集中不等式, 离散傅里叶变换

参考答案

1. (10 分) 完成以下关于 Chernoff bound 的证明.

(1) (0 分) 设随机变量 $(X_1,\dots,X_n)\sim(\mathrm{Bern}(p))^n$, 即它们独立地服从 $\mathrm{Bern}(p)$. 对任意 $t>0$,

$$\Pr\Big[\frac{X_1+\cdots+X_n}{n}\geq q\Big]=\Pr[e^{t(X_1+\cdots+X_n)}\geq e^{tqn}]\underset{\text{Markov's bound}}{\leq}\frac{\mathbb{E}[e^{t(X_1+\cdots+X_n)}]}{e^{tqn}}=\Big(\frac{\mathbb{E}[e^{tX_1}]}{e^{tq}}\Big)^n.$$

当 $0\leq p\leq q\leq 1$ 时, 请选取合适的 t 使得上式最紧. 得到的结果应为

$$\Pr\Big[\frac{X_1+\cdots+X_n}{n}\geq q\Big]\leq\exp(-n\cdot d(q\|p)).$$

Remark: 对称地, 当 $0\leq q\leq p\leq 1$ 时, 可以证明

$$\Pr\Big[\frac{X_1+\cdots+X_n}{n}\leq q\Big]\leq\exp(-n\cdot d(q\|p)).$$

(2) 设 $(X_1,\dots,X_n)\sim P_1P_2\dots P_n$, 即它们相互独立. 每个 P_i 都是 $[0,1]$ 上的期望等于 p 的分布. 证明当 $0\leq p\leq q\leq 1$ 时,

$$\Pr\Big[\frac{X_1+\cdots+X_n}{n}\geq q\Big]\leq\exp(-n\cdot d(q\|p)).$$

提示: 比较 $\mathbb{E}_{X\sim P_i}[e^{tX}]$ 和 $\mathbb{E}_{X\sim\mathrm{Bern}(p)}[e^{tX}]$ 的大小.

(3) 有 $m>n$ 个球, 其中 pm 个是白球. 从中无放回的随机选取 n 个球. 用随机变量 $(X_1,\dots,X_n)$ 表示这 n 次选取的结果. $X_i=1$ 表示第 i 个球是白球, $X_i=0$ 表示第 i 个球不是白球. 显然 $\mathbb{E}[X_i]=p$. 证明当 $0\leq p\leq q\leq 1$ 时,

$$\Pr\Big[\frac{X_1+\cdots+X_n}{n}\geq q\Big]\leq\exp(-n\cdot d(q\|p)).$$

解

(1) 定义 $f(t)=\ln\frac{\mathbb{E}[e^{tX_1}]}{e^{tq}}=\ln(pe^t+1-p)-tq$. 对 f 求导

$$f'(t)=\frac{pe^t}{pe^t+1-p}-q.$$

f' 单调递增, 且存在唯一 t^* 使得 $f'(t^*)=0$. 这说明 f 的最小值点为 $t^*=\ln(\frac{q}{1-q}\frac{1-p}{p})>0$.

$$\min_{t>0}f(t)=f(t^*)=\ln\Big(p\frac{q}{1-q}\frac{1-p}{p}+1-p\Big)-q\ln\Big(\frac{q}{1-q}\frac{1-p}{p}\Big)=-d(p\|q).$$

(2) 只需说明 $\mathbb{E}_{X\sim P_i}[e^{tX}]\leq\mathbb{E}_{X\sim\mathrm{Bern}(p)}[e^{tX}]$. 证明的其余部分和 Bernoulli 分布的情况相同.

不妨定义一个从 $[0,1]$ 到 $\{0,1\}$ 的 kernel $P_{Y|X}$ 使得 $P_{Y|X=x}=\mathrm{Bern}(x)$. 换言之,

$$P_{Y|X}(y|x)=\begin{cases}x, & \text{if } y=1\\ 1-x, & \text{if } y=0\end{cases}$$

对任意 $[0,1]$ 上期望等于 p 的分布 P_X, 考虑 $(X,Y)\sim P_XP_{Y|X}$. 因为 $\mathbb{E}[Y]=\mathbb{E}[\mathbb{E}[Y|X]]=\mathbb{E}[X]=p$, 所以 $Y\sim\mathrm{Bern}(p)$. 对任何 $x\in\mathrm{Supp}(P_i)$, 因为指数函数的凸性,

$$\mathbb{E}[e^{tY}|X=x]\geq e^{t\,\mathbb{E}[Y|X=x]}=e^{tx}.$$

进而

$$\mathbb{E}[e^{tY}]=\mathbb{E}[\mathbb{E}[e^{tY}|X]]\geq\mathbb{E}[e^{tX}].$$

(3) 只需证明对任何 $t>0$,

$$\mathbb{E}[e^{t(X_1+\cdots+X_n)}]\leq\mathop{\mathbb{E}}_{X\sim\mathrm{Bern}(p)}[e^{tX}]^n.$$

证明的其余部分和 Bernoulli 分布的情况相同.

我们递归地证明这个命题. 当 $n=1$ 时, 命题显然成立. 下面假设命题对 $n-1$ 成立, 我们证明命题对 n 也成立. 考虑 X_n 的条件分布,

$$\Pr[X_n=1|X_1+\cdots+X_{n-1}=s]=\frac{pm-s}{m-n+1}.$$

不难看出 s 越大, X_n 的条件期望就越小. 类似地, $e^{t(X_1+\cdots+X_{n-1})}$ 越大, $e^{t(X_n)}$ 的条件期望就越小. 因此

$$\begin{aligned}\mathbb{E}[e^{t(X_1+\cdots+X_n)}]&=\mathbb{E}[e^{t(X_1+\cdots+X_{n-1})}\,\mathbb{E}[e^{t(X_n)}|e^{t(X_1+\cdots+X_{n-1})}]]\\&\leq\mathbb{E}[e^{t(X_1+\cdots+X_{n-1})}]\,\mathbb{E}[e^{t(X_n)}]\leq\mathop{\mathbb{E}}_{X\sim\mathrm{Bern}(p)}[e^{tX}]^n.\end{aligned}$$

第一个不等号可以抽象化为一个类似排序不等式的引理: 对任意非负实数上的分布 P 和任意单调递减的函数 f, $\mathbb{E}_{X\sim P}[Xf(X)]\leq\mathbb{E}_{X\sim P}[X]\,\mathbb{E}_{X\sim P}[f(X)]$. 其中 X 对应 $e^{t(X_1+\cdots+X_n)}$, $f(v):=\mathbb{E}[e^{t(X_n)}|e^{t(X_1+\cdots+X_{n-1})}=v]$. 引理的证明如下:

$$\begin{aligned}\mathop{\mathbb{E}}_{X\sim P}[X]\mathop{\mathbb{E}}_{X\sim P}[f(X)]&=\sum_x P(x)x\sum_y P(y)f(y)\\&=\sum_x P^2(x)xf(x)+\sum_{x<y}P(x)P(y)xf(y)+\sum_{x>y}P(x)P(y)xf(y)\\&=\sum_x P^2(x)xf(x)+\sum_{x<y}\Big(P(x)P(y)xf(y)+P(x)P(y)yf(x)\Big)\\&\leq\sum_x P^2(x)xf(x)+\sum_{x<y}\Big(P(x)P(y)xf(x)+P(x)P(y)yf(y)\Big)\\&=\sum_x P^2(x)xf(x)+\sum_{x\neq y}P(x)P(y)xf(x)\\&=\sum_x\sum_y P(x)P(y)xf(x)\\&=\sum_x P(x)xf(x)\\&=\mathop{\mathbb{E}}_{X\sim P}[Xf(X)]\end{aligned}$$

2. (10 分) 根据 Sanov's Theorem 我们可以看出, Chernoff bound 对于

$$\Pr_{(X_1,\ldots,X_n)\sim(\mathrm{Bern}(p))^n}\Big[\frac{X_1+\cdots+X_n}{n}\ge q\Big]$$

的估计已经很精确, 指数上的系数是紧的. 这个估计对非 Bernoulli 分布是否也同样精确?

考虑有限个正实数上的分布 P. 记 $\mathrm{Supp}(P)=\{v_1,\ldots,v_T\}\subseteq\mathbb{R}^+$. 记 $p_i:=P(v_i)>0$. 这个分布的期望是 $\bar{v}=\sum p_iv_i$. 考虑任意 $b\in(\bar{v},\max_i v_i)$, 定义

$$Q^*=\arg\min_{\substack{\text{分布 } Q\\ \mathbb{E}_{X\sim Q}[X]\ge b}} D(Q\|P).$$

根据 Sanov's Theorem,

$$\Pr_{(X_1,\ldots,X_n)\sim P^n}\Big[\frac{X_1+\cdots+X_n}{n}\ge b\Big]\le(n+1)^T\cdot\exp(-n\cdot D(Q^*\|P)).$$

而根据 Chernoff bound,

$$\Pr_{(X_1,\ldots,X_n)\sim P^n}\Big[\frac{X_1+\cdots+X_n}{n}\ge b\Big]\le\min_{t>0}\Big(\frac{\mathbb{E}_{X\sim P}[e^{tX}]}{e^{tb}}\Big)^n.$$

请问是否存在 P 和 $b\in(\bar{v},\max_i v_i)$ 使得 Chernoff bound 的估计要弱于 Sanov's Theorem?

提示: 拉格朗日乘数.

解 不存在.

首先考虑 Sanov's Theorem 一边. 用 $q_1,\ldots,q_T$ 表示 Q 分布下的概率. Q^* 是如下优化问题的解

$$\begin{aligned}\text{最小化}\quad & f(q_1,\ldots,q_T):=\sum_i q_i\ln\Big(\frac{q_i}{p_i}\Big)=D(Q\|P)/\log e\\ \text{约束:}\quad & (1)\ \sum_i q_iv_i\ge b\\ & (2)\ \sum_i q_i=1\\ & (3)\ \forall i\quad q_i\ge 0\end{aligned}$$

约束条件是有界闭集而 f 连续, 所以最小值一定存在. 对最小值点 Q^* 来说, 条件 (2) 显然是紧的. 条件 (1) 也是紧的, 因为从任何 P 到 Q 的连线上, $f(\varepsilon Q+(1-\varepsilon)P)$ 都单调地增长 ($\varepsilon\in[0,1]$). 而条件 (3) 是松的, 不妨考虑任何一个满足 (1)(2) 的分布 Q 且满足 $q_i=0$, 我们来说明 Q 不是极小值点. 计算偏导

$$\frac{\partial f}{\partial q_i}=\ln\Big(\frac{q_i}{p_i}\Big)+1,$$

在 $q_i=0$ 的位置, $\frac{\partial f}{\partial q_i}(Q)=-\infty$, 这提示我们微调 q_i 的值会使函数值更小. 严格来说, 可以分两种情况讨论

- 如果存在 j,k 使得 $q_j,q_k>0$: 定义 Q_ε 为

$$Q_\varepsilon(v_x)=\begin{cases}\varepsilon, & \text{if } x=i\\ q_j+C_j\varepsilon, & \text{if } x=j\\ q_k+C_k\varepsilon, & \text{if } x=k\\ Q(v_x)=q_x, & \text{otherwise}\end{cases}\qquad \text{其中 } C_j,C_k \text{ 是 } \begin{array}{c}1+C_j+C_k=0\\ v_i+v_jC_j+v_kC_k=0\end{array} \text{ 的解}$$

这样对足够小的 $\varepsilon \geq 0$, Q_ε 满足 (1)(2)(3). 同时 $Q_0 = Q$, $\frac{d}{d\varepsilon} f(Q_\varepsilon)|_{\varepsilon=0} = -\infty$. 因此 Q 不是极小值点.

- 如果存在 k 使得 $q_k = 1$: 根据现有的条件, 这说明 $v_k = b \in (\min_x v_x, \max_x v_x)$. 因此一定存在 j 使得 $v_i < v_k < v_j$ 或 $v_i > v_k > v_j$. 定义 Q_ε 为

$$Q_\varepsilon(v_x) = \begin{cases} C_i\varepsilon, & \text{if } x = i \\ C_j\varepsilon, & \text{if } x = j \\ 1-\varepsilon, & \text{if } x = k \\ Q(v_x) = 0, & \text{otherwise} \end{cases} \qquad \text{其中 } C_i, C_j \text{ 是 } \begin{matrix} C_i + C_j - 1 = 0 \\ v_iC_i + v_jC_j - v_k = 0 \end{matrix} \text{ 的解}$$

  同样的论证可以说明 Q 不是极小值点.

使用拉格朗日乘数法, 存在 A, B 使得

$$\frac{\partial}{\partial q_i}\Big(f(Q) + A\sum_j q_jv_j + B\sum_j q_j\Big) = \ln\Big(\frac{q_i}{p_i}\Big) + 1 + Av_i + B$$

在极小值点等于 0. 不妨令 $t = -A$, 那么在极小值点处

$$q_i = p_ie^{-1-Av_i-B} \propto p_ie^{tv_i}.$$

因为 Q 是概率, 所以 B 的取值一定会令

$$q_i = \frac{p_ie^{tv_i}}{\sum_j p_je^{tv_j}}.$$

定义 $C_t = \sum_j p_je^{tv_j}$, 定义 Q_t 为 $Q_t(v_i) = p_ie^{tv_i}/C_t$. 已经证明存在 t 使得 $Q^* = Q_t$. 注意到 Q_t 的期望随着 t 严格单调增加, 因此存在唯一的 $t^* > 0$ 满足

$$b = \mathop{\mathbb{E}}_{X\sim Q_{t^*}}[X] = \frac{\sum_i v_ip_ie^{t^*v_i}}{\sum_i p_ie^{t^*v_i}}$$

同一个 t 使得 $Q^* = Q_{t^*}$ 成立.

再考虑 Chernoff bound 一边.

$$\Pr_{(X_1,\ldots,X_n)\sim P^n}\Big[\frac{X_1+\cdots+X_n}{n} \geq b\Big] \leq \min_{t>0}\Big(\frac{\mathbb{E}_{X\sim P}[e^{tX}]}{e^{tb}}\Big)^n \leq \Big(\frac{\mathbb{E}_{X\sim P}[e^{t^*X}]}{e^{t^*b}}\Big)^n.$$

只需再证明

$$\ln\Big(\frac{\mathbb{E}_{X\sim P}[e^{t^*X}]}{e^{t^*b}}\Big) = -D(Q^*\|P)/\log e.$$

简单验证即可

$$\text{左边} = \ln(\mathop{\mathbb{E}}_{X\sim P}[e^{t^*X}]) - t^*b = \ln\Big(\sum_i p_ie^{t^*v_i}\Big) - t^*b = \ln C_{t^*} - t^*b$$

$$\text{右边} = -\sum_i Q^*(v_i)\ln\frac{Q^*(v_i)}{p_i} = -\sum_i Q^*(v_i)\ln\frac{e^{t^*v_i}}{C_{t^*}} = \ln C_{t^*} - t^*\sum_i Q^*(v_i)v_i = \ln C_{t^*} - t^*b$$

3. (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.

解

(1) 对于单调布尔函数 f,

$$\begin{aligned}\sum_i \mathrm{Influence}_i(f) &= \frac{1}{2^n}\sum_i\sum_x \mathbb{1}\big[f(x) \neq f(x\oplus e_i)\big] \\ &= \frac{1}{2^{n-1}} \sum_{i,x \text{ s.t. } x_i=0} \big(f(x\oplus e_i) - f(x)\big) \\ &= \frac{1}{2^{n-1}} \sum_{i,x} \begin{cases} f(x), & \text{if } x_i = 1 \\ -f(x), & \text{if } x_i = 0 \end{cases} \\ &= \frac{1}{2^{n-1}} \sum_x \big(x \text{ 中 1 的个数} - x \text{ 中 0 的个数}\big).\end{aligned}$$

因此, 使 $\sum_i \mathrm{Influence}_i(f)$ 最大的 f 即为 majority 函数.

(2) 把 n 个 bit 分成 a 组 $G_1,\cdots,G_a$, 每组大约为 b 个 bit, 忽略常数项, 即 $ab=n$. 取 $f(x)$ 为 $[\exists i \in [a], \forall j \in G_i, x_j = 1]$. 从而有,

$$\Pr_{x\leftarrow\{0,1\}^n}[f(x)=1] = 1-(1-2^{-b})^a = \frac{1}{2}.$$

可推得 $b = \log n - \log(\log n) + O(1)$, 对于 $\mathrm{Influence}_i$ 仅需考虑 i 所在的组全是 1 的 x, 因此,

$$\max_i \mathrm{Influence}_i(f) = 2^{1-b} = O(\frac{\log n}{n}).$$

4. (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 $g_s(x) := f(x \oplus s)$.

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

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

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

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

(6) $\mathrm{Var}[f(X)]$, 其中 X 服从均匀分布.

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

解

(1) $\widehat{g_s}(x) = \mathbb{E}_y\left[(-1)^{\langle x,y\rangle} f(y\oplus s)\right] = (-1)^{\langle x,s\rangle}\mathbb{E}_y\left[(-1)^{\langle x,y\rangle} f(y)\right] = (-1)^{\langle x,s\rangle}\widehat{f}(x)$.

(2) $\widehat{g_s}(x) = \mathbb{E}_y\left[(-1)^{\langle x\oplus s,y\rangle} f(y)\right] = \widehat{f}(x\oplus s)$.

(3) $\widehat{f_i}(x) = \widehat{f}(x) - \widehat{g_{e_i}}(x) = (1-(-1)^{x_i})\,\widehat{f}(x)$.

(4) $\widehat{g_k}(x) = \mathbb{E}_y\left[(-1)^{\langle x,y\rangle} f(y,0)\right] = \mathbb{E}_y\left[\sum_z (-1)^{\langle y,x\oplus z\rangle}\sum_w \widehat{f}(z,w)\right] = \sum_z\sum_w \widehat{f}(z,w)\mathbb{E}_y\left[(-1)^{\langle y,x\oplus z\rangle}\right] = \sum_w \widehat{f}(x,w)$.

(5) $\widehat{g_a}(x) = \mathbb{E}_y\left[(-1)^{\langle x,y\rangle}\mathbb{E}_z\left[(-1)^{\langle a,z\rangle} f(y,z)\right]\right] = \widehat{f}(x,a)$.

(6) $\mathrm{Var}[f] = \mathbb{E}_x\left[f(x)^2\right] - \mathbb{E}_x[f(x)]^2 = \sum_{x\neq 0}\widehat{f}(x)^2$.

5. (5 分) 对于函数 $f:\{0,1\}^n\to\mathbb{R}$, 它的傅里叶系数 $\hat{f}:\{0,1\}^n\to\mathbb{R}$ 满足

$$\hat{f}(y) = \frac{1}{2^n}\sum_x f(x)\chi_y(x), \qquad f(x) = \sum_y \hat{f}(y)\chi_y(x).$$

考虑一个"噪音算子"$T_\rho : (\{0,1\}^n\to\mathbb{R})\to(\{0,1\}^n\to\mathbb{R})$, 其中 $\rho\in[0,1]$.

$$T_\rho(f)(x) = \mathop{\mathbb{E}}_{y\sim(\mathrm{Bern}(\rho))^n}\Big[f(x\oplus y)\Big].$$

求 $\widehat{T_\rho(f)}(y)$. 化简后的表达式不应该出现 $\sum$ 或 $\mathbb{E}$.

解

$$\begin{aligned}
\widehat{T_\rho(f)}(y) &= \mathop{\mathbb{E}}_x\Big[T_\rho(f)(x)\chi_y(x)\Big] \\
&= \mathop{\mathbb{E}}_x \mathop{\mathbb{E}}_{z\sim(\mathrm{Bern}(\rho))^n}\Big[f(x+z)\chi_y(x)\Big] \\
&= \mathop{\mathbb{E}}_x \mathop{\mathbb{E}}_{z\sim(\mathrm{Bern}(\rho))^n}\Big[f(x)\chi_y(x+z)\Big] \\
&= \mathop{\mathbb{E}}_x \mathop{\mathbb{E}}_{z\sim(\mathrm{Bern}(\rho))^n}\Big[f(x)\chi_y(x)\chi_y(z)\Big] \\
&= \mathop{\mathbb{E}}_x\Big[f(x)\chi_y(x)\Big]\mathop{\mathbb{E}}_{z\sim(\mathrm{Bern}(\rho))^n}\Big[\chi_y(z)\Big] \\
&= \hat{f}(y)\prod_i \mathop{\mathbb{E}}_{z_i\sim\mathrm{Bern}(\rho)}[(-1)^{y_i z_i}] \\
&= \hat{f}(y)(1-2\rho)^{\|y\|_1}
\end{aligned}$$

6. (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 $\sum_i c_iX_i$ 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

$$\Delta\Big(\sum_i c_iX_i, \sum_i c_iX_i+1\Big)\le 2^{-\lambda}. \tag{*}$$

一种显然的做法, 是令 $n=\lambda/2$, 令 $c_i=2^{i-1}$; 这样 $\sum_i c_iX_i$ 服从 $\{0,1,\ldots,2^\lambda-1\}$ 上的均匀分布, 而 $\sum_i c_iX_i+1$ 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

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

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 λ, 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 λ 选取.

解 在本答案中, 我们定义傅立叶变换和逆傅立叶变换为,

$$\hat f(y)=\sum_{x\in\mathbb{Z}_N} f(x)e^{-2\pi i\frac{xy}{N}}, \qquad f(x)=\frac{1}{N}\sum_{y\in\mathbb{Z}_N}\hat f(y)e^{2\pi i\frac{xy}{N}}.$$

为了能使用离散傅立叶分析, 我们选取一个足够大的 $N>nB+1$. 这样, 无论 c_i, X_i 如何取值, 总有 $\sum_i c_iX_i, \sum_i c_iX_i+1<N$. 因此可以把它们看成 $\mathbb{Z}_N$ 上的分布.

用 $\sigma_c:\mathbb{Z}_N\to\mathbb{R}$ 表示 $\{c\}$ 上的退化分布, $\tau_c=\frac{1}{2}(\sigma_0+\sigma_c)$ 表示 $\{0,c\}$ 的均匀分布, 也就是

$$\sigma_c(x)=\begin{cases}1, & \text{if } x=c\\ 0, & \text{otherwise}\end{cases} \qquad \tau_c(x)=\begin{cases}\frac{1}{2}, & \text{if } x\in\{0,c\}\\ 0, & \text{otherwise}\end{cases}$$

那么 $\sum_i c_iX_i$ 的分布就是 $\tau_{c_1}*\tau_{c_2}*\cdots*\tau_{c_n}$, 这里 $*$ 表示卷积. 类似地, $\sum_i c_iX_i+1$ 的分布是 $\tau_{c_1}*\tau_{c_2}*\cdots*\tau_{c_n}*\sigma_1$. 这两个分布间的统计距离就是

$$\Delta\Big(\sum_i c_iX_i, \sum_i c_iX_i+1\Big)=\frac{1}{2}\big\|\tau_{c_1}*\tau_{c_2}*\cdots*\tau_{c_n}-\tau_{c_1}*\tau_{c_2}*\cdots*\tau_{c_n}*\sigma_1\big\|_1=\frac{1}{2}\big\|\tau_{c_1}*\tau_{c_2}*\cdots*\tau_{c_n}*(\sigma_0-\sigma_1)\big\|_1$$

为了估计 L1 距离的上界, 只需对 L2 距离有足够紧的估计. 而计算 L2 距离可以利用傅立叶系数. 令 $f=\tau_{c_1}*\tau_{c_2}*\cdots*\tau_{c_n}*(\sigma_0-\sigma_1)$, 那么 $\hat f=\widehat{\tau_{c_1}}\cdot\widehat{\tau_{c_2}}\cdot\ldots\widehat{\tau_{c_n}}\cdot\widehat{\sigma_0-\sigma_1}$. 对每个 $a\in\mathbb{Z}_N$,

- $|\hat f(a)|\le|\widehat{\sigma_0-\sigma_1}(a)|\le\frac{2\pi}{N}\cdot|a|$, 其中 $|a|:=\min\{a,N-a\}$. 当 $|a|$ 较小时, 这是一个比较好的估计.

- 当 $|a|$ 较大时, 使用另一个估计. 注意到

  $$\widehat{\tau_c}(a) = \frac{1}{2}(1 + e^{-2\pi i \frac{ac}{N}})$$

  而 c 在 $\{1, \ldots, B\}$ 中均匀取值. 这样当 $|a|$ 较大时, $\widehat{\tau_c}(a)$ 在复平面上, 以 $1/2$ 为中心, 以 $1/2$ 为半径的圆上"均匀"分布. 因此以 $\Omega(1)$ 概率, $\widehat{\tau_c}(a)$ 的绝对值小于 $1 - \Omega(1)$.

  具体来说,

  $$\begin{aligned}
  \mathop{\mathbb{E}}_{c \leftarrow \{1,\ldots,B\}}\left[|\widehat{\tau_c}(a)|^2\right] = \frac{1}{B}\sum_{c=1}^{B}|\widehat{\tau_c}(a)|^2 &= \frac{1}{B}\sum_{c=1}^{B}\frac{1}{4}(1 + e^{-2\pi i \frac{ac}{N}})(1 + e^{2\pi i \frac{ac}{N}}) \\
  &= \frac{1}{B}\sum_{c=1}^{B}\frac{1}{4}(2 + e^{-2\pi i \frac{ac}{N}} + e^{2\pi i \frac{ac}{N}}) \\
  &= \frac{1}{2} + \frac{1}{B}\frac{1}{4}\Big(\frac{1 - e^{-2\pi i \frac{aB}{N}}}{e^{2\pi i \frac{a}{N}} - 1} + \frac{1 - e^{2\pi i \frac{aB}{N}}}{e^{-2\pi i \frac{a}{N}} - 1}\Big) \\
  &\le \frac{1}{2} + \frac{1}{B}\frac{1}{|e^{2\pi i \frac{a}{N}} - 1|} \\
  &\le \frac{1}{2} + \frac{1}{B}\frac{1}{2 \cdot |\sin(\pi \frac{a}{N})|}
  \end{aligned}$$

  只要 $|a| \ge N/B$ 且 $B \ge 2$

  $$\mathop{\mathbb{E}}_{c \leftarrow \{1,\ldots,B\}}\left[|\widehat{\tau_c}(a)|^2\right] \le \frac{1}{2} + \frac{1}{B}\frac{1}{2 \cdot |\sin(\pi \frac{a}{N})|} \le \frac{1}{2} + \frac{1}{B}\frac{1}{2\sin(\pi/B)} \le \frac{3}{4}.$$

  根据 Chernoff bound,

  $$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}|\widehat{\tau_{c_i}}(a)|^2 \ge 7/8\right] \le e^{-n/64}.$$

  利用算术平均和几何平均的关系,

  $$\Pr\left[\prod_{i=1}^{n}|\widehat{\tau_{c_i}}(a)|^2 \ge (7/8)^n\right] \le e^{-n/64}.$$

  这样以 $1 - N \cdot e^{-n/64}$ 的概率, 对所有 $a \in [N/B, N - N/B]$, 均有

  $$|\hat{f}(a)| = |\widehat{\sigma_0 - \sigma_1}(a)| \cdot \prod_{i=1}^{n}|\widehat{\tau_{c_i}}(a)| \le \prod_{i=1}^{n}|\widehat{\tau_{c_i}}(a)| \le (7/8)^{n/2}.$$

综合两种情况, 以 $1 - N \cdot e^{-n/64}$ 的概率

$$\begin{aligned}
\|f\|_2^2 = \sum_x f^2(x) = \frac{1}{N}\sum_a \hat{f}^2(a) &= \frac{1}{N}\sum_{a:|a|<N/B}\hat{f}^2(a) + \frac{1}{N}\sum_{a:|a|\ge N/B}\hat{f}^2(a) \\
&\le \frac{1}{N}\sum_{a:|a|<N/B}\Big(\frac{2\pi}{N}\cdot|a|\Big)^2 + \frac{1}{N}\sum_{a:|a|\ge N/B}(7/8)^n \\
&\le \frac{8\pi^2}{B^3} + (7/8)^n.
\end{aligned}$$

根据 Cauchy-Schwarz 不等式, 此时

$$\mathrm{Err}(c_1, \ldots, c_n) = \frac{1}{2}\|f\|_1 \le \frac{1}{2}\sqrt{N}\cdot\|f\|_2 \le \frac{1}{2}\sqrt{N\Big(\frac{8\pi^2}{B^3} + (7/8)^n\Big)}$$

为了达到题设的要求, 只需要 n, N, B 满足

$$1 - N \cdot e^{-n/64} \geq 1 - 2^{-\lambda}, \qquad \frac{1}{2}\sqrt{N\left(\frac{8\pi^2}{B^3} + (7/8)^n\right)} \leq 2^{-\lambda}, \qquad N > nB + 1, \qquad B \geq 2.$$

不难验证, 存在 $n = O(\lambda), B = O(2^\lambda \lambda), N = nB + 1$ 使得需要的条件均被满足.