_{n\to\infty}\binom{n}{k}\frac{(2n-2k-1)!!}{(2n-1)!!}\Pr[C=S\mid C\subseteq S]\\
&=\lim_{n\to\infty}\binom{n}{k}\frac{(2n-2k-1)!!}{(2n-1)!!}\cdot\lim_{n\to\infty}\Pr[C=\varnothing]\\
&=\frac{1}{k!}2^{-k}e^{-1/2}.
\end{aligned}$$

另一种解法: 基于前面第二问的观察, $\lim_{n\to\infty}\Pr[|C|=k]=\frac{1}{k!}2^{-k}\lim_{n\to\infty}\Pr[C=\varnothing]$.

同时我们知道 $\mathbb{E}[|C|]=\frac{n}{2n-1}<1$. 由 Markov bound $\Pr[|C|\geq m]<\frac{1}{m}$. 那么对任意常数 m

$$(1-1/m,1]\ni\lim_{n\to\infty}\Pr[|C|\leq m]=\sum_{k=0}^{m}\frac{1}{k!}2^{-k}\lim_{n\to\infty}\Pr[C=\varnothing].$$

那么将 $m\to\infty$ 便得出

$$\lim_{n\to\infty}\Pr[C=\varnothing]=1\Big/\sum_{k=0}^{\infty}\frac{1}{k!}2^{-k}=e^{-1/2}.$$

评分标准: 第一问 3 分, 结果 1 分证明 2 分, 只完成粗估扣一分; 第二问 2 分, 结果 1 份证明 2 分

11. (5 分) 这里我们设计一种高效纠错码.

(1) 考虑压缩函数 $f:\{0,1\}^n\to\{0,1\}^m$ 和解压缩函数 $g:\{0,1\}^m\to\{0,1\}^n$. 固定一个常数概率 $p<1/2$, 求出

$$\alpha^*=\inf\left\{\alpha\middle|\begin{array}{c}\exists f:\{0,1\}^n\to\{0,1\}^{\alpha n},g:\{0,1\}^{\alpha n}\to\{0,1\}^n\\ \Pr\limits_{\mathbf{x}\sim\mathrm{Bern}(p)^n}[g(f(\mathbf{x}))=\mathbf{x}]=1-2^{\Theta(n)}\end{array}\right\}$$

不需写出证明和计算过程.

(2) 如果额外限制 f 为 $\mathbb{F}_2$ 上的线性函数, 那么 α^* 将如何变化? 证明你的结论.

(3) 符合前一问的压缩和解压缩函数可以非常高效 ($O(n\log n)$ 电路大小). 据此设计用于 BSC(p) 信道的高效编码 $E:\{0,1\}^n\to\{0,1\}^\ell, D:\{0,1\}^\ell\to\{0,1\}^n$. 要求对任意消息 $\mathbf{w}\in\{0,1\}^n$,

$$\Pr_{Y\sim(P_{Y|X})^\ell(\cdot|E(\mathbf{w}))}[D(Y)=\mathbf{w}]\geq 1-\exp(-\Theta(n)).$$

在这样的要求下, 码率 n/ℓ 可以达到多少? 请写出 E,D 的构造并证明结论.

BSC(p) 的定义: $P_{Y|X}(0|0)=P_{Y|X}(1|1)=1-p$, $P_{Y|X}(0|1)=P_{Y|X}(1|0)=p$.

解 答案中对数以 2 为基底.

(1) $h(p)$.

(2) 选取随机的线性压缩函数 f. 固定一个小常数 ε 和集合

$$S=\{\mathbf{x}\in\{0,1\}^n\mid\|\mathbf{x}\|_1<n(p+\varepsilon)\}$$

解压缩函数为

$$g(\mathbf{y})=\begin{cases}\mathbf{x}, & \text{如果存在唯一 }\mathbf{x}\in S\text{ 满足 }f(\mathbf{x})=\mathbf{y}\\ \perp, & \text{otherwise}\end{cases}$$

解压缩失败的可能性有两种:

- 一是 $\mathbf{x}\sim\mathrm{Bern}(p)^n$ 没有落入集合 S 中. 根据 Chernoff bound, 这个概率不超过 $\exp(-\varepsilon^2 n/2)$.
- 一是存在另一个 $\mathbf{x}'\in S$ 满足 $f(\mathbf{x})=f(\mathbf{x}')$. 对于任何一个 $\mathbf{x}'\neq\mathbf{x}$, 碰撞发生的概率为 2^{-m}. 根据 union bound, 碰撞的概率不超过

$$\frac{|S|-1}{2^m}\leq n\cdot\frac{2^{h(p+\varepsilon)n}}{2^{\alpha n}}$$

因此只要 $\alpha>h(p)$, 便可适当选取 ε 使得解码错误概率指数小.

(3) 根据上一问, 对任意 $\alpha>h(p)$, 有线性压缩函数 $f:\{0,1\}^\ell\to\{0,1\}^{\alpha\ell}$. 考虑 f 的零空间 (也就是 kernel) 是一个 $(1-\alpha)\ell$ 维的子空间. 令 $n=(1-\alpha)\ell$, 令 E 是 $\{0,1\}^n$ 到 $\mathrm{Ker}\,f$ 的线性双射.

$\mathrm{BSC}(p)^\ell$ 信道可以看作 $Y = \mathbf{x} + Z$, 其中 $\mathbf{x}$ 是信道输入, Y 是信道输出, $Z \sim \mathrm{Bern}(p)^\ell$. 那么当输入 $\mathbf{x} = E(\mathbf{w})$ 时,

$$g(f(Y)) = g(f(\mathbf{x} + Z)) = g(f(Z)) \overset{w.h.p.}{=} Z$$

可以大概率解出噪声 Z. 将噪声 Z 减去即完成解码. 解码算法为

$$D(\mathbf{y}) = E^{-1}(\mathbf{y} - g(f(\mathbf{y})))$$

码率为 $n/\ell = 1 - \alpha$, 可以任意逼近信道容量 $C = 1 - h(p)$.

评分标准: 第一问 1 分; 第二三问各 2 分

12. (6 分) 给定一个奇素数 p, 我们将 $x \in \mathbb{Z}_p^*$ 分为两类: 如果 $\exists y \in \mathbb{Z}_p^*$, $x = y^2$, 那么称 x 为二次剩余, 否则称 x 为二次非剩余. 不难说明 $\mathbb{Z}_p^*$ 中恰好有一半是二次剩余. 判断一个元素是否为二次剩余可以使用勒让德符号 (Legendre symbol)

$$\left(\frac{x}{p}\right) = x^{\frac{p-1}{2}} = \begin{cases} +1, & \text{if } x \text{ 是二次剩余} \\ -1, & \text{if } x \text{ 是二次非剩余} \\ 0, & \text{if } x = 0 \end{cases}$$

不难说明, $x \mapsto (\frac{x}{p})$ 是一个 $\mathbb{Z}_p^*$ 到 $\{-1, 1\}$ 的群同态.

请证明存在常数 C_0, C_1 使得, 对于任何 $n \in \mathbb{N}$, 任何素数 $p > 2^{C_0 n + C_1}$ 和任何 $(b_1, \ldots, b_n) \in \{0,1\}^n$, 存在 $x \in \mathbb{Z}_p$ 满足

$$\left(\frac{x+1}{p}\right) = (-1)^{b_1}, \left(\frac{x+2}{p}\right) = (-1)^{b_2}, \ldots, \left(\frac{x+n}{p}\right) = (-1)^{b_n}.$$

Weil bound 的大意为: $\mathbb{F}_p$ 上的"简单"曲线经过大约 p 个点. 特别地, 证明需要 Weil bound 的以下推论: 如果 $f \in \mathbb{F}_p[x]$ 满足 $\deg f = d > 0$ 且无平方 (square-free, 等价于在任何扩域中无重根), 那么

$$\left|\left\{(x, y) \in \mathbb{Z}_p^2 \;\middle|\; f(x) = y^2\right\}\right| \in (p - d\sqrt{p}, p + d\sqrt{p}).$$

证明也可以使用这个推论的以下等价表述, 等价性只需注意到 $(\frac{x}{p}) = |\{y \mid y^2 = x\}| - 1$,

$$\sum_{x \in \mathbb{Z}_p} \left(\frac{f(x)}{p}\right) \in (-d\sqrt{p}, d\sqrt{p}).$$

提示: 对任意非空子集 $\Theta \subseteq \mathbb{F}_p$, 考虑函数 $x \mapsto \prod_{i \in \Theta}(x - i)$.

解 令 X 在 $\{0, 1, 2, \ldots, p - n - 1\}$ 中均匀分布. 定义概率函数 P 为

$$P(b_1, \ldots, b_n) = \Pr\left[\left(\frac{X+1}{p}\right) = (-1)^{b_1}, \ldots, \left(\frac{X+n}{p}\right) = (-1)^{b_n}\right].$$

题目等价于证明 P 在 $\mathbb{F}_2^n$ 中任何位置概率都非零. 实际上, 我们将证明 P 足够接近均匀分布. 为此我们考虑 P 的傅立叶变换

$$\hat{P}(\mathbf{s}) = \sum_{\mathbf{b} \in \mathbb{F}_2^n} (-1)^{\langle \mathbf{b}, \mathbf{s} \rangle} P(\mathbf{b})$$

$\hat{P}$ 的输入 $\mathbf{s}$ 也可以视作一个集合. $\mathbf{s}$ 对应 $S = \{i | s_i = 1\}$. 那么傅立叶变换就是

$$\begin{aligned}
\hat{P}(S) &= \sum_{\mathbf{b}\in\mathbb{F}_2^n} (-1)^{\sum_{i\in S} b_i} P(\mathbf{b}) \\
&= \mathop{\mathbb{E}}_{(b_1,\ldots,b_n)\sim P}\left[(-1)^{\sum_{i\in S} b_i}\right] \\
&= \mathop{\mathbb{E}}_{X\leftarrow\{0,1,\ldots,p-n-1\}}\left[\prod_{i\in S}\left(\frac{X+i}{p}\right)\right] \\
&= \mathop{\mathbb{E}}_{X\leftarrow\{0,1,\ldots,p-n-1\}}\left[\left(\frac{\prod_{i\in S}(X+i)}{p}\right)\right] \\
&= \mathop{\mathbb{E}}_{X\leftarrow\{0,1,\ldots,p-n-1\}}\left[\left(\frac{f_S(X)}{p}\right)\right]
\end{aligned}$$

其中 f_S 定义为 $f_S(x) = \prod_{i\in S}(x+i)$. 只要 S 非空, f_S 就满足 Weil bound 推论的条件: 度数大于 0 且 square-free. 因此

$$\begin{aligned}
\hat{P}(S) &= \mathop{\mathbb{E}}_{X\leftarrow\{0,1,\ldots,p-n-1\}}\left[\left(\frac{f_S(X)}{p}\right)\right] \\
&= \frac{1}{p-n}\sum_{x=0}^{p-n-1}\left(\frac{f_S(X)}{p}\right) \\
&= \frac{1}{p-n}\sum_{x=0}^{p-1}\left(\frac{f_S(X)}{p}\right) \\
&\in \left(-\frac{|S|\sqrt{p}}{p-n}, \frac{|S|\sqrt{p}}{p-n}\right)
\end{aligned}$$

可以看出只要 p 足够大, $\hat{P}(S)$ 就足够小. 考虑逆傅立叶变换, 对任意 $\mathbf{b}\in\{0,1\}^n$

$$\begin{aligned}
P(\mathbf{b}) &= \frac{1}{2^n}\sum_{S}(-1)^{\sum_{i\in S} b_i}\hat{P}(S) \\
&\geq \frac{1}{2^n}\left(\hat{P}(\varnothing) - \sum_{S\neq\varnothing}|\hat{P}(S)|\right) \\
&\geq \frac{1}{2^n}\left(1 - \sum_{S\neq\varnothing}\frac{|S|\sqrt{p}}{p-n}\right) \\
&= \frac{1}{2^n}\left(1 - 2^n\frac{\frac{n}{2}\sqrt{p}}{p-n}\right).
\end{aligned}$$

只要 $p > 2^{2n-2}n^2 + 2n$, 便可保证 $P(\mathbf{b}) > 0$.

13. (5 分) 对于随机图上的单调性质, 课上讨论了阈值 (threshold) 与 sharp threshold.

(1) 存在 K_4 (4-clique) 是图的一个单调性质, 因此一定存在阈值. 请找出一个阈值并证明.

(2) 这个性质是否存在 sharp threshold? 请证明你的结论.

本题中请使用 $G(n,p)$ 模型. 阈值和 sharp threshold 的定义分别为.

- $p^*:\mathbb{N}\to\mathbb{R}$ 是性质 $\mathcal{P}$ 的阈值, 如果对任意 $p(n)=o(p^*(n))$, $p'(n)=\omega(p^*(n))$

$$\lim_{n\to\infty}\Pr[G(n,p(n))\in\mathcal{P}]=0,\qquad \lim_{n\to\infty}\Pr[G(n,p'(n))\in\mathcal{P}]=1.$$

- $p^*:\mathbb{N}\to\mathbb{R}$ 是性质 $\mathcal{P}$ 的 sharp threshold, 如果对任意 $\varepsilon>0$

$$\lim_{n\to\infty}\Pr[G(n,(1-\varepsilon)p^*(n))\in\mathcal{P}]=0,\qquad \lim_{n\to\infty}\Pr[G(n,(1+\varepsilon)p^*(n))\in\mathcal{P}]=1.$$

解

(1) 阈值为 $p^*=n^{-2/3}$.

对任意四个点 S, 用 $X_S=\mathbb{1}_{\binom{S}{2}\subseteq E}$ 指示这四个点是否构成 clique. 存在 K_4 等价于 $\sum_S X_S>0$. 首先考虑期望, $\mathbb{E}[\sum_S X_S]=\Theta(n^4p^6)$. 当 $p=o(n^{-2/3})$ 时, $\mathbb{E}[\sum_S X_S]=o(1)$. 由 Markov bound, $\Pr[\sum_S X_S>0]=o(1)$.

反之, 当 $p=\omega(n^{-2/3})$ 时, $\mathbb{E}[\sum_S X_S]=o(1)$. 为了说明这时有 K_4 的概率很大, 我们考虑二阶矩. 当 $p=\omega(n^{-2/3})$ 且 $p=O(n^{-1/2})$ 时

$$\begin{aligned}\operatorname{Var}\Big[\sum_S X_S\Big]&=\sum_S\operatorname{Var}\Big[X_S\Big]+\sum_{S,T\text{ 共用 3 点 3 边}}\operatorname{Cov}(X_S,X_T)+\sum_{S,T\text{ 共用 2 点 1 边}}\operatorname{Cov}(X_S,X_T)\\&=\Theta(n^4)p^6+\Theta(n^5)(p^9-p^{12})+\Theta(n^6)(p^{11}-p^{12})\\&=\Theta(n^4p^6)\end{aligned}$$

因此, 由 Chebyshev bound,

$$\Pr\Big[\sum_S X_S=0\Big]\le\frac{\operatorname{Var}[\sum_S X_S]}{(\mathbb{E}[\sum_S X_S])^2}=o(1).$$

(2) 不存在 sharp threshold.

假设存在 sharp threshold p^*. 那么随着 $n\to\infty$, $G(n,\frac12p^*(n))$ 含有 K_4 的概率趋近于 0, 而 $G(n,2p^*(n))$ 含有 K_4 的概率趋近于 1.

注意到以下方式可以采样 $G(n,\frac12p^*(n))$:

- 第一步: 采样 $G\sim G(n,2p^*(n))$.
- 第二步: 采样 G 的子图 G', 点集不变, 每条边以 $1/4$ 的概率保留.

如果 G 含有 K_4, 那么 G' 含有这个 K_4 的概率不低于 $1/4^6$. 也就是说,

$$\begin{aligned}&\Pr\Big[G(n,\tfrac12p^*(n))\text{ 含有 }K_4\Big]\Big/\Pr\Big[G(n,2p^*(n))\text{ 含有 }K_4\Big]\\&\qquad=\Pr[G'\text{ 含有 }K_4\mid G\text{ 含有 }K_4]\ge1/4^6.\end{aligned}$$

评分标准: 第一问 3 分, 结果 1 分, 证明上下界各 1 分; 第二问 2 分, 其中结果 0.5 分