# 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

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

其中 $\rho$ 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 $\theta$ 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

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

## 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 $\rho$ 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 $\mathcal{N}(\delta;\mathcal{T},\rho) \geq 0$, 有

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

成立.

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

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

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

定义随机变量 $X_\theta = \langle \theta, \boldsymbol{w} \rangle$, 其中 $w_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1)$, 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 $N = \mathcal{N}(\delta; \mathcal{T}, \rho)$, 设 $\{\theta^1, \cdots, \theta^N\}$ 为 $\mathcal{T}$ 的 $\delta$-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

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

□