О точности оценивания ковариационной матрицы многомерного случайного вектора

Ф. Носков Н. Пучкин М. Рахуба В. Спокойный

Национальный исследовательский университет
“Высшая школа экономики”

Содержание

Оценка ковариационной матрицы: постановка задачи

- $\mathbf{X}, \mathbf{X}_1, \dots, \mathbf{X}_n$ – н. о. р. случайные векторы в $\mathbb{R}^d$
- $\mathbb{E}\,\mathbf{X} = \mathbf{0}$, $\mathbb{E}\,\mathbf{X}\mathbf{X}^\top = \Sigma$
- Размерность пространства d велика и может превосходить n

Цель: оценить ковариационную матрицу Σ по выборке $(\mathbf{X}_1, \dots, \mathbf{X}_n)$

Оценка ковариационной матрицы: субгауссовский случай

Стандартная оценка – выборочная ковариационная матрица

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^\top$$

Точность оценивания: если $\mathbf{X}_1, \dots, \mathbf{X}_n$ – субгауссовские случайные векторы, то

$$\left\| \widehat{\Sigma} - \Sigma \right\| \lesssim \|\Sigma\| \left(\sqrt{\frac{d + \ln(1/\delta)}{n}} \vee \frac{d + \ln(1/\delta)}{n} \right)$$

с вероятностью не менее $(1 - \delta)$ (см., например, [Vershynin, 2012, Corollary 5.50])

Обозначение: $f \lesssim g$ и $g \gtrsim f$ равносильны $f = \mathcal{O}(g)$

Замечание: оценка становится бессодержательной, когда $d > n$

Как преодолеть проклятие размерности I: структура Σ

Первый способ преодоления проклятия размерности – сделать предположения о структуре ковариационной матрицы Σ

- разреженные ковариационные матрицы [Bickel and Levina, 2008a], [Cai and Zhou, 2012], [Fan et al., 2015]
- ленточные ковариационные матрицы [Bickel and Levina, 2008b], [Cai et al., 2010]
- теплицевы ковариационные матрицы [Xiao and Wu, 2012], [Cai et al., 2013]
- ковариационные матрицы со структурой произведения Кронекера [Tsiligkaridis and Hero, 2013], [Leng and Pan, 2018]

Как преодолеть проклятие размерности II: распределение X

Предположение L_2-ψ_2-эквивалентности: существует такое $\varkappa > 0$, что для любого $\mathbf{u} \in \mathbb{R}^d$ выполнено неравенство

$$\|\mathbf{u}^\top \mathbf{X}\|_{\psi_2} \leqslant \varkappa \sqrt{\mathbf{u}^\top \Sigma \mathbf{u}}, \quad \text{где} \quad \|\mathbf{u}^\top \mathbf{X}\|_{\psi_2} = \inf\left\{t > 0 : \mathbb{E} e^{(\mathbf{u}^\top \mathbf{X})^2/t^2} \leqslant 2\right\}$$

Безразмерные оценки: если выполнено условие L_2-ψ_2-эквивалентности, то, каково бы ни было $\delta \in (0, 1)$, с вероятностью не менее $(1 - \delta)$, выполнено

$$\left\|\widehat{\Sigma} - \Sigma\right\| \lesssim \|\Sigma\| \left(\sqrt{\frac{\mathrm{r}(\Sigma) + \ln(1/\delta)}{n}} \vee \frac{\mathrm{r}(\Sigma) + \ln(1/\delta)}{n} \right), \quad \text{где} \quad \mathrm{r}(\Sigma) = \frac{\mathrm{Tr}(\Sigma)}{\|\Sigma\|}$$

(см., например, [Koltchinskii and Lounici, 2017])

Правая часть неравенства не зависит от d!

Основные тенденции

- Структурные предположения отошли на второй план
- Больше внимания уделяется безразмерным оценкам
- Техника доказательств значительно эволюционировала за последние 10 лет
- Результаты помогают лучше понять феномены, возникающие в задаче перепараметризованной многомерной регрессии

Содержание

Обзор литературы

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^\top$$

[Kannan et al., 1997]: если $\mathbf{X}_1, \ldots, \mathbf{X}_n$ распределены равномерно на выпуклом теле $\mathcal{K}$,

$$\mathbb{E} \left\| \widehat{\Sigma} - \Sigma \right\| \lesssim \frac{d}{\sqrt{n}}$$

[Rudelson, 1999]: при тех же предположениях

$$\mathbb{E} \left\| \widehat{\Sigma} - \Sigma \right\| \lesssim \sqrt{\frac{\ln d}{n}} \left(\mathbb{E} \|\mathbf{X}\|^{\ln n} \right)^{1/\ln n}$$

Обзор литературы

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^\top$$

Результат [Kannan et al., 1997] был улучшен в ряде работ:

- [Bourgain, 1999]
- [Giannopoulos et al., 2005]
- [Paouris, 2006]
- [Adamczak et al., 2010]

Обзор литературы

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^\top$$

[Adamczak et al., 2011]: если $\mathbf{X}_1, \dots, \mathbf{X}_n$ имеют лог-вогнутое распределение, то

$$\left\|\widehat{\Sigma} - \Sigma\right\| \lesssim \sqrt{\frac{d}{n}} \quad \text{с вероятностью} \geqslant 1 - e^{-\mathcal{O}(\sqrt{d})}$$

[Vershynin, 2012, Corollary 5.50]: если $\mathbf{X}_1, \dots, \mathbf{X}_n$ – субгауссовские случайные векторы,

$$\left\|\widehat{\Sigma} - \Sigma\right\| \lesssim \|\Sigma\| \left(\sqrt{\frac{d + \ln(1/\delta)}{n}} \vee \frac{d + \ln(1/\delta)}{n}\right) \quad \text{с вероятностью} \geqslant (1 - \delta)$$

[Tropp, 2012, Vershynin, 2012]: если $\|\mathbf{X}_i\| \leqslant R$ почти наверное для всех $1 \leqslant i \leqslant n$, то

$$\left\|\widehat{\Sigma} - \Sigma\right\| \lesssim \sqrt{\frac{R\|\Sigma\| \ln(d/\delta)}{n}} \vee \frac{R \ln(d/\delta)}{n} \quad \text{с вероятностью} \geqslant (1 - \delta)$$

Обзор литературы

Предположение L_2-ψ_2-эквивалентности: существует такое $\varkappa > 0$, что для любого $\mathbf{u} \in \mathbb{R}^d$ выполнено неравенство

$$\|\mathbf{u}^\top \mathbf{X}\|_{\psi_2} \leqslant \varkappa \sqrt{\mathbf{u}^\top \Sigma \mathbf{u}}, \quad \text{где} \quad \|\mathbf{u}^\top \mathbf{X}\|_{\psi_2} = \inf\left\{t > 0 : \mathbb{E} e^{(\mathbf{u}^\top \mathbf{X})^2/t^2} \leqslant 2\right\}$$

Если выполнено условие L_2-ψ_2-эквивалентности, то с вероятностью не менее $(1 - \delta)$

- [Adamczak, 2015], [Koltchinskii and Lounici, 2017]:

$$\left\|\widehat{\Sigma} - \Sigma\right\| \lesssim \|\Sigma\| \left(\sqrt{\frac{\mathrm{r}(\Sigma) + \ln(1/\delta)}{n}} \vee \frac{\mathrm{r}(\Sigma) + \ln(1/\delta)}{n}\right), \quad \text{где} \quad \mathrm{r}(\Sigma) = \frac{\mathrm{Tr}(\Sigma)}{\|\Sigma\|}$$

- [Bunea and Xiao, 2015]:

$$\left\|\widehat{\Sigma} - \Sigma\right\|_{\mathrm{F}} \lesssim \|\Sigma\| \mathrm{r}(\Sigma) \left(\sqrt{\frac{\ln(2/\delta)}{n}} \vee \frac{\ln(2/\delta)}{n}\right), \quad \text{где} \quad \mathrm{r}(\Sigma) = \frac{\mathrm{Tr}(\Sigma)}{\|\Sigma\|}$$

Обзор литературы

Предположение L_2-ψ_2-эквивалентности: существует такое $\varkappa > 0$, что для любого $\mathbf{u} \in \mathbb{R}^d$ выполнено неравенство

$$\|\mathbf{u}^\top \mathbf{X}\|_{\psi_2} \leqslant \varkappa \sqrt{\mathbf{u}^\top \Sigma \mathbf{u}}, \quad \text{где} \quad \|\mathbf{u}^\top \mathbf{X}\|_{\psi_2} = \inf\left\{ t > 0 : \mathbb{E} e^{(\mathbf{u}^\top \mathbf{X})^2/t^2} \leqslant 2 \right\}$$

- [Zhivotovskiy, 2024]: если выполнено условие L_2-ψ_2-эквивалентности,

$$\left\| \widehat{\Sigma} - \Sigma \right\| \leqslant 20 \varkappa^2 \|\Sigma\| \left(\sqrt{\frac{\mathtt{r}(\Sigma) + \ln(1/\delta)}{n}} \vee \frac{\mathtt{r}(\Sigma) + \ln(1/\delta)}{n} \right) \quad \text{с вероятностью} \geqslant (1 - \delta)$$

- [Han, 2022]: если $\mathbf{X}_1, \dots, \mathbf{X}_n \sim \mathcal{N}(\mathbf{0}, \Sigma)$, то

$$\mathbb{E} \left\| \widehat{\Sigma} - \Sigma \right\| \leqslant \|\Sigma\| \left(1 + \mathcal{O}\left(\frac{1}{\sqrt{\mathtt{r}(\Sigma)}} \right) \right) \left(2\sqrt{\frac{\mathtt{r}(\Sigma)}{n}} + \frac{\mathtt{r}(\Sigma)}{n} \right) \quad \text{где} \quad \mathtt{r}(\Sigma) = \frac{\mathrm{Tr}(\Sigma)}{\|\Sigma\|}$$

Предположения

Предположение о распределении X: существует $\omega > 0$, такая что случайный вектор $\xi = \Sigma^{-1/2}\mathbf{X}$ удовлетворяет неравенству

$$\left\| \xi^\top V \xi - \mathrm{Tr}(V) \right\|_{\psi_1} \leqslant \omega \|V\|_{\mathrm{F}} \quad \text{для всех } V \in \mathbb{R}^{d \times d} \tag{1}$$

ψ_s-норма Орлича, $s \geqslant 1$: $\|\eta\|_{\psi_s} = \inf\left\{t > 0 : \mathbb{E} e^{|\eta|^s / t^s} \leqslant 2\right\}$

Примеры: данное предположение выполнено для

- многомерного нормального распределения
- распределений, удовлетворяющих нер-ву Хэнсона-Райта [Hanson and Wright, 1971]
- распределений, удовлетворяющих свойству выпуклой концентрации [Adamczak, 2015]

Предположения

Утверждение

Пусть $\varepsilon_1, \ldots, \varepsilon_d$ – независимые радемахеровские случайные величины, $\eta \sim \mathrm{Be}(1/2)$. Рассмотрим случайный вектор $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)^\top$ с компонентами $\xi_i = \sqrt{2}\varepsilon_i \eta$, $1 \leqslant i \leqslant d$. Тогда:

- $\mathbb{E}\boldsymbol{\xi} = \mathbf{0}$ и $\mathbb{E}\boldsymbol{\xi}\boldsymbol{\xi}^\top = I_d$;
- для любого $\mathbf{u} \in \mathbb{R}^d$ выполнено $\left\|\boldsymbol{\xi}^\top \mathbf{u}\right\|_{\psi_2}^2 \leqslant 16\,\mathbb{E}\|\boldsymbol{\xi}^\top \mathbf{u}\|^2/3$;
- существует матрица $V \in \mathbb{R}^{d \times d}$, такая что

$$\mathbb{E}\left(\boldsymbol{\xi}^\top V \boldsymbol{\xi} - \mathbb{E}\boldsymbol{\xi}^\top V \boldsymbol{\xi}\right)^4 = d^2 \, \|V\|_{\mathrm{F}}^4 .$$

Вывод: предположение (1) более строгое, чем требование $L_2 - \psi_2$-эквивалентности

Основные результаты

Теорема: (П., Носков, Спокойный)

Пусть выполнено предположение (1). Зафиксируем $\delta \in (0, 1)$ и обозначим

$$\mathrm{R}(\Sigma, \delta) = \frac{3\mathtt{r}(\Sigma)}{2\sqrt{2}} + \frac{1}{2}\sqrt{\mathtt{r}(\Sigma^2)} + 2\sqrt{e \log(3/\delta)}$$

Допустим, что объем выборки n удовлетворяет неравенствам

$$6\omega \|\Sigma\| \, \mathrm{R}(\Sigma, \delta) \leqslant \sqrt{n} \quad \text{и} \quad \mathtt{r}(\Sigma)^2 \, \mathrm{R}(\Sigma, \delta)^2 \left(\frac{\mathrm{R}(\Sigma, \delta)^2}{4} + 3 \right) \leqslant 36n$$

Тогда с вероятностью не менее $(1 - \delta)$ выполнено

$$\left\|\widehat{\Sigma} - \Sigma\right\|_{\mathrm{F}}^2 - \mathbb{E}\left\|\widehat{\Sigma} - \Sigma\right\|_{\mathrm{F}}^2 < \frac{4\|\Sigma\|^2}{n} \max\left\{ \sqrt{2\left(64\omega^2 \mathtt{r}(\Sigma^2)^2 + (8\omega)^4 \mathtt{r}(\Sigma^4)\right)\log(3/\delta)},\ 4e\tau \log(3/\delta) \right\}$$

Основные результаты

Теорема: (П., Носков, Спокойный)

Пусть выполнено условие (1). Предположим, что Σ, ω и $\delta \in (0,1)$ удовлетворяют соотношениям

$$\frac{\mathrm{Tr}(\Sigma)^2 \vee \log(8/\delta)}{n\rho_{\max}^2} \to 0, \qquad \frac{\mathrm{r}(\Sigma)^6}{n} \to 0 \quad \text{и} \quad \frac{\mathrm{r}(\Sigma)\log(8/\delta)}{\sqrt{n}} \to 0 \quad \text{при } n \to \infty.$$

Тогда с вероятностью не менее $(1-\delta)$ выполнено

$$\left| \frac{\|\widehat{\Sigma} - \Sigma\|_{\mathrm{F}}^2}{\mathbb{E}\|\widehat{\Sigma} - \Sigma\|_{\mathrm{F}}^2} - 1 \right| \lesssim \tau \max\left\{ \frac{\sqrt{\log(8/\delta)}}{\mathrm{r}(\Sigma) - 1}, \frac{\log(8/\delta)}{\mathrm{r}(\Sigma)\big(\mathrm{r}(\Sigma) - 1\big)} \right\}$$

Примеры

$$\mathbf{X} = \Sigma^{1/2}\xi$$

Пример 1: $\xi_i \sim \mathrm{p}(x),\ \mathrm{p}(x) \propto e^{-|x|}\mathbb{1}(|x| \leqslant 6)$ Пример 2: $\xi \sim \mathcal{U}(\mathcal{S}^{d-1})$

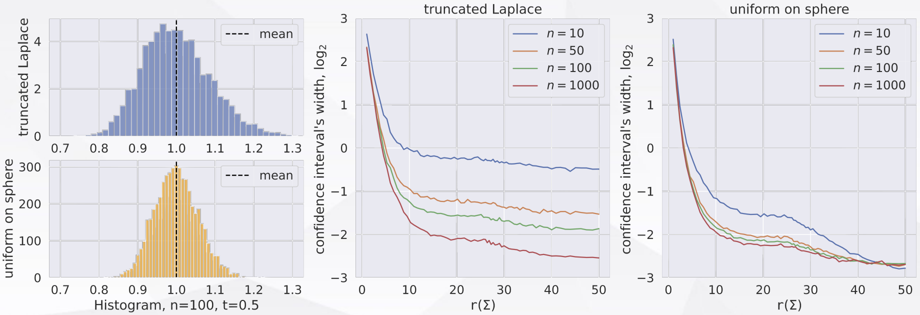

Рис.: Концентрация $\|\widehat{\Sigma} - \Sigma\|_F^2 / \mathbb{E}\|\widehat{\Sigma} - \Sigma\|_F^2$ в окрестности 1 с ростом $\mathrm{r}(\Sigma)$

Пример 3: $\mathbf{X} = \Sigma^{1/2}\xi$, $\xi_i = \sqrt{2}\varepsilon_i\eta$, $\eta \sim \mathrm{Be}(1/2)$, $\varepsilon_1, \ldots, \varepsilon_d$ – независимые радемахеровские случайные величины

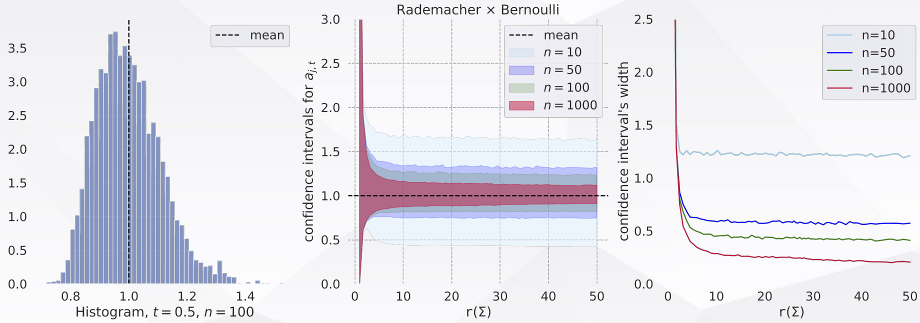

Рис.: Отсутствие концентрации $\|\widehat{\Sigma} - \Sigma\|_{\mathrm{F}}^2 / \mathbb{E}\|\widehat{\Sigma} - \Sigma\|_{\mathrm{F}}^2$ в окрестности 1 с ростом $\mathbf{r}(\Sigma)$

Основные идеи доказательства: шаг 1

Гауссовское распределение – ключ к безразмерным оценкам

Лемма

Зафиксируем произвольные положительные ζ and λ, и пусть Γ – случайная матрица, компоненты которой – независимые стандартные нормальные случайные величины. Тогда, для любой матрицы $A \in \mathbb{R}^{d \times d}$, удовлетворяющей неравенству $\|\Sigma A \Sigma\|_{\mathrm{F}} \leqslant \zeta$, и любого $\beta > 1$, выполнено

$$\mathbb{E}_{\Gamma} \exp\left\{\sqrt{\lambda} \operatorname{Tr}(\Gamma^{\top} \Sigma^{1/2} A \Sigma^{1/2})\right\} \mathbb{1}\left(\|\Sigma^{1/2} \Gamma \Sigma^{1/2}\|_{\mathrm{F}} \leqslant \zeta\sqrt{\lambda} + \sqrt{\beta} \operatorname{Tr}(\Sigma)\right)$$
$$\geqslant \frac{\beta - 1}{\beta} \exp\left\{\lambda \|\Sigma^{1/2} A \Sigma^{1/2}\|_{\mathrm{F}}^2 / 2\right\}$$

Основные идеи доказательства: шаг 1

Взяв $\beta = 2$ и применив лемму к матрице $H = \sqrt{n}\,\Sigma^{-1/2}\left(\widehat{\Sigma} - \Sigma\right)\Sigma^{-1/2}$, получаем

$$
\begin{aligned}
&\mathbb{E} e^{\lambda \|\Sigma^{1/2} H \Sigma^{1/2}\|_{\mathrm{F}}^2/2} \mathbb{1}\left(\|\Sigma H \Sigma\|_{\mathrm{F}} \leqslant \zeta\right) \\
&\leqslant 2\, \mathbb{E} e^{\sqrt{\lambda}\mathrm{Tr}(\Gamma^\top \Sigma^{1/2} H \Sigma^{1/2})} \mathbb{1}\left(\|\Sigma^{1/2}\Gamma\Sigma^{1/2}\|_{\mathrm{F}} \leqslant \zeta\sqrt{\lambda} + \sqrt{2}\,\mathrm{Tr}(\Sigma)\right) \\
&= 2\, \mathbb{E} e^{n\,\phi(\sqrt{\lambda}\Sigma^{1/2}\Gamma\Sigma^{1/2}/\sqrt{n})} \mathbb{1}\left(\|\Sigma^{1/2}\Gamma\Sigma^{1/2}\|_{\mathrm{F}} \leqslant \zeta\sqrt{\lambda} + \sqrt{2}\,\mathrm{Tr}(\Sigma)\right),
\end{aligned}
$$

где

$$
\varphi(U) = \log \mathbb{E} \exp\left\{\xi^\top U \xi\right\}, \quad \xi = \Sigma^{-1/2}\mathbf{X}
$$

Основные идеи доказательства: шаг 2

Лемма

Пусть распределение $\mathbf{X}_1, \dots, \mathbf{X}_n$ удовлетворяет (1). Тогда

$$\left|\left\langle \nabla^3 \varphi(O), U^{\otimes 3} \right\rangle\right| \leqslant (8\omega)^3 \|U\|_{\mathrm{F}}^3,$$
$$\left|\left\langle \nabla^4 \varphi(V), U^{\otimes 4} \right\rangle\right| \leqslant (8\omega)^4 \|U\|_{\mathrm{F}}^4 \quad \text{для всех } \|V\|_{\mathrm{F}} \leqslant 1/(6\omega)$$

Благодаря лемме можем оценить вероятности больших уклонений:

$$\mathbb{P}\left(\|\Sigma^{1/2} H \Sigma^{1/2}\|_{\mathrm{F}} \geqslant \zeta\right) = \sum_{k=1}^{\infty} \mathbb{P}\left(e^{k-1}\zeta \leqslant \|\Sigma^{1/2} H \Sigma^{1/2}\|_{\mathrm{F}} < e^k \zeta\right)$$

$$\leqslant \sum_{k=1}^{\infty} \mathbb{P}\left(\|\Sigma^{1/2} H \Sigma^{1/2}\|_{\mathrm{F}} \geqslant e^{k-1}\zeta \text{ и } \|\Sigma H \Sigma\|_{\mathrm{F}} \leqslant e^k \zeta \|\Sigma\|\right)$$

$$\leqslant 2 \sum_{k=1}^{\infty} e^{-\lambda e^{k-1}\zeta}\, \mathbb{E} e^{n\,\phi(\sqrt{\lambda}\Sigma^{1/2}\Gamma\Sigma^{1/2}/\sqrt{n})} \mathbb{1}\left(\|\Sigma^{1/2}\Gamma\Sigma^{1/2}\|_{\mathrm{F}} \leqslant e^k \zeta \sqrt{\lambda} + \sqrt{2}\,\mathrm{Tr}(\Sigma)\right)$$

Содержание

Напоминание

Два способа преодоления проклятия размерности:

- сделать предположения о структуре ковариационной матрицы Σ ([Bickel and Levina, 2008a], [Cai et al., 2010], [Xiao and Wu, 2012], [Tsiligkaridis and Hero, 2013], [Fan et al., 2015] и др.)
- сделать дополнительные предположения о распределении $\mathbf{X}_1, \dots, \mathbf{X}_n$ ([Adamczak, 2015], [Bunea and Xiao, 2015], [Koltchinskii and Lounici, 2017], [Han, 2022], [Zhivotovskiy, 2024])

Вопрос: можно ли совместить 2 способа преодоления проклятия размерности?

Предположения I

Структурное предположение – модель суммы произведений Кронекера

$$\Sigma = \Phi_1 \otimes \Psi_1 + \ldots + \Phi_K \otimes \Psi_K, \tag{2}$$

где $\Phi_1, \Psi_1, \ldots, \Phi_K, \Psi_K$ – симметричные положительно полуопределенные матрицы, $\Phi_j \in \mathbb{R}^{p \times p}$, $\Psi_j \in \mathbb{R}^{q \times q}$ для всех $j \in \{1, \ldots, K\}$ и $pq = d$

Произведение Кронекера матриц $A \in \mathbb{R}^{p \times q}$ и $B \in \mathbb{R}^{r \times s}$:

$$A \otimes B = \begin{pmatrix} a_{11}B & \ldots & a_{1q}B \\ \vdots & \ddots & \vdots \\ a_{p1}B & \ldots & a_{pq}B \end{pmatrix} \in \mathbb{R}^{pr \times qs}$$

Предположения II

Предположение о распределении X: существует $\omega > 0$, такая что случайный вектор $\boldsymbol{\xi} = \Sigma^{-1/2}\mathbf{X}$ удовлетворяет неравенству

$$\ln \mathbb{E} \exp\left\{\boldsymbol{\xi}^\top V(\Sigma^{-1/2}\boldsymbol{\xi}) - \mathrm{Tr}(V)\right\} \leqslant \omega^2 \|V\|_{\mathrm{F}}^2 \tag{3}$$

при всех $V \in \mathbb{R}^{d \times d}$ с нормой Фробениуса $\|V\|_{\mathrm{F}} \leqslant 1/\omega$

Замечание. Условие эквивалентно

$$\left\|\boldsymbol{\xi}^\top V \boldsymbol{\xi} - \mathrm{Tr}(V)\right\|_{\psi_1} \leqslant \omega' \|V\|_{\mathrm{F}} \quad \text{для всех } V \in \mathbb{R}^{d \times d}$$

Оператор переформировки

Переформировка – стандартный способ работы с произведениями Кронекера

Определение: оператор переформировки

Пусть M – матрица размера $pq \times pq$ с блочной структурой

$$M = \begin{pmatrix} M(1,1) & \dots & M(1,p) \\ \vdots & \ddots & \vdots \\ M(p,1) & \dots & M(p,p) \end{pmatrix}, \quad \text{где } M(i,j) \in \mathbb{R}^{q\times q} \text{ для всех } i,j \in \{1,\dots,p\}.$$

Оператор переформировки $\mathcal{R} : \mathbb{R}^{pq\times pq} \to \mathbb{R}^{p^2\times q^2}$ определен следующим образом: $((j-1)p+i)$-ая строка $\mathcal{R}(M)$ равна $\mathbf{vec}(M(i,j))^\top$

Здесь и далее **vec** – оператор, соединяющий столбцы матрицы в один вектор

Оператор переформировки

Свойства оператора переформировки $\mathcal{R}$:

- $\mathcal{R}$ – линейный оператор
- $\mathcal{R}$ задает изометрию между $(\mathbb{R}^{pq\times pq}, \|\cdot\|_F)$ и $(\mathbb{R}^{p^2\times q^2}, \|\cdot\|_F)$
- существует обратный оператор $\mathcal{R}^{-1}$
- если $M = A \otimes B$, где $A \in \mathbb{R}^{p\times p}$ и $B \in \mathbb{R}^{q\times q}$, то $\mathcal{R}(M) = \mathbf{vec}(A)\mathbf{vec}(B)^\top$.

Следствия:

- $\mathcal{R}(\Sigma)$ – матрица малого ранга
- для любой оценки $\tilde{\Sigma}$ выполнено равенство $\left\|\tilde{\Sigma} - \Sigma\right\|_F = \left\|\mathcal{R}(\tilde{\Sigma}) - \mathcal{R}(\Sigma)\right\|_F$

Оценки пенализованного метода наименьших квадратов:

Ранговая пенализация [Masak et al., 2022]:

$$\check{\Sigma} = \mathcal{R}^{-1}(\check{R}), \quad \text{где} \quad \check{R} \in \operatorname*{argmin}_{R \in \mathbb{R}^{p^2 \times q^2}} \left\{ \left\| R - \mathcal{R}(\widehat{\Sigma}) \right\|_{\mathrm{F}}^2 + \lambda \operatorname{rank}(R) \right\}, \quad \widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^\top$$

Пенализация ядерной нормой [Tsiligkaridis and Hero, 2013]:

$$\widetilde{\Sigma} = \mathcal{R}^{-1}(\widetilde{R}), \quad \text{где} \quad \widetilde{R} \in \operatorname*{argmin}_{R \in \mathbb{R}^{p^2 \times q^2}} \left\{ \left\| R - \mathcal{R}(\widehat{\Sigma}) \right\|_{\mathrm{F}}^2 + \lambda \|R\|_* \right\}, \quad \widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^\top \tag{4}$$

Замечание: обе оценки $\check{R}$ и $\widetilde{R}$ имеют явный вид и могут быть выражены через сингулярные числа и векторы матрицы $\mathcal{R}(\widehat{\Sigma})$

Точность оценивания

[Tsiligkaridis and Hero, 2013]: если $\mathbf{X}_1, \dots, \mathbf{X}_n \sim \mathcal{N}(\mathbf{0}, \Sigma)$ и Σ представима в виде суммы K произведений Кронекера (2), то неравенство

$$\left\| \widetilde{\Sigma} - \Sigma \right\|_{\mathrm{F}}^2 \lesssim \frac{K\left(p^2 + q^2 + \ln(1/\delta)\right)}{n} \vee \frac{K\left(p^2 + q^2 + \ln(1/\delta)\right)^2}{n^2}$$

выполнено с вероятностью не менее $1 - \delta$

Теоретическая верхняя оценка на $\|\check{\Sigma} - \Sigma\|_{\mathrm{F}}^2$ из работы [Masak et al., 2022] асимптотическая и не дает улучшений по сравнению с неструктурированным случаем

Вопрос: можно ли доказать аналогичную безразмерную оценку?

Основной результат

Теорема: (П., Рахуба, 2024)

Предположим, что центрированный случайный вектор $\mathbf{X}$ удовлетворяет предположению (1) и что ковариационная матрица Σ представима в виде суммы K произведений Кронекера (2). Зафиксируем $\delta \in (0, 1)$, для которого выполнено

$$\frac{1}{2n}\left(\max_{1 \leqslant j \leqslant K} \mathtt{r}(\Phi_j)^2 + \max_{1 \leqslant j \leqslant K} \mathtt{r}(\Psi_j)^2\right) + \frac{\ln(4/\delta)}{n} \leqslant 1$$

и выберем произвольное $\lambda \in \mathbb{R}$, такое что

$$\lambda \geqslant 2\omega \left(\sum_{j=1}^{K} \|\Phi_j\| \|\Psi_j\|\right) \sqrt{\frac{13}{2n}\left(\max_{1 \leqslant j \leqslant K} \mathtt{r}(\Phi_j)^2 + \max_{1 \leqslant j \leqslant K} \mathtt{r}(\Psi_j)^2\right) + \frac{13\ln(4/\delta)}{n}}$$

Тогда, с вероятностью не менее $(1 - \delta)$, оценка $\widetilde{\Sigma}$, определенная в (4), удовлетворяет неравенству

$$\left\|\widetilde{\Sigma} - \Sigma\right\|_{\mathrm{F}}^2 < \frac{3\lambda^2 K}{2}$$

Замечания

- При правильном выборе параметра $\lambda > 0$ получаем оценку

$$\left\| \widetilde{\Sigma} - \Sigma \right\|_{\mathrm{F}}^2 \lesssim \frac{K}{n} \left(\sum_{j=1}^{K} \|\Phi_j\| \|\Psi_j\| \right)^2 \left(\max_{1 \leqslant j \leqslant K} \mathtt{r}(\Phi_j)^2 + \max_{1 \leqslant j \leqslant K} \mathtt{r}(\Psi_j)^2 + \ln(1/\delta) \right)$$

- Безразмерная оценка согласуется с верхней оценкой [Tsiligkaridis and Hero, 2013]

- Скорость сходимости заметно быстрее по сравнению с неструктурированным случаем: вместо

$$\mathrm{Tr}(\Sigma)^2 = \left(\sum_{j=1}^{K} \|\Phi_j\| \|\Psi_j\| \mathtt{r}(\Phi_j) \mathtt{r}(\Psi_j) \right)^2 \quad \text{(см., например, [Bunea and Xiao, 2015])}$$

имеем

$$K \left(\sum_{j=1}^{K} \|\Phi_j\| \|\Psi_j\| \right)^2 \left(\max_{1 \leqslant j \leqslant K} \mathtt{r}(\Phi_j)^2 + \max_{1 \leqslant j \leqslant K} \mathtt{r}(\Psi_j)^2 \right)$$

Основные идеи доказательства: шаг 1

Теорема: [Tsiligkaridis and Hero, 2013, Theorem 2]

Если $\lambda \geqslant 2\|\mathcal{R}(\widehat{\Sigma} - \Sigma)\|$, то

$$\left\|\widetilde{R} - \mathcal{R}(\Sigma)\right\|_{\mathrm{F}}^2 \leqslant \inf_R \left\{\|R - \mathcal{R}(\Sigma)\|_{\mathrm{F}}^2 + \frac{(1+\sqrt{2})^2}{4}\lambda^2 \mathrm{rank}(R)\right\}$$
$$\leqslant \frac{(1+\sqrt{2})^2}{4}\lambda^2 K.$$

Для доказательства безразмерной верхней оценки достаточно ограничить операторную норму $\mathcal{R}(\widehat{\Sigma} - \Sigma)$

Основные идеи доказательства: шаг 2

С помощью несложных алгебраических преобразований $\|\mathcal{R}(\hat{\Sigma} - \Sigma)\|$ можно представить в виде супремума эмпирического процесса

Лемма

Пусть случайные матрицы $\mathbb{X}, \mathbb{X}_1, \ldots, \mathbb{X}_n$ размера $(q \times p)$ таковы, что $\mathbf{vec}(\mathbb{X}) = \mathbf{X}$ и $\mathbf{vec}(\mathbb{X}_i) = \mathbf{X}_i$ для всех $i \in \{1, \ldots, n\}$. Тогда

$$\left\|\mathcal{R}(\hat{\Sigma} - \Sigma)\right\| = \sup_{\substack{U \in \mathbb{R}^{p \times p}, V \in \mathbb{R}^{q \times q} \\ \|U\|_{\mathrm{F}} = \|V\|_{\mathrm{F}} = 1}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathrm{Tr}\left(\mathbb{X}_i^\top V^\top \mathbb{X}_i U\right) - \mathbb{E}\mathrm{Tr}\left(\mathbb{X}^\top V^\top \mathbb{X} U\right) \right\}$$

Лемма: PAC-байесовское вариационное неравенство

Пусть $\mathbf{X}, \mathbf{X}_1, \dots, \mathbf{X}_n$ – н. о. р. случайные элементы на измеримом пространстве $\mathcal{X}$. Пусть Θ – пространство параметров с заданной на нем (априорным) распределением μ. Пусть $f : \mathcal{X} \times \Theta \to \mathbb{R}$. Тогда с вероятностью не менее $1 - \delta$ неравенство

$$\mathbb{E}_{\boldsymbol{\theta} \sim \rho} \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{X}_i, \boldsymbol{\theta}) \leqslant \mathbb{E}_{\boldsymbol{\theta} \sim \rho} \ln \mathbb{E}_{\mathbf{X}} e^{f(\mathbf{X}, \boldsymbol{\theta})} + \frac{\mathcal{KL}(\rho, \mu) + \ln(1/\delta)}{n}$$

выполнено одновременно для всех $\rho \ll \mu$.

Ключевой шаг: PAC-байесовское вариационное неравенство

$$\left\| \mathcal{R}(\widehat{\Sigma} - \Sigma) \right\| = \sup_{\substack{U \in \mathbb{R}^{p \times p}, V \in \mathbb{R}^{q \times q} \\ \|U\|_{\mathrm{F}} = \|V\|_{\mathrm{F}} = 1}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathrm{Tr}\left(\mathbb{X}_i^\top V^\top \mathbb{X}_i U \right) - \mathbb{E}\,\mathrm{Tr}\left(\mathbb{X}^\top V^\top \mathbb{X} U \right) \right\}$$

Основная идея – рассмотреть U и V как математические ожидания усеченной гауссовской меры

$$\rho_{U,V}(X, Y) \propto \mu(X - U, Y - V) \mathbb{1}\left[(X, Y) \in \Upsilon_{U,V} \right],$$

где $\Upsilon_{U,V}$ – центрально симметричное отностительно (U, V) множество,

$$\mu(X, Y) = \frac{\alpha^{p^2/2} \beta^{q^2/2}}{(2\pi)^{p^2/2 + q^2/2}} \exp\left\{ -\frac{\alpha}{2} \|X\|_{\mathrm{F}}^2 - \frac{\beta}{2} \|Y\|_{\mathrm{F}}^2 \right\}, \quad X \in \mathbb{R}^{p \times p},\ Y \in \mathbb{R}^{q \times q},$$

$$\alpha = \max_{1 \leqslant j \leqslant K} \mathtt{r}(\Phi_j)^2, \quad \beta = \max_{1 \leqslant j \leqslant K} \mathtt{r}(\Psi_j)^2$$

Ключевой шаг: PAC-байесовское вариационное неравенство

Зафиксируем $\nu > 0$

Согласно вариационному PAC-байесовскому неравенству,

$$\sup_{\substack{U \in \mathbb{R}^{p \times p}, V \in \mathbb{R}^{q \times q} \\ \|U\|_{\mathrm{F}} = \|V\|_{\mathrm{F}} = 1}} \left\{ \frac{\nu}{n} \sum_{i=1}^{n} \mathrm{Tr}\left(\mathbb{X}_i^\top V^\top \mathbb{X}_i U\right) - \nu \mathbb{E}_{\mathbb{X}} \mathrm{Tr}\left(\mathbb{X}^\top V^\top \mathbb{X} U\right) \right\}$$

$$\leqslant \sup_{\substack{U \in \mathbb{R}^{p \times p}, V \in \mathbb{R}^{q \times q} \\ \|U\|_{\mathrm{F}} = \|V\|_{\mathrm{F}} = 1}} \left\{ \mathbb{E}_{(P,Q) \sim \rho_{U,V}} \ln \left[\mathbb{E}_{\mathbb{X}} \exp\left\{ \nu \mathrm{Tr}\left(\mathbb{X}^\top Q^\top \mathbb{X} P\right) - \nu \mathbb{E}_{\mathbb{X}} \mathrm{Tr}\left(\mathbb{X}^\top Q^\top \mathbb{X} P\right) \right\} \right] \right.$$

$$\left. + \frac{\mathcal{KL}(\rho_{U,V}, \mu) + \ln(1/\delta)}{n} \right\}$$

с вероятностью не менее $(1 - \delta)$

Заключительный шаг

В условиях Теоремы, выполнено

$$\mathbb{E}_{(P,Q)\sim\rho_{U,V}} \ln \mathbb{E}_{\mathbb{X}} \exp\left\{\nu \mathrm{Tr}\left(\mathbb{X}^\top Q^\top \mathbb{X} P\right) - \nu \mathbb{E}_{\mathbb{X}} \mathrm{Tr}\left(\mathbb{X}^\top Q^\top \mathbb{X} P\right)\right\} \leqslant 13\omega^2\nu^2 \left(\sum_{j=1}^K \|\Phi_j\| \|\Psi_j\|\right)^2$$

для всех ν, удовлетворяющих неравенству

$$13\omega^2\nu^2 \left(\sum_{j=1}^K \|\Phi_j\| \|\Psi_j\|\right)^2 \leqslant 1$$

Кроме того,

$$\mathcal{KL}(\rho_{U,V}, \mu) \leqslant 2\ln 2 + \frac{1}{2}\left(\max_{1\leqslant j\leqslant K} \mathtt{r}(\Phi_j)^2 + \max_{1\leqslant j\leqslant K} \mathtt{r}(\Psi_j)^2\right) \quad \text{для всех } \|U\|_{\mathrm{F}} = \|V\|_{\mathrm{F}} = 1$$

Список литературы I

[Adamczak, 2015] Adamczak, R. (2015).
A note on the Hanson-Wright inequality for random vectors with dependencies.
Electronic Communications in Probability, 20:1–13.

[Adamczak et al., 2010] Adamczak, R., Litvak, A. E., Pajor, A., and Tomczak-Jaegermann, N. (2010).
Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles.
Journal of the American Mathematical Society, 23(2):535–561.

[Adamczak et al., 2011] Adamczak, R., Litvak, A. E., Pajor, A., and Tomczak-Jaegermann, N. (2011).
Sharp bounds on the rate of convergence of the empirical covariance matrix.
Comptes Rendus Mathématique. Académie des Sciences. Paris, 349(3-4):195–200.

Список литературы II

[Bickel and Levina, 2008a] Bickel, P. J. and Levina, E. (2008a).
Covariance regularization by thresholding.
The Annals of Statistics, 36(6):2577–2604.

[Bickel and Levina, 2008b] Bickel, P. J. and Levina, E. (2008b).
Regularized estimation of large covariance matrices.
The Annals of Statistics, 36(1):199–227.

[Bourgain, 1999] Bourgain, J. (1999).
Random points in isotropic convex sets.
In *Convex geometric analysis (Berkeley, CA, 1996)*, volume 34 of *Mathematical Sciences Research Institute Publications*, pages 53–58. Cambridge University Press, Cambridge.

[Bunea and Xiao, 2015] Bunea, F. and Xiao, L. (2015).
On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA.
Bernoulli, 21(2):1200–1230.

Список литературы III

[Cai et al., 2013] Cai, T. T., Ren, Z., and Zhou, H. H. (2013).
Optimal rates of convergence for estimating Toeplitz covariance matrices.
Probability Theory and Related Fields, 156(1-2):101–143.

[Cai et al., 2010] Cai, T. T., Zhang, C.-H., and Zhou, H. H. (2010).
Optimal rates of convergence for covariance matrix estimation.
The Annals of Statistics, 38(4):2118–2144.

[Cai and Zhou, 2012] Cai, T. T. and Zhou, H. H. (2012).
Optimal rates of convergence for sparse covariance matrix estimation.
The Annals of Statistics, 40(5):2389–2420.

[Fan et al., 2015] Fan, J., Rigollet, P., and Wang, W. (2015).
Estimation of functionals of sparse covariance matrices.
The Annals of Statistics, 43(6):2706–2737.

Список литературы IV

[Giannopoulos et al., 2005] Giannopoulos, A., Hartzoulaki, M., and Tsolomitis, A. (2005).
Random points in isotropic unconditional convex bodies.
Journal of the London Mathematical Society, 72(3):779–798.

[Han, 2022] Han, Q. (2022).
Exact spectral norm error of sample covariance.
Preprint, arXiv:2207.13594.

[Hanson and Wright, 1971] Hanson, D. L. and Wright, F. T. (1971).
A Bound on Tail Probabilities for Quadratic Forms in Independent Random Variables.
The Annals of Mathematical Statistics, 42(3):1079–1083.

[Kannan et al., 1997] Kannan, R., Lovász, L., and Simonovits, M. (1997).
Random walks and an $O^*(N^5)$ volume algorithm for convex bodies.
Random Structures & Algorithms, 11(1):1—-50.

Список литературы V

[Koltchinskii and Lounici, 2017] Koltchinskii, V. and Lounici, K. (2017).
Concentration inequalities and moment bounds for sample covariance operators.
Bernoulli, 23(1):110–133.

[Leng and Pan, 2018] Leng, C. and Pan, G. (2018).
Covariance estimation via sparse Kronecker structures.
Bernoulli, 24(4B):3833–3863.

[Masak et al., 2022] Masak, T., Sarkar, S., and Panaretos, V. M. (2022).
Separable expansions for covariance estimation via the partial inner product.
Biometrika, 110:225–247.

[Paouris, 2006] Paouris, G. (2006).
Concentration of mass on convex bodies.
Geometric and Functional Analysis, 16(5):1021–1049.

Список литературы VI

[Rudelson, 1999] Rudelson, M. (1999).
Random vectors in the isotropic position.
Journal of Functional Analysis, 164(1):60–72.

[Tropp, 2012] Tropp, J. A. (2012).
User-friendly tail bounds for sums of random matrices.
Foundations of Computational Mathematics, 12(4):389–434.

[Tsiligkaridis and Hero, 2013] Tsiligkaridis, T. and Hero, A. O. (2013).
Covariance estimation in high dimensions via Kronecker product expansions.
IEEE Transactions on Signal Processing, 61(21):5347–5360.

[Vershynin, 2012] Vershynin, R. (2012).
Introduction to the non-asymptotic analysis of random matrices.
In *Compressed Sensing: Theory and Applications*, pages 210–268, Cambridge. Cambridge University Press.

Список литературы VII

[Xiao and Wu, 2012] Xiao, H. and Wu, W. B. (2012).
Covariance matrix estimation for stationary time series.
The Annals of Statistics, 40(1):466–493.

[Zhivotovskiy, 2024] Zhivotovskiy, N. (2024).
Dimension-free bounds for sums of independent matrices and simple tensors via the variational principle.
Electronic Journal of Probability, 29:1–28.

Подробности в наших публикациях:

N. Puchkin, M. Rakhuba
“Dimension-free Structured Covariance Estimation”
Proceedings of Thirty Seventh Conference on Learning Theory, PMLR 247:4276-4306, 2024
https://proceedings.mlr.press/v247/puchkin24a.html

N. Puchkin, F. Noskov, V. Spokoiny
“Sharper dimension-free bounds on the Frobenius distance between sample covariance and its expectation”
Bernoulli (to appear), 2025+
https://arxiv.org/abs/2308.14739

Спасибо за внимание!