第五讲

王伟文　暨南大学

1 Sub-exponential 变量

定义 1.1: Sub-exponential 变量

设随机变量 X 的期望为 $\mathbb{E}X = \mu$, 若存在非负参数 (ν, α) 满足

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\frac{\nu^2\lambda^2}{2}} \quad \forall |\lambda| < \frac{1}{\alpha}.$$

则称 X 为 *sub-exponential* 变量.

评论. Sub-Gaussian 变量一定是 sub-exponential 变量, 反之不成立.

例 设 $Z \sim \mathcal{N}(0,1)$, $X = Z^2$, $\mathbb{E}X = \mathbb{E}Z^2 = 1$.

$$\mathbb{E}\left[e^{\lambda(X-1)}\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{\lambda(z^2-1)}e^{-z^2/2}dz = \frac{e^{-\lambda}}{\sqrt{1-2\lambda}}$$

若 $\lambda \geq \frac{1}{2}$, 矩母函数不存在, 这表明 X 不是 sub-Gaussian 变量.

另一方面, 利用 $\ln(1-x)$ 的泰勒级数展开, 可证明

$$\frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \leq e^{2\lambda^2} = e^{4\lambda^2/2}, \quad \forall\, |\lambda| < \frac{1}{4}.$$

故 X 是以 $(2,4)$ 为参数的 sub-exponential 变量.

引理 1.1

设 $X_i, i = 1, 2, \ldots, n$ 为相互独立的、以 (ν_i, α_i) 为参数的 *sub-exponential* 随机变量, $\mathbb{E}X_i = \mu_i$, 则 $\sum\limits_{i=1}^{n}(X_i - \mu_i)$ 是以 $(\tilde{\nu}, \tilde{\alpha})$ 为参数的 *sub-exponential* 随机变量, 其中 $\tilde{\alpha} = \max\limits_{i=1,\ldots,n} \alpha_i, \tilde{\nu} = \sqrt{\sum\limits_{i=1}^{n}\nu_i^2}$.

证明. 结论直接由定义得到. □

2 Sub-exponential 变量尾界

命题 2.1: Sub-exponential 变量尾界

设 X 是以 (ν, α) 为参数的 sub-exponential 变量, 则

$$\mathbb{P}\left[X-\mu \geq t\right]=\begin{cases} e^{-\frac{t^2}{2\nu^2}}, & 0 \leq t \leq \dfrac{\nu^2}{\alpha}, \\ e^{-\frac{t}{2\alpha}}, & t \geq \dfrac{\nu^2}{\alpha}. \end{cases}$$

证明. 不失一般性假设 $\mu=0$. 由 Chernoff 技巧, $\forall \lambda \in (0, \frac{1}{\alpha})$ 有

$$\mathbb{P}\left[X \geq t\right] \leq e^{-\lambda t} \leq \exp\left(-\lambda t+\frac{\nu^2\lambda^2}{2}\right).$$

记 $h(\lambda)=-\lambda t+\frac{\nu^2\lambda^2}{2}$. 当 $\lambda=\frac{t}{\nu^2}$ 时, $g(\lambda)$ 取得最小值 $g(\frac{t}{\nu^2})=-\frac{t^2}{2\nu^2}$.

故若 $0 \leq \frac{t}{\nu^2} \leq \frac{1}{\alpha}$, 即 $0 \leq t \leq \frac{\nu^2}{\alpha}$,

$$\mathbb{P}\left[X \geq t\right] \leq e^{-\frac{t^2}{2\nu^2}}.$$

若 $\frac{t}{\nu^2} > \frac{1}{\alpha}$, 即 $t > \frac{\nu^2}{\alpha}$,

$$\inf_{\lambda \in (0, \frac{1}{\alpha})} g(\lambda)=g(\frac{1}{\alpha})=-\frac{t}{\alpha}+\frac{\nu^2}{2\alpha} \leq -\frac{t}{\alpha}+\frac{t}{2\alpha}=-\frac{t}{2\alpha}.$$

此时

$$\mathbb{P}\left[X \geq t\right] \leq e^{-\frac{t}{2\alpha}}.$$

$\square$

设 $X_i, i=1,2,\ldots,n$ 为相互独立的、以 (ν_i, α_i) 为参数的 sub-exponential 随机变量, $\mathbb{E}X_i=\mu_i$, 由命题 2.1 有

$$\mathbb{P}\left[\frac{1}{n}\sum(X_i-\mu_i) \geq t\right]=\begin{cases} e^{-\frac{nt^2}{2(\tilde{\nu}^2/n)}}, & 0 \leq t \leq \dfrac{\nu^2}{n\tilde{\alpha}}, \\ e^{-\frac{t}{2\tilde{\alpha}}}, & t \geq \dfrac{\nu^2}{\tilde{\alpha}}. \end{cases}$$

3 Bernstein 型尾界

定义 3.1: Bernstein 条件

给定随机变量 X, 期望 $\mathbb{E}X = \mu$, 方差 $\sigma^2 = \mathbb{E}X^2 - \mu^2$, 称 X 以 $b(>0)$ 为参数满足 *Bernstein* 条件, 若

$$\left|\mathbb{E}\left[(X-\mu)^k\right]\right| \le \frac{1}{2}k!\sigma^2 b^{k-2}$$

关于 $k = 2, 3, 4, \ldots$ 成立.

命题 3.1

若 X 满足以 b 为参数的 *Bernstein* 条件, 则 X 是以 $(\sqrt{2}\sigma, 2b)$ 为参数的 *sub-exponential* 变量.

证明.

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda(X-\mu)}\right] &= \mathbb{E}\left[\sum_{k=0}^{\infty}\frac{\lambda^k (X-\mu)^k}{k!}\right] \\
&= 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty}\frac{\lambda^k \mathbb{E}(X-\mu)^k}{k!} \\
&\overset{(i)}{\le} 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty}\frac{\lambda^k\sigma^2 b^{k-2}}{2} \\
&\le 1 + \frac{\lambda^2\sigma^2}{2}\left(1 + \sum_{k=3}^{\infty}(|\lambda| b)^{k-2}\right) \\
&\overset{(ii)}{=} 1 + \frac{\lambda^2\sigma^2}{2}\cdot\frac{1}{1-|\lambda| b}
\end{aligned}$$

其中利用 Bernstein 条件, 有不等式 (i) 成立; 若 $|\lambda|b < 1$ 即 $|\lambda| < \frac{1}{b}$, 由 $\frac{1}{1-x}$ 的泰勒级数展开得到等式 (ii).

又因为 $1 + x \le e^x$ 对任意 $x \in \mathbb{R}$ 成立, 故

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda(X-\mu)}\right] &\le 1 + \frac{\lambda^2\sigma^2}{2}\cdot\frac{1}{1-|\lambda| b} \le \exp\left(\frac{\lambda^2\sigma^2}{2-2|\lambda| b}\right) \\
&\overset{(iii)}{\le} \exp\left(\lambda^2(\sqrt{2}\sigma)^2/2\right).
\end{aligned}$$

其中不等式 (iii) 成立若 $|\lambda| \le \frac{1}{2b}$. □

评论. 上述证明过程表明, 命题 3.1 中的 sub-exponential 参数不唯一.

由命题 3.1 结合 Chernoff 技巧, 有如下结论.

定理 3.1: Bernstein 型尾界

对任意满足以 b 为参数的 *Bernstein* 条件的随机变量 X, 有

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq \exp\left(\frac{\lambda^2\sigma^2}{2-2|\lambda|b}\right) \quad \forall|\lambda| < \frac{1}{b}$$

且满足集中不等式

$$\mathbb{P}\left[|X-\mu| \geq t\right] \leq e^{-\frac{t^2}{2(\sigma^2+bt)}} \quad \forall t \geq 0.$$

证明. 1. 定理中的第一个不等式由命题 3.1 的证明过程得到.

2. 从 Chernoff 技巧出发, 由第一个不等式取 $\lambda = \frac{t}{bt+\sigma^2} \in [0, \frac{1}{b})$ 得证. □

评论. *对于有界随机变量即满足 $|X-\mu| \leq b$, 可以证明 $X-\mu$ 是以 b 为参数的 sub-Gaussian 随机变量, 从而得到 Hoeffding 型尾界. 另一方面, X 满足 Bernstein 条件, 从而有 Bernstein 型尾界. 由有界性, $\sigma^2 = \mathbb{E}(X-\mu)^2 \leq b$. 若 t 很小, 在忽略 bt 影响的情形下, Bernstein 型尾界是比 Hoeffding 型尾界更好的估计, 尤其是在 $\sigma^2 \ll b^2$ 的情形下.*

4 应用

χ^2 **变量.** 设 $Z_k \sim \mathcal{N}(0,1)$, 则 $Y = \sum\limits_{k=1}^{n} Z_k^2$ 为自由度是 n 的 χ^2 随机变量. 第一节中的例子表明 Z_k^2 是以 $(2,4)$ 为参数的 sub-exponential 随机变量, Y 是以 $(2\sqrt{n},4)$ 为参数的 sub-exponential 随机变量. 由前述内容可知

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{k=1}^{n} Z_k^2 - 1\right| \geq t\right] \leq 2e^{-nt^2/8}, \quad \forall t \in (0,1).$$

下述 Johnson-Lindenstrauss 引理表明高维空间中随机投影在欧式距离度量下具有近似保距离的性质.

引理 4.1: Johnson-Lindenstrauss 引理

给定 p 维空间中 N 个点 $\{u_i\}_{i=1}^{N} \subset \mathbb{R}^p$. 存在投影 $F: \mathbb{R}^p \to \mathbb{R}^d$, 取 $\delta \in (0,1)$, 当 $d \geq \frac{8}{\delta^2}\log\frac{N^2}{\epsilon}$ 时,

$$\forall (u_i, u_j), i \neq j, (1-\delta)\|u_i - u_j\|^2 \leq \|Fu_i - Fu_j\|^2 \leq (1+\delta)\|u_i - u_j\|^2$$

至少以 $1-\epsilon$ 成立.

证明. 设 $P \in \mathbb{R}^{d\times p}$ 为随机矩阵, 其中每一个元素 $p_{ij} \sim \mathcal{N}(0,1)$ 且相互独立. 记 p_i 表示矩阵 P 第 i 行.

对于任意向量 $v \in \mathbb{R}^p$, $\langle p_i, \frac{v}{\|v\|}\rangle \sim N(0,1)$. 因此

$$Y := \frac{\|Xv\|^2}{\|v\|^2} = \sum_{i=1}^{d} \langle p_i, \frac{v}{\|v\|}\rangle$$

为自由度为 d 的 χ^2 变量.

给定 $u_i, u_j, i \neq j$, 由 χ^2 变量的集中不等式, 有

$$\mathbb{P}\left[\left|\frac{1}{d}\frac{\|Xu_i - Xu_j\|^2}{\|u_i - u_j\|^2} - 1\right| \geq \delta\right] \leq 2e^{-d\delta^2/8}.$$

故

$$\begin{aligned}\mathbb{P}\left[\exists (i,j), i \neq j : \left|\frac{1}{d}\frac{\|Xu_i - Xu_j\|^2}{\|u_i - u_j\|^2} - 1\right| \geq \delta\right] &\leq \sum_{i \neq j} \mathbb{P}\left[\left|\frac{1}{d}\frac{\|Xu_i - Xu_j\|^2}{\|u_i - u_j\|^2} - 1\right| \geq \delta\right] \\ &\leq \binom{N}{2} 2e^{-d\delta^2/8} \leq N^2 e^{-d\delta^2/8}\end{aligned}$$

令 $N^2 e^{-d\delta^2/8} \leq \epsilon$, 解得 $d \geq \frac{8}{\delta^2}\log\frac{N^2}{\epsilon}$, 此时

$$\mathbb{P}\left[\forall (i,j), i \neq j : \left|\frac{1}{d}\frac{\|Xu_i - Xu_j\|^2}{\|u_i - u_j\|^2} - 1\right| \leq \delta\right] \geq 1 - \epsilon.$$

定义随机投影 $F : u \in \mathbb{R}^p \mapsto \frac{Xu}{\sqrt{d}} \in \mathbb{R}^d$, 得证. □