Оценка спектральных проекторов выборочных ковариационных матриц с помощью бутстреп-метода

Шумовская Валентина Сергеевна

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: valentinashumovskaya@gmail.com

Задача восстановления спектральных проекторов (собственных векторов и собственных подпространств) ковариационных матриц высокой размерности по выборке наблюдений является одной из ключевых задач статистики и имеет непосредственное отношение к задачам снижения размерности. В частности, широко распространенный метод главных компонент состоит в том, что данные в высокой размерности проецируются на подпространства, натянутые на первые собственные векторы (отвечающие наибольшим собственным значениям). Однако, проблема восстановления спектральных проекторов ковариационных матриц высокой размерности достаточно плохо изучена. Недавно, в работе [1] были получены неасимптотические оценки нормы Фробениуса расстояния между выборочным и истинным проекторами, а также исследовано асимптотическое поведение. Эта работа позволяет построить асимптотические доверительные интервалы для истинного проектора. Однако, хорошо известно, что такие асимптотические результаты применяются только для выборок очень большого объема. В частности, это связано с тем, что скорость сходимости нормализованных U-статистик, возникающих в данной задаче, к предельному закону слишком медленная. Цель настоящей работы - построить доверительные множества для истинного проектора с помощью бутстреп-метода. Предлагаемый подход к построению доверительных множеств с помощью бутстреп-метода работает хорошо даже в случае, когда размерность пространства значительно превышает объем выборки.

Пусть $X_1, \ldots, X_n \in \mathbb{R}^p$ — независимые одинаково распределенные случайные векторы с нулевым средним и неизвестной ковариационной матрицей $\mathbf{\Sigma}$. Классической оценкой неизвестной матрицы $\mathbf{\Sigma}$ служит выборочная ковариационная матрица:

$$\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{j=1}^{n} X_j X_j^{\mathsf{T}}. \tag{1}$$

Заметим, что $\mathbb{E}\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma}$. Рассмотрим также бутстреп-версию выборочной ковариационной матрицы:

$$\mathbf{\Sigma^o} := \frac{1}{n}\sum_{j=1}^{n} \omega_j X_j X_j^{\mathsf{T}}, \tag{2}$$

где $\omega_1, \ldots, \omega_n$ — н.о.р. случайные величины, такие, что $\mathbb{E}\omega_j = 1, Var(\omega_j) = 1$. Беря математическое ожидание $\mathbb{E}^o$ относительно $\omega_1, \ldots, \omega_n$ при фиксированных $X_1, \ldots, X_n$ получим, что $\mathbb{E}^o\mathbf{\Sigma^o} = \widehat{\mathbf{\Sigma}}$, т.е. в мире "бутстрепа" математическое ожидание известно. Пусть $\mu_1 > \mu_2 > \ldots$ – различные собственные значения матрицы $\mathbf{\Sigma}$, а $\mathbf{P}_r$ - проектор на собственное подпространство, отвечающее μ_r. В предположении существования спектральной дырки (spectral gap condition), можно определить соответствующие выборочные и бутстреп проекторы, $\widehat{\mathbf{P}}_r, \mathbf{P}_r^o$ соответственно. В работе численно сравниваются распределения

$$n\|\mathbf{P}_r - \widehat{\mathbf{P}_r}\|_2^2 \tag{3}$$

и

$$n\|\mathbf{P}_r^o - \widehat{\mathbf{P}_r}\|_2^2 \tag{4}$$

Численные эксперименты произведены на суперкомпьютере «Ломоносов». Результат одного из них показан на Рис. 1. Основные характеристики ковариационной матрицы представлены в описаниии к иллюстрациии.

Иллюстрации

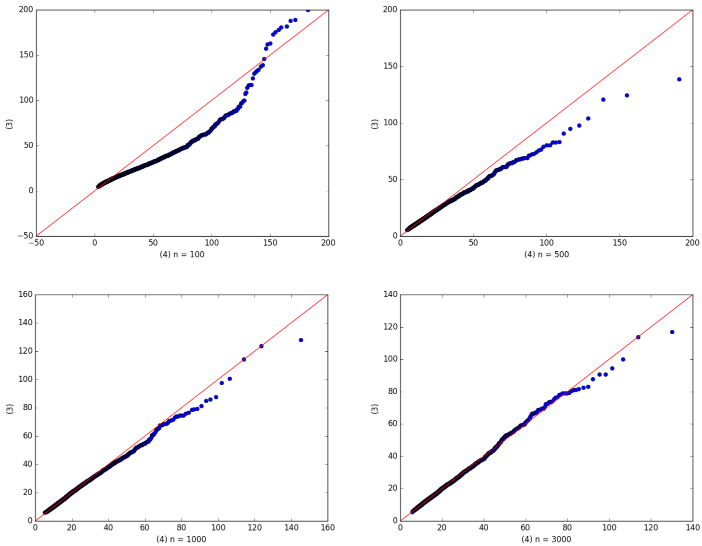

Рис. 1. QQ-plots, p = 100,
$\mu_1 = 26.7, \mu_2 = 15.8, \mu_3 = 10.1, \mu_4 = 5.9, \mu_5 = 3.5, |\mu_j| < 1.5, j > 5$

Литература

1. Koltchinskii V and Lounici K., 2015b // Normal approximation and concentration of spectral projectors of sample covariance, arXiv:1504.07333