第六讲

王伟文　暨南大学

1　有界差不等式

我们称函数 $\mathbb{R}^d \to \mathbb{R}$ 满足以 $(L_1, L_2, \cdots, L_d)$ 为参数的有界差不等式, 若对于任意坐标 $k \in [d]$, 存在参数 $L_k > 0$ 使得

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}^{\setminus k})| \leq L_k \quad \forall \boldsymbol{x} \in \mathbb{R}^d, \forall \boldsymbol{x}^{\setminus k} \in \{(x_1, x_2, \cdots, \underbrace{x'}_{k}, \cdots, x_d) : x' \in \mathbb{R}\}$$

引理 1.1

若 f 满足以 $(L_1, L_2, \cdots, L_d)$ 为参数的有界差分不等式, 随机向量 $X = (X_1, X_2, \cdots, X_d)$ 各分量相互独立, 则

$$\mathbb{P}\left[|f(X) - \mathbb{E}\left[f(X)\right]| \geq t\right] \leq 2e^{-\frac{2t^2}{\sum_{i=1}^{d} L_i^2}} \quad \forall t \geq 0.$$

例 1.1. 设 $X_i \in [a, b]$ 且相互独立 $(i = 1, 2, \cdots, n)$, $\mathbb{E}X_i = \mu_i$. 记 $X = (X_1, \ldots, X_n)$, $f(X) = \sum_{i=1}^{n}(X_i - \mu_i)$.

$$\begin{aligned}|f(x_1, \cdots, x_k, \cdots, x_n) - f(x_1, \cdots, x_k', \cdots, x_n)| &= |(x_k - \mu_k) - (x_k' - \mu_k)| \\ &= |x_k - x_k'| \leq b - a \quad \forall k \in [n].\end{aligned}$$

故 f 满足有界差不等式, 因此

$$\mathbb{P}\left[|f(X) - \mathbb{E}\left[f(X)\right]| \geq t\right] = \mathbb{P}\left[\left|\sum_{i=1}^{n}(X_i - \mu_i)\right| \geq t\right] \leq 2e^{-\frac{2t^2}{n(b-a)^2}} \quad \forall t \geq 0.$$

例 1.2. 随机图中的圈数. 给定无向图 $G = (V, E)$, 其中 $V = \{1, 2, \ldots, d\}$ 表示图中顶点集合, E 表示图中连边的集合. 图上的圈 C 为顶点集合的子集, 其中任意两个元素存在连边. 图的圈数 $C(G)$ 为图上最大圈的元素个数. $C(G) \in [1, d]$.

在 Erdöds-Rényi 模型中, 随机图由 $m := \binom{d}{2}$ 维的随机向量 $Z = (X_{ij})_{i<j}$ 刻画, $X_{ij} \sim \text{Ber}(p)$ 且相互独立, 指示节点 i 与 j 之间是否存在连边, $C(G)$ 为随机变量. 显然 Z 唯一确定 $C(G)$, 因此 $C(G)$ 可以表示为 Z 的函数 $f : \{0, 1\}^m \to [d]$. 容易确定, 对任意 $k \in [m]$

$$|f(z_1, \cdots, z_k, \cdots, z_m) - f(z_1, \cdots, z_k', \cdots, z_m)| \leq 1.$$

故 f 满足有界差不等式, 由引理 1.1

$$\mathbb{P}\left[\frac{1}{m}|C(G)-\mathbb{E}\left[C(G)\right]|\geq\delta\right]\leq 2e^{-2m\delta^2}\quad\forall\delta\geq 0.$$

例 1.3. Rademacher 复杂度. 设 $\mathcal{A}$ 为 n 维实空间的子集 $\mathcal{A}\subset\mathbb{R}^n$, $\boldsymbol{\epsilon}=(\epsilon_i)_{i\in[n]}$ 为一组相互独立的 Rademacher 随机变量, 定义随机变量

$$Z=\sup_{\boldsymbol{a}\in\mathcal{A}}[\langle\boldsymbol{a},\boldsymbol{\epsilon}\rangle]=\sup_{\boldsymbol{a}\in\mathcal{A}}\sum_{i=1}^{n}a_i\epsilon_i$$

称 $\mathcal{R}(\mathcal{A}):=\mathbb{E}Z$ 为集合 $\mathcal{A}$ 的 Rademacher 复杂度.

记 $Z=f(\epsilon_1,\ldots,\epsilon_n)$.

$$\langle\boldsymbol{a},\boldsymbol{\epsilon}\rangle-\langle\boldsymbol{a},\boldsymbol{\epsilon}^{\backslash k}\rangle=a_k(\epsilon_k-\epsilon_k')$$

故

$$\sup_{\boldsymbol{a}\in\mathcal{A}}\{\langle\boldsymbol{a},\boldsymbol{\epsilon}\rangle-\langle\boldsymbol{a},\boldsymbol{\epsilon}^{\backslash k}\rangle\}=\sup_{\boldsymbol{a}\in\mathcal{A}}a_k(\epsilon_k-\epsilon_k')\leq 2\sup_{\boldsymbol{a}\in\mathcal{A}}|a_k|$$

因为

$$\langle\boldsymbol{a},\boldsymbol{\epsilon}\rangle-\sup_{\boldsymbol{a}'\in\mathcal{A}}\langle\boldsymbol{a}',\boldsymbol{\epsilon}^{\backslash k}\rangle\leq\langle\boldsymbol{a},\boldsymbol{\epsilon}\rangle-\langle\boldsymbol{a},\boldsymbol{\epsilon}^{\backslash k}\rangle$$

$$\sup_{\boldsymbol{a}\in\mathcal{A}}\{\langle\boldsymbol{a},\boldsymbol{\epsilon}\rangle-\sup_{\boldsymbol{a}'\in\mathcal{A}}\langle\boldsymbol{a}',\boldsymbol{\epsilon}^{\backslash k}\rangle\}\leq\sup_{\boldsymbol{a}\in\mathcal{A}}\{\langle\boldsymbol{a},\boldsymbol{\epsilon}\rangle-\langle\boldsymbol{a},\boldsymbol{\epsilon}^{\backslash k}\rangle\}\leq 2\sup_{\boldsymbol{a}\in\mathcal{A}}|a_k|$$

即

$$\sup_{\boldsymbol{a}\in\mathcal{A}}\langle\boldsymbol{a},\boldsymbol{\epsilon}\rangle-\sup_{\boldsymbol{a}'\in\mathcal{A}}\langle\boldsymbol{a}',\boldsymbol{\epsilon}^{\backslash k}\rangle\leq 2\sup_{\boldsymbol{a}\in\mathcal{A}}|a_k|$$

所以

$$f(\epsilon_1,\ldots,\epsilon_k,\ldots\epsilon_n)-f(\epsilon_1,\ldots,\epsilon_k',\ldots,\epsilon_n)\leq 2\sup_{\boldsymbol{a}\in\mathcal{A}}|a_k|$$

同理得到

$$f(\epsilon_1,\ldots,\epsilon_k',\ldots\epsilon_n)-f(\epsilon_1,\ldots,\epsilon_k,\ldots,\epsilon_n)\leq 2\sup_{\boldsymbol{a}\in\mathcal{A}}|a_k|$$

f 满足以 $\left(2\sup_{\boldsymbol{a}\in\mathcal{A}}|a_k|\right)_{k\in[n]}$ 为参数的有界差不等式.

$$\mathbb{P}\left[|Z-\mathbb{E}Z|\geq t\right]\leq 2e^{-\frac{t^2}{2\sum_{k=1}^{n}(a_k^*)^2}}\quad\forall t\geq 0,$$

其中 $a_k^*=\sup_{\boldsymbol{a}\in\mathcal{A}}|a_k|$.

2 高斯变量的 Lipschitz 函数

引理 2.1

设 $f:\mathbb{R}^n\to\mathbb{R}$ 可微, 对于任意凸函数 $\phi:\mathbb{R}\to\mathbb{R}$, 有

$$\mathbb{E}\left[\phi\left(f(X)-\mathbb{E}[f(X)]\right)\right]\le\mathbb{E}\left[\phi\left(\frac{\pi}{2}\langle\nabla f(X),Y\rangle\right)\right],$$

其中 $X,Y\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_n)$, 且 X 与 Y 独立.

定理 2.1

设 $f:\mathbb{R}^n\to\mathbb{R}$ 可微且 *L-Lipschitz* 连续, $X=(X_1,X_2,\dots,X_n)$ 为随机向量, 其中元素相互独立且服从标准高斯 *(*正态*)* 分布, 则 $f(X)-\mathbb{E}[f(X)]$ 为以 $\frac{\pi L}{2}$ 为参数的 *sub-Gaussian* 变量, 故有

$$\mathbb{P}\left[|f(X)-\mathbb{E}[f(X)]|\ge t\right]\le 2e^{-\frac{2t^2}{\pi^2L^2}}\quad\forall t\ge 0.$$

证明. 对任意 $\lambda\in\mathbb{R}$, 定义 $\phi(\cdot)=\exp(\lambda\cdot)$, ϕ 为凸函数, 由引理 2.1,

$$\mathbb{E}\left[\exp\left(\lambda(f(X)-\mathbb{E}[f(X)])\right)\right]\le\mathbb{E}_{X,Y}\left[\exp\left(\langle\frac{\lambda\pi}{2}\nabla f(X),Y\rangle\right)\right].$$

因为 X 与 Y 相互独立, 且 $Y\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_n)$,

$$\begin{aligned}\mathbb{E}_{X,Y}\left[\exp\left(\langle\frac{\lambda\pi}{2}\nabla f(X),Y\rangle\right)\right]&=\mathbb{E}_X\mathbb{E}_Y\left[\exp\left(\langle\frac{\lambda\pi}{2}\nabla f(X),Y\rangle\right)\right]\\&=\mathbb{E}_X\left[\frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n}\exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}+\langle\frac{\lambda\pi}{2}\nabla f(X),y\rangle\right)\mathrm{d}\boldsymbol{y}\right]\\&=\mathbb{E}_X\left[\exp\left(\frac{\lambda^2\pi^2}{8}\|\nabla f(X)\|^2\right)\right]\\&\le\exp\left(\frac{\lambda^2\pi^2}{8}L^2\right)\qquad(L-\text{Lipschitz})\end{aligned}$$

故

$$\mathbb{E}\left[\exp\left(\lambda(f(X)-\mathbb{E}[f(X)])\right)\right]\le\exp\left(\frac{\lambda^2\pi^2}{8}L^2\right)$$

即 $f(X)-\mathbb{E}[f(X)]$ 是以 $\frac{\pi L}{2}$ 为参数的 sub-Gaussian 变量. □

例 2.1. **χ^2 集中不等式.** 设随机向量 $Z=(Z_k)_{k\in[n]}$ 中元素为相互独立的标准正态分布随机变量, 则 $Y=\|Z\|^2=\sum_{k=1}^{n}Z_k^2$ 服从自由度是 n 的 χ^2 分布.

定义函数 $f:\boldsymbol{v}\in\mathbb{R}^n\mapsto\frac{\|\boldsymbol{v}\|}{\sqrt{n}}$,$f$ 为 $\frac{1}{\sqrt{n}}$-Lipshitz 连续函数, 由定理 2.1

$$\mathbb{P}\left[\frac{\sqrt{Y}}{\sqrt{n}}-\frac{1}{\sqrt{n}}\mathbb{E}[\sqrt{Y}]\ge t\right]\le e^{-\frac{2nt^2}{\pi^2}}\quad\forall t\ge 0.$$

因为 $\frac{1}{\sqrt{n}}\mathbb{E}[\sqrt{Y}] \le \frac{1}{\sqrt{n}}\sqrt{\mathbb{E}[Y]} = 1$, 故有

$$\mathbb{P}\left[\frac{\sqrt{Y}}{\sqrt{n}} - 1 \ge t\right] \le \mathbb{P}\left[\frac{\sqrt{Y}}{\sqrt{n}} - \frac{1}{\sqrt{n}}\mathbb{E}[\sqrt{Y}] \ge t\right] \le e^{-\frac{2nt^2}{\pi^2}} \quad \forall t \ge 0.$$

因此

$$\mathbb{P}\left[Y \ge n(1+t)^2\right] \le e^{-\frac{2nt^2}{\pi^2}} \quad \forall t \ge 0.$$

因为 $(1+t)^2 = 1 + 2t + t^2 \le 1 + 3t, t \in (0,1)$

$$\mathbb{P}\left[Y \ge n(1+3t)\right] \le \mathbb{P}\left[Y \ge n(1+t)^2\right] \quad \forall t \in (0,1).$$

令 $\delta = 3t$, 可整理得到

$$\mathbb{P}\left[Y \ge n(1+\delta)\right] \le e^{-\frac{2n\delta^2}{9\pi^2}} \le e^{-\frac{n\delta^2}{72}} \quad \forall \delta \in (0,3).$$

例 2.2. Gaussian 复杂度. 设随机向量 $\boldsymbol{w} = (w_i)_{i\in[n]}$ 中元素相互独立且服从标准高斯分布, 对于集合 $\mathcal{A} \subset \mathbb{R}^n$, 定义随机变量

$$Z := \sup_{\boldsymbol{a}\in\mathcal{A}} \langle \boldsymbol{a}, \boldsymbol{w} \rangle$$

称 $\mathcal{G}(Z) := \mathbb{E}Z$ 为集合 $\mathcal{A}$ 的 Gaussian 复杂度.

定义函数 $f : \boldsymbol{w} \mapsto \sup_{\boldsymbol{a}\in\mathcal{A}} \langle \boldsymbol{a}, \boldsymbol{w} \rangle$. 由 Cauchy-Schwartz 不等式

$$\sup_{\boldsymbol{a}\in\mathcal{A}} \langle \boldsymbol{a}, \boldsymbol{w} - \boldsymbol{w}' \rangle \le (\sup_{\boldsymbol{a}\in\mathcal{A}} \|\boldsymbol{a}\|) \cdot \|\boldsymbol{w} - \boldsymbol{w}'\|.$$

又因为

$$f(\boldsymbol{w}) - f(\boldsymbol{w}') = \sup_{\boldsymbol{a}\in\mathcal{A}} \langle \boldsymbol{a}, \boldsymbol{w} \rangle - \sup_{\boldsymbol{a}'\in\mathcal{A}} \langle \boldsymbol{a}', \boldsymbol{w}' \rangle \le \sup_{\boldsymbol{a}\in\mathcal{A}} \langle \boldsymbol{a}, \boldsymbol{w} - \boldsymbol{w}' \rangle$$

故

$$f(\boldsymbol{w}) - f(\boldsymbol{w}') \le (\sup_{\boldsymbol{a}\in\mathcal{A}} \|\boldsymbol{a}\|) \cdot \|\boldsymbol{w} - \boldsymbol{w}'\|$$

同理

$$f(\boldsymbol{w}) - f(\boldsymbol{w}) \le (\sup_{\boldsymbol{a}\in\mathcal{A}} \|\boldsymbol{a}\|) \cdot \|\boldsymbol{w} - \boldsymbol{w}'\|$$

记 $D(\mathcal{A}) = \sup_{\boldsymbol{a}\in\mathcal{A}} \|\boldsymbol{a}\|$, 函数 f 为 $D(\mathcal{A})$-Lipschitz 连续. 由定理 2.1

$$\mathbb{P}\left[|Z - \mathbb{E}Z| \ge t\right] \le 2e^{-\frac{2t^2}{\pi^2 D^2(\mathcal{A})}} \quad \forall t \ge 0.$$

例 2.3. Gaussian 混沌. 设 $\boldsymbol{w} \in \mathbb{R}^n$ 和 $\tilde{\boldsymbol{w}} \in \mathbb{R}^n$ 为相互独立的标准 Gaussian 变量. 给定对称矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, 定义随机变量

$$Z = \boldsymbol{w}^T \boldsymbol{A} \tilde{\boldsymbol{w}}$$

称 Z 为 Gaussian 混沌.

容易知道

- $\mathbb{E}[Z]=0$
- 在给定 $\tilde{\boldsymbol{w}}$ 条件下, Z 为均值为 0, 方差为 $\|\boldsymbol{A}\tilde{\boldsymbol{w}}\|^2$ 的 Gaussian 随机变量, 因此

$$\mathbb{P}[|Z|\geq t|\tilde{\boldsymbol{w}}]\leq \exp\left(-\frac{t^2}{2\|\boldsymbol{A}\tilde{\boldsymbol{w}}\|^2}\right)\quad \forall t\geq 0.$$

定义随机变量 $Y:=\|\boldsymbol{A}\tilde{\boldsymbol{w}}\|$, 记 $Y=f(\tilde{\boldsymbol{w}})$, 则函数是 $\|\boldsymbol{A}\|_2$-Lipschitz 连续. 由定理 2.1 有

$$\mathbb{P}\left[Y-\mathbb{E}[Y]\geq\delta\right]\leq \exp\left(-\frac{2\delta^2}{\pi^2\|\boldsymbol{A}\|_2^2}\right)\quad \forall\delta\geq 0.$$

注意到 $\mathbb{E}[Y^2]=\mathbb{E}\left[\tilde{\boldsymbol{w}}^T\boldsymbol{A}^T\boldsymbol{A}\tilde{\boldsymbol{w}}\right]=\operatorname{trace}(\boldsymbol{A}^T\boldsymbol{A})=\|\boldsymbol{A}\|_F^2$, 由 Jensen 不等式

$$\mathbb{E}[Y]=\mathbb{E}[\sqrt{Y^2}]\leq\sqrt{\mathbb{E}[Y^2]}=\|\boldsymbol{A}\|_F$$

故

$$\mathbb{P}\left[Y-\|\boldsymbol{A}\|_F\geq\delta\right]\leq\mathbb{P}\left[Y-\mathbb{E}[Y]\geq\delta\right]\leq \exp\left(-\frac{2\delta^2}{\pi^2\|\boldsymbol{A}\|_2^2}\right)\quad \forall\delta\geq 0.$$

$$\mathbb{P}\left[Y^2\geq(\delta+\|\boldsymbol{A}\|_F)^2\right]\leq \exp\left(-\frac{2\delta^2}{\pi^2\|\boldsymbol{A}\|_2^2}\right)\quad \forall\delta\geq 0.$$

因为 $(\delta+\|\boldsymbol{A}\|_F)^2\leq 2\delta^2+2\|\boldsymbol{A}\|_F^2$,

$$\mathbb{P}\left[Y^2\geq 2\delta^2+2\|\boldsymbol{A}\|_F^2\right]\leq\mathbb{P}\left[Y^2\geq(\delta+\|\boldsymbol{A}\|_F)^2\right]\leq \exp\left(-\frac{2\delta^2}{\pi^2\|\boldsymbol{A}\|_2^2}\right)\quad \forall\delta\geq 0.$$

令 $\delta^2=t\|\boldsymbol{A}\|_2$,

$$\mathbb{P}\left[Y^2\geq 2t\|\boldsymbol{A}\|_2+2\|\boldsymbol{A}\|_F^2\right]\leq \exp\left(-\frac{2t}{\pi^2\|\boldsymbol{A}\|_2}\right)\quad \forall t\geq 0.$$

综上

$$\begin{aligned}\mathbb{P}[|Z|\geq t]&\leq 2\exp\left(-\frac{t^2}{4t\|\boldsymbol{A}\|_2+4\|\boldsymbol{A}\|_F^2}\right)+\exp\left(-\frac{2t}{\pi^2\|\boldsymbol{A}\|_2}\right)\\&\leq 4\exp\left(-\frac{t^2}{8t\|\boldsymbol{A}\|_2+8\|\boldsymbol{A}\|_F^2}\right)\quad \forall t\geq 0.\end{aligned}$$