第十讲

王伟文　暨南大学

1 sub-Gaussian 过程

定义 1.1

一组随机变量 $\{X_\theta : \theta \in \mathcal{T}\}$, 其中 $\mathbb{E}[X_\theta] = 0$, $\forall \theta \in \mathcal{T}$, 若满足

$$\mathbb{E}[e^{\lambda(X_\theta - X_{\bar{\theta}})}] \le e^{\lambda^2\rho^2(\theta,\bar{\theta})/2} \quad \forall \theta, \bar{\theta} \in \mathcal{T}, \forall \lambda \in \mathbb{R},$$

其中 ρ 为定义在 $\mathcal{T}$ 上的度量, 则称 $\{X_\theta : \theta \in \mathcal{T}\}$ 为 sub-Gaussian 过程.

由上述定义, 给定 θ 及 $\bar{\theta}$, $(X_\theta - X_{\bar{\theta}})$ 是均值为 0、以 $\rho(\theta, \bar{\theta})$ 为参数的 sub-Gaussian 随机变量, 故

$$\mathbb{P}[|X_\theta - X_{\bar{\theta}}| \ge \delta] \le 2e^{-\frac{\delta^2}{2\rho^2(\theta,\bar{\theta})}} \quad \forall \delta \ge 0.$$

2 单步离散化上界

定理 2.1

设 $\{X_\theta : \theta \in \mathcal{T}\}$ 是均值为 0、在度量 ρ 下的 sub-Gaussian 过程. 记 $D := \sup\limits_{\theta,\bar{\theta}\in\mathcal{T}} \rho(\theta, \bar{\theta})$. 对任意 $\delta \in [0, D]$ 使得 $\mathcal{N}(\delta; \mathcal{T}, \rho) \ge 0$, 有

$$\mathbb{E}\left[\sup_{\theta,\bar{\theta}\in\mathcal{T}} (X_\theta - X_{\bar{\theta}})\right] \le 2\left[\sup_{\substack{\gamma,\gamma'\in\mathcal{T}\\ \rho(\gamma,\gamma')\le\delta}} (X_\gamma - X_{\gamma'})\right] + 4\sqrt{D^2 \log \mathcal{N}(\delta; \mathcal{T}, \rho)}$$

成立.

定理 2.1 提供了 $\mathbb{E}[\sup\limits_{\theta\in\mathcal{T}} X_\theta]$ 的上界:

$$\mathbb{E}[\sup_{\theta\in\mathcal{T}} X_\theta] = \mathbb{E}[\sup_{\theta\in\mathcal{T}}(X_\theta - \mathbb{E}[X_{\theta_0}])] = \mathbb{E}[\sup_{\theta\in\mathcal{T}}(X_\theta - X_{\theta_0})] \le \mathbb{E}\left[\sup_{\theta,\bar{\theta}\in\mathcal{T}} (X_\theta - X_{\bar{\theta}})\right]$$

例 2.1. 考虑单位球 $\mathcal{B}_2 = \{\theta \in \mathbb{R}^d : \|\theta\|_2 \le 1\}$, 已知其 δ-覆盖数 $\mathcal{N}(\delta; \mathcal{B}_2, \|\cdot\|_2) \le d\log(1 + \frac{2}{\delta})$, $D := \sup\limits_{\theta,\bar{\theta}\in\mathcal{B}_2} \rho(\theta, \bar{\theta}) = 2$.

定义随机变量 $X_\theta = \langle \theta, \boldsymbol{w} \rangle$, 其中 $w_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, 则

$$\begin{aligned}
\sup_{\substack{\gamma,\gamma' \in \mathcal{B}_2 \\ \|\gamma-\gamma'\|_2 \le \delta}} (X_\gamma - X_{\gamma'}) &= \sup_{\substack{\gamma,\gamma' \in \mathcal{B}_2 \\ \|\gamma-\gamma'\|_2 \le \delta}} \langle \gamma - \gamma', \boldsymbol{w} \rangle \\
&\le \sup_{\substack{\gamma,\gamma' \in \mathcal{B}_2 \\ \|\gamma-\gamma'\|_2 \le \delta}} \|\boldsymbol{w}\|_2 \|\gamma - \gamma'\|_2 \\
&\le \delta \|\boldsymbol{w}\|_2
\end{aligned}$$

故

$$\mathbb{E}\left[\sup_{\substack{\gamma,\gamma' \in \mathcal{B}_2 \\ \|\gamma-\gamma'\|_2 \le \delta}} (X_\gamma - X_{\gamma'})\right] \le \mathbb{E}\left[\delta \|\boldsymbol{w}\|_2\right] \le \sqrt{d}\delta$$

由定理 2.1, 欧氏单位球的 Gaussian 复杂度

$$\mathscr{G}(\mathcal{B}_2) = \mathbb{E}[\sup_{\theta \in \mathcal{B}_2} X_\theta] \le 2\sqrt{d}\left(\delta + 2\sqrt{2\log(1+\frac{2}{\delta})}\right) \quad \forall \delta \in (0,2].$$

不等式右边关于 $\delta \in (0,2]$ 最小化得到欧氏单位球的 Gaussian 复杂度上界.

3 定理 2.1 证明

证明. 给定 $\delta > 0$, 记 $N = \mathscr{N}(\delta; \mathcal{T}, \rho)$, 设 $\{\theta^1, \cdots, \theta^N\}$ 为 $\mathcal{T}$ 的 δ-覆盖. 故 $\forall \theta \in \mathcal{T}$, $\exists \theta^i$ 使得 $\rho(\theta, \theta^i) \le \delta$.

$$\begin{aligned}
X_\theta - X_{\theta^1} &= X_\theta - X_{\theta^i} + X_{\theta^i} - X_{\theta^1} \\
&\le \sup_{\substack{\gamma,\gamma' \in \mathcal{T} \\ \rho(\gamma,\gamma') \le \delta}} (X_\gamma - X_{\gamma'}) + \max_{i=1,\ldots,N} |X_{\theta^i} - X_{\theta^1}|
\end{aligned}$$

类似地

$$X_{\theta^1} - X_{\bar{\theta}} \le \sup_{\substack{\gamma,\gamma' \in \mathcal{T} \\ \rho(\gamma,\gamma') \le \delta}} (X_\gamma - X_{\gamma'}) + \max_{i=1,\ldots,N} |X_{\theta^i} - X_{\theta^1}|$$

因此

$$X_\theta - X_{\bar{\theta}} \le 2 \sup_{\substack{\gamma,\gamma' \in \mathcal{T} \\ \rho(\gamma,\gamma') \le \delta}} (X_\gamma - X_{\gamma'}) + 2 \max_{i=1,\ldots,N} |X_{\theta^i} - X_{\theta^1}|$$

另一方面, $X_{\theta^i} - X_{\theta^1}$ 是以 $\rho(\theta^i, \theta^1)$ 为参数的、均值为 0 的 sub-Gaussian 变量, 故对任意 $i \in [N]$, $X_{\theta^i} - X_{\theta^1}$ 是以 D 为参数的 sub-Gaussian, 因此

$$\mathbb{E} \max_{i=1,\ldots,N} |X_{\theta^i} - X_{\theta^1}| \le 2\sqrt{D^2 \log N} \quad N \ge 3.$$

□

4 Dudley's 积分不等式

定理 4.1: Dudley's 积分不等式

设 $\{X_\theta : \theta \in \mathcal{T}\}$ 是均值为 0、在度量 ρ 下的 sub-Gaussian 过程，记 $D := \sup_{\theta,\bar{\theta}\in\mathcal{T}} \rho(\theta,\bar{\theta})$，则

$$\mathbb{E}\sup_{\theta\in\mathcal{T}} X_\theta \le 8D\int_0^\infty \sqrt{\log \mathcal{N}(\epsilon;\mathcal{T},\rho)}\mathrm{d}\epsilon$$

证明. 不失一般性假设 $D=1$, $\mathcal{T}$ 为有限集 (无限集不等式右端为无穷), 定义

$$\epsilon = 2^{-k}, \quad k\in\mathbb{Z}.$$

选择 $\mathcal{T}$ 的一个 ϵ_k-覆盖 $\mathcal{T}_k$, 满足

$$\texttt{card}(\mathcal{T}_k) = \mathcal{N}(\epsilon_k;\mathcal{T},\rho).$$

因为 $\mathcal{T}$ 为有限集, 故存在 $\kappa\in\mathbb{Z}$ 使得, 有 $\theta_0\in\mathcal{T}$ 满足

$$T_\kappa = \{\theta_0\};$$

存在 $K\in\mathbb{Z}$, 使得

$$T_K = T.$$

$\forall\theta\in\mathcal{T}$, 定义

$$\pi_k(\theta) = \arg\min_{\tilde{\theta}\in\mathcal{T}_k} \rho(\theta,\tilde{\theta}) \quad \forall k\in\mathbb{Z}.$$

又因为 $\mathcal{T}_k$ 为 $\mathcal{T}$ 的一个 ϵ_k-覆盖, 故

$$\rho(\theta,\pi_k(\theta)) \le \epsilon_k$$

$X_\theta - X_{\theta_0}$ 可展开为有限项的和

$$X_\theta - X_{\theta_0} = (X_{\pi_\kappa(\theta)} - X_{\theta_0}) + (X_{\pi_{\kappa+1}(\theta)} - X_{\pi_\kappa(\theta)}) + \cdots + (X_\theta - X_{\pi_K(\theta)})$$

因为 $\mathcal{T}_K = \mathcal{T}$,

$$\pi_K(\theta) = \arg\min_{\tilde{\theta}\in\mathcal{T}_K} \rho(\theta,\tilde{\theta}) = \theta$$

$\mathcal{T}_\kappa = \{\theta_0\}$,

$$\pi_\kappa(\theta) = \arg\min_{\tilde{\theta}\in\mathcal{T}_K} \rho(\theta,\tilde{\theta}) = \theta_0$$

故 $X_{\pi_\kappa(\theta)} - X_{\theta_0} = 0$, $X_\theta - X_{\pi_K(\theta)} = 0$.

$$X_\theta - X_{\theta_0} = \sum_{k=\kappa+1}^{K} \left(X_{\pi_k(\theta)} - X_{\pi_{k-1}(\theta)}\right)$$

因此

$$\mathbb{E}\sup_{\theta\in\mathcal{T}}(X_\theta - X_{\theta_0}) \le \sum_{k=\kappa+1}^{K}\mathbb{E}\sup_{\theta\in\mathcal{T}}\left(X_{\pi_k(\theta)} - X_{\pi_{k-1}(\theta)}\right)$$

接下来, 需要控制不等式右端求和各项.

因为 $\{X_\theta : \theta\in\mathcal{T}\}$ 为 sub-Gaussian 过程, 故 $X_{\pi_k(\theta)} - X_{\pi_{k-1}(\theta)}$ 是以 $\rho(\pi_k(\theta), \pi_{k-1}(\theta))$ 为参数的 sub-Gaussian 变量.

由三角不等式

$$\begin{aligned}\rho(\pi_k(\theta), \pi_{k-1}(\theta)) &\le \rho(\pi_k(\theta), \theta) + \rho(\theta, \pi_{k-1}(\theta)) \\ &\le \epsilon_k + \epsilon_{k-1} \\ &\le 2\epsilon_{k-1}\end{aligned}$$

另一方面,

$$\mathtt{card}\left(\{X_{\pi_k(\theta)} - X_{\pi_{k-1}(\theta)} : \theta\in\mathcal{T}\}\right) = \mathtt{card}(\mathcal{T}_k)\cdot\mathtt{card}(\mathcal{T}_{k-1}) \le \mathtt{card}(\mathcal{T}_k)^2$$

由 sub-Gaussian 最大值知

$$\mathbb{E}\sup_{\theta\in\mathcal{T}}\left(X_{\pi_k(\theta)} - X_{\pi_{k-1}(\theta)}\right) \le 2\epsilon_{k-1}\sqrt{\log\mathtt{card}(T_k)}$$

将 ϵ_{k-1} 及 $\mathtt{card}(\mathcal{T}_k)$ 的值代入得到

$$\mathbb{E}\sup_{\theta\in\mathcal{T}}(X_\theta - X_{\theta_0}) \le 4\sum_{k=\kappa+1}^{K} 2^{-k}\sqrt{\log\mathscr{N}(2^{-k};\mathcal{T},\rho)}$$

又因为 $2^{-k} = 2\int_{2^{-k-1}}^{2^{-k}}\mathrm{d}\epsilon$; 若 $\epsilon\le 2^{-k}$, $\mathscr{N}(\epsilon;\mathcal{T},\rho)\ge\mathscr{N}(2^{-k};\mathcal{T},\rho)$.

故

$$\begin{aligned}\mathbb{E}\sup_{\theta\in\mathcal{T}}(X_\theta - X_{\theta_0}) &\le 4\sum_{k=\kappa+1}^{K} 2^{-k}\sqrt{\log\mathscr{N}(2^{-k};\mathcal{T},\rho)} \\ &\le 8\sum_{k=\kappa+1}^{K}\int_{2^{-k-1}}^{2^{-k}}\sqrt{\log\mathscr{N}(\epsilon;\mathcal{T},\rho)}\mathrm{d}\epsilon \\ &\le 8\int_0^\infty\sqrt{\log\mathscr{N}(\epsilon;\mathcal{T},\rho)}\mathrm{d}\epsilon\end{aligned}$$

又因为 $\mathbb{E}X_{\theta_0} = 0$,

$$\mathbb{E}\sup_{\theta\in\mathcal{T}} X_\theta = \mathbb{E}\sup_{\theta\in\mathcal{T}}(X_\theta - \mathbb{E}X_{\theta_0}) = \mathbb{E}\sup_{\theta\in\mathcal{T}}(X_\theta - X_{\theta_0}),$$

结论得证. □

例 4.1. 考虑集合 $\mathcal{B}_2^d(1) = \{\theta\in\mathbb{R}^d : \|\theta\|_2\le 1\}$, 计算 $\mathscr{G}(\mathcal{B}_2^d(1))$.

$$\mathscr{G}(\mathcal{B}_2^d(1)) = \mathbb{E}\sup_{\theta\in\mathcal{B}_2^d(1)}\langle\theta, \boldsymbol{w}\rangle = \mathbb{E}\|\boldsymbol{w}\|_2 = \mathbb{E}\sqrt{\sum_{i=1}^d w_i^2} \le \sqrt{\sum_{i=1}^d \mathbb{E}w_i^2} = \sqrt{d}$$

由体积比例定理, $\mathscr{N}(\epsilon;\mathcal{B}_2^d(1),\|\cdot\|_2)\leq(1+\frac{2}{\epsilon})^d\leq(\frac{3}{\epsilon})^d,\forall\epsilon\in(0,1]$.

若 $\epsilon>1,\mathscr{N}(\epsilon;\mathcal{B}_2^d(1),\|\cdot\|_2)=1$, 有 Dudley 积分不等式

$$\mathscr{G}(\mathcal{B}_2^d(1))=\mathbb{E}\sup_{\theta\in\mathcal{B}_2^d(1)}\leq 16\int_0^1\sqrt{d\log\frac{3}{\epsilon}}\mathrm{d}\epsilon\leq C\sqrt{d}.$$