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 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

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

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 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 $\mathcal{N}(\delta; \mathcal{T}, \rho) \geq 0$, 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 δ-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 $w_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, 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

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

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

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

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,

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

由 sub-Gaussian 最大值知

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

将 ϵ_{k-1} 及 $\texttt{card}(\mathcal{T}_k)$ 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 $\mathscr{G}(\mathcal{B}_2^d(1))$.

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

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

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

5 Dudley 积分不等式应用: Lipschitz 连续函数的 Monte-Carlo 法

考虑积分

$$\int_0^1 f(x)\mathrm{d}x = \mathbb{E}_{X \sim \mathtt{Unif}[0,1]}[f(X)] \approx \frac{1}{n}\sum_{i=1}^n f(X_i)$$

定理 5.1

设函数类

$$\mathcal{F} \triangleq \{f : [0,1] \to \mathbb{R} \text{且} f \text{为} L - \text{Lipschitz 连续}\},$$

若 $X, X_1, X_2, \ldots, X_n$ 为取值处于区间 $[0,1]$ 的独立同分布样本, 则存在常数 $C > 0$, 使得

$$\mathbb{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)] \right| \leq \frac{CL}{\sqrt{n}}.$$

应用 Dudley 积分不等式的定理 5.1 证明思路.

- 定义随机过程

$$X_f = \frac{1}{2\sqrt{n}} \sum_{i=1}^n \left(f(X_i) - \mathbb{E}[f(X)]\right), \quad f \in \mathcal{F}.$$

- 证明随机过程 $\{X_f : f \in \mathcal{F}\}$ 是度量 $\|\cdot\|_\infty$ 下的 sub-Gaussian 过程, 其中

$$\|f - g\|_\infty = \sup_{x \in [0,1]} |f(x) - g(x)| \quad \forall f, g \in \mathcal{F}.$$

- 计算度量空间 $(\mathcal{F}, \|\cdot\|_\infty)$ 的度量熵:

$$\log \mathscr{N}(\epsilon; \mathcal{F}, \|\cdot\|_\infty) \quad \forall \epsilon > 0.$$

为简化定理 5.1 的证明过程, 考虑简单函数类

$$\mathcal{F}_{0,1} \triangleq \{f : [0,1] \to \mathbb{R}, f(0) = 0, \text{且} f \text{为} 1 - \text{Lipschitz 连续}\},$$

引理 5.1

随机过程 $\{X_f : f \in \mathcal{F}_{0,1}\}$ 是度量 $\|\cdot\|_\infty$ 下的 sub-Gaussian 过程.

证明. 只需证明 $\forall f, g \in \mathcal{F}_{0,1}$, 随机变量

$$X_{f,i} - X_{g,i} = \frac{1}{2\sqrt{n}}\left(f(X_i) - \mathbb{E}[f(X_i)] - (g(X_i) - \mathbb{E}[g(X_i)])\right)$$

有界.

因为

$$\begin{aligned}|X_{f,i} - X_{g,i}| &= \frac{1}{2\sqrt{n}}\left(f(X_i) - \mathbb{E}[f(X_i)] - (g(X_i) - \mathbb{E}[g(X_i)])\right) \\ &\leq \frac{1}{\sqrt{n}}\|f-g\|_\infty\end{aligned}$$

故 $X_{f,i} - X_{g,i}$ 是以 $\frac{1}{\sqrt{n}}\|f-g\|_\infty$ 为参数的 sub-Gaussian 随机变量.

由 $\{X_i\}$ 相互独立, $\{X_{f,i} - X_{g,i}\}$ 为相互独立的 sub-Gaussian 随机变量. 故

$$\begin{aligned}X_f - X_g &= \frac{1}{2\sqrt{n}}\sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X)]) - \frac{1}{2\sqrt{n}}\sum_{i=1}^n (g(X_i) - \mathbb{E}[g(X)]) \\ &= \sum_{i=1}^n (X_{f,i} - X_{g,i})\end{aligned}$$

是以

$$\sqrt{\sum_{i=1}^n \frac{1}{n}\|f-g\|_\infty^2} = \|f-g\|_\infty$$

为参数的 sub-Gaussian 变量.

因此 $\{X_f : f \in \mathcal{F}\}$ 为度量 $\|\cdot\|_\infty$ 下的 sub-Gaussian 过程.

□

下面引理 5.2 给出度量空间 $(\mathcal{F}_{0,1}, \|\cdot\|_\infty)$ 的 ϵ-覆盖数上界估计.

引理 5.2

存在常数 $c > 0$, 使得

$$\mathcal{N}(\epsilon; \mathcal{F}_{0,1}, \|\cdot\|_\infty)\begin{cases}\leq e^{c/\epsilon}, & \epsilon \in [0,1]; \\ = 1, & \epsilon > 1.\end{cases}$$

证明. 若 $\epsilon > 1$. 取常数函数 $h(x) = 0, \forall x \in (0,1]$, 则

$$\|f-h\|_\infty = \sup_{x\in[0,1]} |f(x)| = \sup_{x\in[0,1]} |f(x) - f(0)| \leq \sup_{x\in[0,1]} |x| = 1 < \epsilon.$$

此时 $\{h\}$ 为 $\mathcal{F}_{0,1}$ 的一个最小 ϵ-覆盖, 故

$$\mathcal{N}(\epsilon; \mathcal{F}_{0,1}, \|\cdot\|_\infty) = 1.$$

若 $0 < \epsilon \leq 1$, 定义函数类

$$
\begin{aligned}
\mathcal{F}_{0,1}^{\epsilon} \triangleq \{f\text{为分片线性函数}: f(0)=0, f(1/\epsilon) \in \{\pm 1/\epsilon, 0\}, \\
f(2/\epsilon) \in f(1/\epsilon) + \{\pm 1/\epsilon, 0\}, \\
\cdots, \\
f(n/\epsilon) \in f((n-1)/\epsilon) + \{\pm 1/\epsilon, 0\}, \\
f(1) \in f(n/\epsilon) + \{\pm 1/\epsilon, 0\}\}
\end{aligned}
$$

其中 $n = \lfloor 1/\epsilon \rfloor$.

有 $\mathcal{F}_{0,1}^{\epsilon} \subseteq \mathcal{F}_{0,1}$, 且 $\mathcal{F}_{0,1}^{\epsilon}$ 为 $\mathcal{F}_{0,1}$ 的一个 ϵ-覆盖. 由 1-Lipschitz 连续性, 对任意 $f \in \mathcal{F}_{0,1}$, 存在 $g \in \mathcal{F}_{0,1}^{\epsilon}$, 使得

$$
|f(x) - g(x)| \leq \epsilon \quad \forall x \in \bigcup_{i \in \{0, n-1\}} [\frac{i}{\epsilon}, \frac{i+1}{\epsilon}] \bigcup [\frac{n}{\epsilon}, 1],
$$

即

$$
\|f - g\|_\infty = \sup_{x \in [0,1]} |f(x) - g(x)| \leq \epsilon.
$$

因此, 存在 $c > 0$ 使得

$$
\mathscr{N}(\epsilon; \mathcal{F}_{0,1}, \|\cdot\|_\infty) \leq \mathtt{card}(\mathcal{F}_{0,1}^{\epsilon}) = 3^{n+1} \leq e^{c/\epsilon}.
$$

□

定理 5.1 **证明**

Step 1 由引理 5.2, 随机过程 $\{X_f : f \in \mathcal{F}_{0,1}\}$ 是以 0 为均值的 sub-Gaussian 过程.

Step 2 由引理 5.2, 存在 $c > 0$

$$
\log \mathscr{N}(\epsilon; \mathcal{F}_{0,1}^{\epsilon}, \|\cdot\|_\infty) \leq \frac{c}{\epsilon} \quad \epsilon \in (0, 1].
$$

Step 3

$$
\begin{aligned}
D := \sup_{f,g \in \mathcal{F}_{0,1}} \|f - g\|_\infty &= \sup_{f,g \in \mathcal{F}_{0,1}} \sup_{x \in [0,1]} |f(x) - g(x)| \\
&\leq \sup_{f,g \in \mathcal{F}_{0,1}} \left(\sup_{x \in [0,1]} |f(x) - f(0)| + \sup_{x \in [0,1]} |g(x) - g(0)| \right) \\
&\leq 2 \cdot 1.
\end{aligned}
$$

应用 Dudley 积分不等式,

$$
\mathbb{E} \sup_{f \in \mathcal{F}_{0,1}} \left[\left| \frac{1}{2\sqrt{n}} \sum_{i=1}^{n} (f(X_i) - \mathbb{E}[f(X)]) \right| \right] \leq C_1 \cdot 1 \int_0^1 \sqrt{c/\epsilon}\, \mathtt{d}\epsilon \triangleq \frac{C \cdot 1.}{2}
$$

上式可改写为

$$
\mathbb{E} \sup_{f \in \mathcal{F}_{0,1}} \left[\left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)] \right| \right] \leq \frac{C \cdot 1}{\sqrt{n}}
$$

6 Monte-Carlo 法在 d-维空间推广

考虑函数类

$$\mathcal{F}^d \triangleq \{f : [0,1]^d \to \mathbb{R}\text{且}f\text{为}L - \text{Lipschitz 连续}\},$$

引理 6.1

存在常数 $c > 0$, 使得

$$\mathcal{N}(\epsilon; \mathcal{F}^d, \|\cdot\|_\infty) \begin{cases} \leq e^{(c/\epsilon)^d}, & \epsilon \in [0,1]; \\ = 1, & \epsilon > 1. \end{cases}$$

若 $d \geq 2$, 奇异积分

$$\int_0^1 \sqrt{(c/\epsilon)^d} = \infty,$$

此时 Dudley 积分不等式不再适用.

分析 Dudley 积分不等式的证明过程, 可通过设定一个稍微大一点的 $\delta > 0$ 作为 Chaining 的分解点, 避免出现较小的 ϵ, 即, 取 $f_0 \in \mathcal{F}^d$

$$\begin{aligned} \mathbb{E} \sup_{f\in\mathcal{F}^d} |X_f - X_{f_0}| &\lesssim L \int_\delta^\infty \sqrt{\log \mathcal{N}(\epsilon; \mathcal{F}^d, \|\cdot\|_\infty)} + \mathbb{E} \sup_{\substack{f,g\in\mathcal{F} \\ \|f-g\|_\infty \leq \delta}} |X_f - X_g| \\ &\lesssim Lc^d \delta^{-d/2+1} + \sqrt{1}\sqrt{n}\delta \end{aligned}$$

其中"$\lesssim$" 表示省略无关常数.

故

$$\begin{aligned} \mathbb{E} \sup_{f\in\mathcal{F}^d} |2\sqrt{n}X_f| &\leq \mathbb{E} \sup_{f\in\mathcal{F}^d} |2\sqrt{n}(X_f - \mathbb{E}[X_{f_0}])| \\ &\leq 2\sqrt{n}\mathbb{E} \sup_{f\in\mathcal{F}^d} |X_f - X_{f_0}| \\ &\lesssim Lc^d \delta^{-d/2+1} \frac{1}{\sqrt{n}} + \delta \end{aligned}$$

不等式右端关于 δ 最小化, 最优 $\delta^* = \frac{c^2 L^{2/d}}{n^{1/d}}(d/2-1)^{1/d}$, 代入原式得到

$$\mathbb{E} \sup_{f\in\mathcal{F}^d} |2\sqrt{n}X_f| \lesssim \left(\frac{L^2}{n}\right)^{1/d}$$

因此, 存在 $C > 0$, 使得

$$\mathbb{E} \sup_{f\in\mathcal{F}_{0,1}} \left[\left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right|\right] \leq C \left(\frac{L^2}{n}\right)^{1/d}$$

令 $C\left(\frac{L^2}{n}\right)^{1/d} \leq \Delta$, 解得

$$n \geq \frac{1}{L^2}\left(\frac{C}{\Delta}\right)^d \quad \forall d \geq 2.$$

这表明要使用 Monte-Carlo 方法要得到“较好”的近似, 样本量随着维度指数增长, 此时即“维数灾难”(Curse of Dimensionality).