Math 547: Mathematical Foundations of Statistical Learning Theory Fall 2017

Lecture 40 — November 27

Instructor: Stas Minsker *Scribe: Mose Wintner*

40.1 Matrix Bernstein Continued

Theorem 1. $X_1, \ldots, X_n$ independent, $X_j \in \mathbb{R}^{d\times d}$, $X = X^T$, $\mathbb{E}X_j = 0$, $||X_j|| \le M$, $\sigma^2 = \left|\left|\sum \mathbb{E}X_j^2\right|\right|$. Then

$$P(\left|\left|\sum X_j\right|\right| \ge t) \le \exp\left(\frac{-t^2/2}{\sigma^2 + Mt/3}\right).$$

Proof: (a)

$$P(\lambda_{max}(\sum X_j) \ge t)) \le \inf_{\theta>0} \frac{\mathbb{E}\operatorname{tr} e^{\theta}\sum X_j}{e^{\theta t}}.$$

(b)

$$E\operatorname{tr} e^{\theta\sum X_j} \le d\left|\left|\exp(\sum \log \mathbb{E}e^{\theta X_j}\right|\right|) \le d\exp(\left|\left|\sum \log \mathbb{E}e^{\theta X_j}\right|\right|).$$

(c)

$$\begin{aligned}
\log \mathbb{E}e^{\theta X_j} &\preceq \frac{\mathbb{E}X_1^2}{M^2}(e^{\theta M} - M - 1) \\
\sum \log \mathbb{E}e^{\theta X_j} &\preceq (\sum \mathbb{E}X_j^2)\frac{e^{\theta M} - \theta M - 1}{M^2} \\
\left|\left|\sum \log \mathbb{E}e^{\theta X_j}\right|\right| &\le \theta^2\sigma^2 g(\theta M).
\end{aligned}$$

Combining (a),(b),(c), we get

$$P(\lambda_{max}(\sum X_j) \ge t) \le \inf_{\theta>0} de^{\theta^2\sigma^2 g(\theta M) - \theta t}.$$

Note that

$$g(\theta M) = \frac{e^{\theta M} - \theta M - 1}{(\theta M)^2} = \frac{1}{\theta^2 M^2}\sum_{j=2}^{\infty}\frac{(\theta M)^j}{j!} \le \frac{1}{\theta^2 M^2}\sum \frac{(\theta M)^{j-2}}{2\cdot 3^{j-2}}9 \le \frac{9/2}{M^2}\frac{1}{1-\theta M/3}$$

whenever $\theta \in (0, 3/M)$...

Example 1. Covariance estimation. $Y_1, \ldots, Y_n$ iid copies of $Y \in \mathbb{R}^d$, $\mathbb{E}Y = 0$, $||Y||_2 \le M$ almost surely, and $\Sigma = YY^T$ the covariance matrix. Define $\Sigma_n = \frac{1}{n}\sum Y_jY_j^T$. Let $X_j = \frac{Y_jY_j^T - \Sigma}{n}$. Then $\mathbb{E}X_j = 0$, $||X_j|| \le \frac{2M^2}{n}$, $\sigma^2 = n\,||EX_1^2||$.

$$\mathbb{E}X_j^2 = \frac{1}{n^2}\left|\left|\mathbb{E}(YY^T - \Sigma)^2\right|\right|..$$

We end up with

$$P(||\Sigma_n - \Sigma|| \ge t) \le 2d\exp\left(\frac{-t^2/2}{\frac{1}{n}M^2\,||\Sigma|| + 6tM^2/n}\right).$$