# 离散傅里叶变换, 马尔科夫链

请在 12 月 19 日课前提交纸质作业.

1. (6 分) 对于一个布尔函数 $f:\{0,1\}^n \to \{0,1\}$, 函数第 $i$ 位的影响被定义为

$$\mathrm{Influence}_i(f) := \Pr_{x\leftarrow\{0,1\}^n}\big[f(x) \neq f(x \oplus e_i)\big],$$

其中 $e_i = (0,\ldots,0,1,0,\ldots,0)$ 只在第 $i$ 位等于 1.

(1) 布尔函数 $f$ 被称为单调（monotone）, 如果 $x \geq y \implies f(x) \geq f(y)$. （这里 $x \geq y$ 表示 $\forall i, x_i \geq y_i$.）请寻找一个单调的布尔函数 $f$, 使得 $\sum_i \mathrm{Influence}_i(f)$ 最大, 并证明.

(2) 布尔函数 $f$ 被称为平衡（balanced）, 如果 $\Pr_{x\leftarrow\{0,1\}^n}[f(x)=1] = \frac{1}{2}$. 请寻找一个平衡的布尔函数 $f$, 使得 $\max_i \mathrm{Influence}_i(f)$ 尽量小, 可以忽略常数系数.

*Remark:* 证明 $\max_i \mathrm{Influence}_i(f)$ 的下界需要非常有技巧地使用傅里叶变换. 本题不需要证明结果最优.

2. (12 分) 对于函数 $f:\{0,1\}^n \to \mathbb{R}$, 用 $f$ 的傅里叶系数表示如下量

(1) $\widehat{g_s}(x)$, 其中 $g(x) := f(x \oplus s)$.

(2) $\widehat{g_y}(x)$, 其中 $g(x) := (-1)^{\langle x,y\rangle} f(x)$.

(3) $\widehat{f_i}(x)$, 其中 $f_i(x) := f(x) - f(x \oplus e_i)$.

(4) $\hat{g}(x)$, 其中 $g(x_1,\ldots,x_k) := f(x_1,\ldots,x_k,0,\ldots,0)$.

(5) $\widehat{g_a}(x)$, 其中 $g(x_1,\ldots,x_k) := \mathbb{E}_{y\leftarrow\{0,1\}^{n-k}}[f(x,y)\chi_a(y)]$.

(6) $\mathrm{Var}[f]$.

这里约定 $\hat{f}(a) = \mathbb{E}_x[f(x)\overline{\chi_a(x)}]$, $f(x) = \sum_a \hat{f}(a)\chi_a(x)$.

3. (4 分) 给定一个函数 $f:\{0,1\}^n \to [-1,1]$, 你能看到若干独立采样 $(x, f(x))$（其中 $x$ 在 $\{0,1\}^n$ 中均匀分布）.

(1) 需要多少个采样才能以 $1-\delta$ 的信心和不错过 $\varepsilon$ 的绝对误差估计 $\hat{f}(a)$. 换言之, 需要设计一个估计方法使得 $\Pr[|(\text{估计值}) - \hat{f}(a)| > \varepsilon] \leq \delta$.

(2) 问题改为估计 $\sum_{a\in\{0,1\}^n} \hat{f}(a)$, 需要多少个采样.

这里约定 $\hat{f}(a) = \mathbb{E}_x[f(x)\overline{\chi_a(x)}]$, $f(x) = \sum_a \hat{f}(a)\chi_a(x)$.

提示：可以使用上一次作业中 Chernoff bound 的变种.

4. (10 分) 令 $X_1,\ldots,X_{2n}$ i.i.d. 服从 Bern(1/2) 分布. 我们要选取一组系数 $c_1,\ldots,c_{2n} \in \mathbb{Z}$, 使得 $\sum_i c_i X_i$ 接近均匀分布. 当然, 并不存在 $\mathbb{Z}$ 上的均匀分布, 我们实际的要求是统计距离

$$\Delta\Big(\sum_i c_i X_i, \sum_i c_i X_i + 1\Big) \leq 2^{-\lambda}. \tag{*}$$

一种显然的做法, 是令 $n = \lambda/2$, 令 $c_i = 2^{i-1}$; 这样 $\sum_i c_i X_i$ 服从 $\{0, 1, \ldots, 2^\lambda - 1\}$ 上的均匀分布, 而 $\sum_i c_i X_i + 1$ 服从 $\{1, 2, \ldots, 2^\lambda\}$ 上的均匀分布, 满足我们对统计距离的要求.

进一步, 假设 $X_1, \ldots, X_n$ 中有一半值被泄漏. 要求即使已经看到泄露值, (*) 仍然成立.

为了方便分析, 我们令 $c_1, \ldots, c_n$ 是 i.i.d. 从某个分布 $P_C$ 中选取的. 这样不管哪部分值泄露, 分析都相同. 不失一般性, 可以假设前一半值没有泄露. 定义函数

$$\mathrm{Err}(c_1, \ldots, c_n) = \Delta\Big(\sum_i c_i X_i, \sum_i c_i X_i + 1\Big)$$

我们要求当 $c_1, \ldots, c_n$ 是从 $(P_C)^n$ 选取时, 以 $1 - 2^{-\lambda}$ 的概率（这个随机性只依赖于 $c_1, \ldots, c_n$）

$$\mathrm{Err}(c_1, \ldots, c_n) \le 2^{-\lambda}.$$

请根据 $\lambda$, 选取合适的 $n$ 以及分布 $P_C$, 使得要求被满足. 请让 $n$ 的取值尽量小, 可以忽略常数系数. 建议选取 $P_C$ 为 $\{1, 2, 3, \ldots, B\}$ 上的均匀分布, 其中 $B = 2^{O(\lambda)}$ 根据 $\lambda$ 选取.

5. (6 分) 考虑一个有限状态空间 $\Omega$ 上不可约的（irreducible）马尔可夫核 $P$. 我们知道 $P$ 存在唯一的稳态分布（stationary distribution）$\pi$. 证明对于任意初始分布 $\mu$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mu P^j = \pi.$$

6. (4 分) 考虑大小为 $n$ 的有限状态空间 $\Omega$ 上的一个不可约（irreducible）马尔可夫核 $P$, 令 $\pi$ 是稳态分布.

   (1) 证明 $P$ 只有一个特征值等于 1.

   (2) 假设 $P$ 有周期 $T > 1$. 这里 $T = \gcd\{t : \exists x, P^t(x|x) > 0\}$. 不难证明, 周期性说明状态空间可以划分为 $T$ 个非空子集 $\Omega_0, \Omega_1, \ldots, \Omega_{T-1}$ 满足

   $$\forall x, y \in \Omega,\ \forall j \in \mathbb{Z}_T,\ P(y|x) > 0 \wedge x \in \Omega_j \implies y \in \Omega_{j+1}.$$

   证明 1 的所有 $T$ 次根 $e^{2\pi i \frac{k}{T}}$（for $k \in \mathbb{Z}_T$）都是 $P$ 的特征值.