Application of matrix Bern. ≠ : Covariance estimation

Note Title 2015-11-18

- X: Random vector in $\mathbb{R}^n$

Problem Estimate mean $\mu := \mathbb{E}X$, & covariance $\Sigma := \mathbb{E}XX^T - \mu\mu^T$ from sample $X_1, \dots, X_m$ of iid copies of X.

① **Mean** Set $\mu_m := \frac{1}{m}\sum_{i=1}^m X_i$

sample mean

$$\boxed{\text{LLN} \quad \mu_m \to \mu \text{ as } m \to \infty}$$

Non-asymptotic error:

$$\begin{aligned}
\mathbb{E}\|\mu - \mu_m\|_2^2 &= \frac{1}{m}\mathbb{E}\|X - \mu\|_2^2 \\
&= \frac{1}{m}\left(\mathbb{E}\|X\|_2^2 - \|\mu\|_2^2\right) \\
&= \frac{1}{m}\operatorname{tr}\left(\mathbb{E}XX^T - \mu\mu^T\right) \\
&= \frac{1}{m}\operatorname{tr}(\Sigma) \\
&\le \|\Sigma\| \frac{n}{m}
\end{aligned}$$

Check these

$\Rightarrow$ **Prop** Assume $\Sigma = I$. Then

$$\boxed{\sqrt{\mathbb{E}\|\mu - \mu_n\|_2^2} \le \sqrt{\frac{n}{m}}}$$

"A sample size of $m \gtrsim n$ is sufficient to estimate a random vector in $\mathbb{R}^n$."

② **Covariance matrix**

Assume $\mu = 0$ for simplicity.

Sample covariance matrix:

$$\Sigma_m := \frac{1}{m}\sum_{i=1}^m X_i X_i^T$$

A) Estimation in Frobenius norm:

$$\|\Sigma_m - \Sigma\|_F = \|vec(\Sigma_m) - vec(\Sigma)\|_2$$

vectorize

Proceed as in estimating μ. Now $\Sigma \in \mathbb{R}^{n^2}$.
$\Rightarrow$ Need $m \gtrsim n^2$ to estimate Σ.

B) Estimation in operator norm:

We bound $\|\Sigma_m - \Sigma\| = \|\frac{1}{m}\sum_{i=1}^{m} X_i X_i^T - \Sigma\| = \|\frac{1}{m}\sum_{i=1}^{m} Z_i\|$

where $Z_i = (X_i X_i^T - \Sigma)$.

We use matrix Bernstein $\neq$. Assume, for simplicity that $\Sigma = I$.

i) Bound the range K:

Since $\Sigma = I$, $\mathbb{E}\|X\|_2^2 = tr(I) = n$

$\boxed{\text{Assume } \|X\|_2 \le C\sqrt{n} \text{ a.s.}}$

Then, $\|Z_i\| = \|X_i X_i^T - I\|$

$= \max(\|X_i\|^2 - 1, 1)$ $\}$ check

$\le cn =: K \cdot m$

• Bound variance $\sigma^2 := \|\sum_{i=1}^{n} \mathbb{E} Z_i^2\|$

Compute $\mathbb{E} Z_i^2 = \mathbb{E}(XX^T - I)^2$.

$= \mathbb{E}\left[(XX^T)^2 - I^2\right]$

$= \mathbb{E}\left(XX^T XX^T - I\right)$

$= \mathbb{E}\|X\|_2^2 XX^T - I$

$\|X\|_2^2 \le cn$, $\|X\|_2^2 XX^T \succeq 0$

$$I \preceq \mathbb{E}\|X\|_2^2 XX^T - I \preceq cn\, \mathbb{E}\, XX^T = cnI$$

$$\Rightarrow -mI \preceq \sum_{i=1}^{m} \mathbb{E} Z_i^2 \preceq m \cdot cn\, I$$

$\Rightarrow$ Take $m^2\sigma^2 = cmn$

Apply matrix Bernstein $\neq$ with $K := \frac{cn}{m}$, $\sigma^2 = \frac{cn}{m}$

$$P(\|\Sigma_n - \Sigma\| > t) \le 2n \exp\left(\frac{-t^2 m}{cn(1+t)}\right)$$

$$\le 2n\exp\left(-cm \min\left(\frac{t^2}{n}, \frac{t}{n}\right)\right)$$

$$\Rightarrow \mathbb{E}\|\Sigma_n - \Sigma\| = \int_0^\infty P(\|\Sigma_n - \Sigma\| > t) \le c\sqrt{\frac{n}{m}\log n}$$

(Assuming RHS <1) (Check this)

Thm (Covariance estimation) Let $X_1, \dots X_n$ be iid copies of an isotropic vector $X \in \mathbb{R}^n$. Assume that

$$\|X\|_2 \le c\sqrt{n} \quad \text{a.s.}$$

Then

$$\mathbb{E}\left\|\frac{1}{m}\sum_{i=1}^{m} X_i X_i^T - I\right\| \le c\sqrt{\frac{n}{m}\log n}$$

as long as $c\sqrt{\frac{n}{m}\log n} \le 1$

"If $m \ge c' n \log n$, covariance matrix can be estimated."

"Structured random matrix is well conditioned when $m \ge n \log n$."

Remark: • Very general: Almost no assumption on X.
• Requirement $m > n\log n$ is tight.