Notes on SPARCs

Yong YANG*

October 7, 2020

* YangYong@bupt.edu.cn

目录

0.1 一些数学上的准备

Lemma 0.1. *Let $u \in \mathbb{R}^N$ and $v \in \mathbb{R}^n$ be deterministic vectors such that $\lim_{n\to\infty} \|u\|^2/n$ and $\lim_{n\to\infty} \|v\|^2/n$ both exist and are finite. Let $\tilde{A} \in \mathbb{R}^{n\times N}$ be a matrix with independent $\mathcal{N}(0,1/n)$ entries. Then*

(a)

$$\tilde{A}u \stackrel{d}{=} \frac{\|u\|}{\sqrt{n}} Z_u \quad \text{and} \quad \tilde{A}^{\mathsf{T}} v \stackrel{d}{=} \frac{\|v\|}{\sqrt{n}} Z_v \tag{1}$$

where $Z_u \in \mathbb{R}^n$ and $Z_v \in \mathbb{R}^N$ are each i.i.d. standard Normal random vectors. Consequently,

$$\lim_{n\to\infty} \frac{\left\|\tilde{A}u\right\|^2}{n} \stackrel{a.s.}{=} \lim_{n\to\infty} \frac{\|u\|^2}{n} \sum_{j=1}^{n} \frac{Z_{u,j}^2}{n} \stackrel{a.s.}{=} \lim_{n\to\infty} \frac{\|u\|^2}{n} \tag{2}$$

$$\lim_{n\to\infty} \frac{\left\|\tilde{A}^{\mathsf{T}}v\right\|^2}{n} \stackrel{a.s.}{=} \lim_{n\to\infty} \frac{\|v\|^2}{n} \sum_{j=1}^{N} \frac{Z_{v,j}^2}{n} \stackrel{a.s.}{=} \lim_{n\to\infty} \frac{\|v\|^2}{n} \tag{3}$$

(b) Let $\mathcal{W}$ be a d-dimensional subspace of $\mathbb{R}^n$ for $d \leqslant n$. Let $(w_1, \cdots, w_d)$ be an orthogonal basis of $\mathcal{W}$ with $\|w_j\|^2 = n$ for $j \in [d]$, and let $\mathsf{P}_{\mathcal{W}}$ denote the orthogonal projection operator onto $\mathcal{W}$. Then for $D = [w_1 \mid \cdots \mid w_d]$, we have $\mathsf{P}_{\mathcal{W}}\tilde{A}u \stackrel{d}{=} \frac{\|u\|}{\sqrt{n}}\mathsf{P}_{\mathcal{W}}Z_u \stackrel{d}{=} \frac{\|u\|}{\sqrt{n}}Dx$ where $x \in \mathbb{R}^n$ is a random vector with i.i.d. $\mathcal{N}(0,1/n)$ entries. Therefore $\lim_{n\to\infty} n^{\delta}\|x\| \stackrel{a.s.}{=} 0$ for any constant $\delta \in [0, 0.5)$. (The limit is taken with d fixed.)

证明. (a) 这个是 SLLN.

□

Lemma 0.2 (Strong Law of Triangular Arrays)**.** 假设 $\{X_{n,j} : j \in [n], n \geqslant 1\}$ 是零均值、行内独立的三角 r.v. 组列, 并且满足

$$\sum_{j=1}^{n} \mathrm{E}|X_{n,j}|^{2+\kappa} \leqslant cn^{1+\kappa/2},\ \exists\kappa \in (0,1),\ c < +\infty. \tag{4}$$

则

$$\frac{1}{n}\sum_{j=1}^{n} X_{n,j} \xrightarrow{a.s.} 0, \quad \text{as } n \to \infty. \tag{5}$$

证明. 我们采用截尾方法来证明它. 令

$$\bar{X}_{n,j} = X_{n,j}\mathbf{1}_{(|X_{n,j}|\leqslant n)},\ Y_{n,j} = \bar{X}_{n,j} - \mathrm{E}\bar{X}_{n,j}, \tag{6}$$

其中 $\mathbf{1}_{(\cdot)}$ 是示性函数. 则 $\{Y_{n,j}\}$ 行内独立, 并且 $\mathrm{E}Y_{n,j} = 0$, $|Y_{n,j}| \leqslant 2n$. 利用 C_r 不等式和 Jensen 不等式可以证明[2]:

$$\mathrm{E}|Y_{n,j}|^r \leqslant 2^r\mathrm{E}\left|\bar{X}_{n,j}\right|^r \leqslant 2^r\mathrm{E}|X_{n,j}|^r, r \geqslant 1.$$

[2]对任意的随机变量 X, $\mathrm{E}|X - \mathrm{E}X|^r \leqslant 2^{r-1}\left(\mathrm{E}|X|^r + |\mathrm{E}X|^r\right) \leqslant 2^r\mathrm{E}|X|^r, r \geqslant 1.$

由 Markov 不等式,

$$
\begin{aligned}
&\sum_{n=1}^{+\infty} P\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_{n,j}\right| \geqslant \varepsilon\right) \leqslant \sum_{n=1}^{+\infty} \frac{1}{n^4\varepsilon^4}\mathrm{E}\left(\sum_{j=1}^{n} Y_{n,j}\right)^4 \\
&= \sum_{n=1}^{+\infty} \frac{1}{n^4\varepsilon^4}\left[\sum_{j=1}^{n}\mathrm{E}Y_{n,j}^4 + \left(\sum_{j=1}^{n}\mathrm{E}Y_{n,j}^2\right)^2\right] \\
&\leqslant \sum_{n=1}^{+\infty} \frac{1}{n^4\varepsilon^4}\left[(2n)^{2-\kappa}\sum_{j=1}^{n}\mathrm{E}|Y_{n,j}|^{2+\kappa} + \left(4\sum_{j=1}^{n}\mathrm{E}X_{n,j}^2\right)^2\right] \\
&\leqslant \sum_{n=1}^{+\infty} \frac{1}{n^4\varepsilon^4}\left[(2n)^{2-\kappa}2^{2+\kappa}\sum_{j=1}^{n}\mathrm{E}|X_{n,j}|^{2+\kappa} + \left(4\sum_{j=1}^{n}\mathrm{E}(1+|X_{n,j}|^{2+\kappa})\right)^2\right] \\
&\leqslant \sum_{n=1}^{+\infty} \frac{16}{n^4\varepsilon^4}\left[n^{2-\kappa}cn^{1+\kappa/2} + (n+cn^{1+\kappa/2})^2\right] \\
&< +\infty.
\end{aligned}
$$

所以

$$
\frac{1}{n}\sum_{j=1}^{n} Y_{n,j} \xrightarrow{\text{a.s.}} 0, \quad (n\to\infty).
$$

另一方面, 利用概率的次可加性可以知道,

$$
\begin{aligned}
&\sum_{n=1}^{+\infty} P\left(\left|\frac{1}{n}\sum_{j=1}^{n} X_{n,j}\mathbf{1}_{(|X_{n,j}|>n)}\right| \geqslant \varepsilon\right) \leqslant \sum_{n=1}^{+\infty} P\left(\frac{1}{n}\sum_{j=1}^{n} X_{n,j}\mathbf{1}_{(|X_{n,j}|>n)} \neq 0\right) \\
&\leqslant \sum_{n=1}^{+\infty} P\left(\bigcup_{j=1}^{n}\{|X_{n,j}|>n\}\right) \leqslant \sum_{n=1}^{+\infty}\sum_{j=1}^{n} P\left(|X_{n,j}|>n\right) \\
&\leqslant \sum_{n=1}^{+\infty}\sum_{j=1}^{n} \frac{\mathrm{E}|X_{n,j}|^{2+\kappa}}{n^{2+\kappa}} \leqslant \sum_{n=1}^{+\infty} \frac{cn^{1+\kappa/2}}{n^{2+\kappa}} \\
&< +\infty.
\end{aligned}
$$

所以又有

$$
\frac{1}{n}\sum_{j=1}^{n} X_{n,j}\mathbf{1}_{(|X_{n,j}|>n)} \xrightarrow{\text{a.s.}} 0, \quad (n\to\infty).
$$

最后,

$$
\begin{aligned}
\left|\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}X_{n,j}\mathbf{1}_{(|X_{n,j}|>n)}\right| &\leqslant \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}|X_{n,j}|\mathbf{1}_{(|X_{n,j}|>n)} \\
&\leqslant \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\frac{|X_{n,j}|^{2+\kappa}}{n^{1+\kappa}} \leqslant \frac{cn^{1+\kappa/2}}{n^{2+\kappa}} \to 0,\ (n\to\infty).
\end{aligned}
$$

综上所述,

$$
\frac{1}{n}\sum_{j=1}^{n} X_{n,j} = \frac{1}{n}\sum_{j=1}^{n} Y_{n,j} + \frac{1}{n}\sum_{j=1}^{n} X_{n,j}\mathbf{1}_{(|X_{n,j}|>n)} - \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}X_{n,j}\mathbf{1}_{(|X_{n,j}|>n)} \xrightarrow{\text{a.s.}} 0.
$$

□

Lemma 0.3. 设 $\{V, V_n; n \geqslant 2\}$ 是独立同分布的随机序列, 并且 $\mathrm{E}V^2 < +\infty$. 对任意的函数 ψ, 如果 $\psi: \mathbb{R} \to \mathbb{R}$ 满足二阶伪 *Lipschitz* 条件, 则

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \psi(V_j) \overset{a.s.}{=} \mathrm{E}\left[\psi(V)\right]. \tag{7}$$

证明. 因为 ψ 满足二阶伪 Lipschitz 条件, 所以

$$\mathrm{E}|\psi(V)| \leqslant C \cdot \mathrm{E}(1 + V^2) < +\infty.$$

注意到 $\psi(V_1), \psi(V_2), \cdots$ 独立同分布, 由 Kolmogorov 强大数定律知结论成立. □

Lemma 0.4 (Stein)**.** 设随机向量 (X, Y) 服从二元正态分布 $\mathcal{N}(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$. 假设函数 $f: \mathbb{R} \to \mathbb{R}$ 绝对连续, 并且 $\mathrm{Cov}(f(X), Y)$ 和 $\mathrm{E}(f'(X))$ 均存在, 则

$$\mathrm{Cov}(f(X), Y) = \mathrm{Cov}(X, Y)\mathrm{E}f'(X). \tag{8}$$

证明. 记标准正态分布的概率密度函数为 $\varphi(\cdot)$, 利用分部积分公式得到

$$\begin{aligned}
\mathrm{E}f'(X) &= \int_{-\infty}^{\infty} f'(x) \frac{1}{\sigma_1} \varphi(\frac{x-\mu_1}{\sigma_1}) \mathrm{d}x \\
&= \left[f(x) \frac{1}{\sigma_1} \varphi(\frac{x-\mu_1}{\sigma_1}) \right] \Bigg|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} f(x) \left[\frac{1}{\sigma_1} \varphi(\frac{x-\mu_1}{\sigma_1}) \right]' \mathrm{d}x \\
&= \frac{1}{\sigma_1^2} \int_{-\infty}^{+\infty} f(x)(x-\mu_1) \frac{1}{\sigma_1} \varphi(\frac{x-\mu_1}{\sigma_1}) \mathrm{d}x = \frac{1}{\sigma_1^2} \mathrm{E}\left[(X-\mu_1) f(X)\right]
\end{aligned}$$

由于 $Y - \dfrac{\rho\sigma_2}{\sigma_1} X$ 与 X 独立, 所以 $Y - \dfrac{\rho\sigma_2}{\sigma_1} X$ 与 $f(X)$ 也独立. 由于独立一定不相关,

$$\mathrm{Cov}\left(f(X), Y - \frac{\rho\sigma_2}{\sigma_1} X \right) = 0.$$

因此,

$$\begin{aligned}
\mathrm{Cov}(f(X), Y) &= \frac{\rho\sigma_2}{\sigma_1} \mathrm{Cov}\left(f(X), X\right) = \frac{\rho\sigma_2}{\sigma_1} \mathrm{E}(X-\mu_1) f(X) \\
&= \rho\sigma_1\sigma_2 \mathrm{E}f'(X) = \mathrm{Cov}(X, Y)\mathrm{E}f'(X).
\end{aligned}$$

□

Lemma 0.5 (正态分布的尾概率估计)**.** 对于 $x > 0$, 有

$$\left(x^{-1} - x^{-3}\right) \varphi(x) \leqslant \int_{x}^{+\infty} \varphi(u) \mathrm{d}u \leqslant x^{-1} \varphi(x), \tag{9}$$

其中 $\varphi(\cdot)$ 是标准正态分布的概率密度函数.

证明. 替换 $u = x + z$, 并且利用 $\exp(-z^2/2) \leqslant 1$, 得到:

$$\int_x^{+\infty} \varphi(u)\mathrm{d}u \leqslant \varphi(x) \int_0^{+\infty} \mathrm{e}^{-xz}\mathrm{d}z = \frac{1}{x}\varphi(x).$$

关于不等式另一边的证明, 我们只需注意到恒等式:

$$\int_x^{+\infty} (1 - 3y^{-4})\mathrm{e}^{-y^2/2}\mathrm{d}y = (x^{-1} - x^{-3})\mathrm{e}^{-x^2/2}.$$

□

Lemma 0.6 (Bernoulli 不等式)**.** 对于自然数 n, 总有

$$\forall x \geqslant -1, (1 + x)^n \geqslant 1 + nx. \tag{10}$$

证明. 利用数学归纳法即可. □

Lemma 0.7. 假设随机变量 $Z_1, Z_2, \cdots$ 独立同分布于标准正态分布, 对任何常数 $K > 2$, 以概率 1, 对充分大的 M 有

$$\max_{j \in [M]} |Z_j| \leqslant \sqrt{2K \ln M}. \tag{11}$$

证明. 记 $\mathcal{Q}(x) = \int_x^{+\infty} \varphi(u)\mathrm{d}u$. 利用 Bernoulli 不等式和引理0.5, 我们有

$$\begin{aligned} P\left(\max_{j \in [M]} |Z_j| > \sqrt{2K \ln M}\right) &= 1 - \left[1 - 2\mathcal{Q}(\sqrt{2K \ln M})\right]^M \\ &\leqslant 2M\mathcal{Q}(\sqrt{2K \ln M}) \\ &\leqslant \frac{2M}{\sqrt{2K \ln M}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-K \ln M} \\ &= \frac{1}{\sqrt{\pi} M^{K-1} \ln^{1/2} M}. \end{aligned}$$

因此, $\sum_{M=1}^{+\infty} P\left(\max_{j \in [M]} |Z_j| > \sqrt{2K \ln M}\right) < +\infty$. 根据 Borel-Cantelli 引理,

$$P\left(\max_{j \in [M]} |Z_j| > \sqrt{2K \ln M},\ i.o.\right) = 0.$$

这说明我们可以以概率 1 保证, $\{\max_{j \in [M]} |Z_j| > \sqrt{2K \ln M}\}$ 至多对有限个 M 发生. □